View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Wales Trinity Saint David

The Contemporary Relevance of Kant's Transcendental Psychology

Deborah Maxwell Alamé-Jones BA (Hons)

Supervisor: Dr David Morgans

Submitted in partial fulfilment for the award of the degree of Doctor of Philosophy

UNIVERSITY OF WALES TRINITY SAINT DAVID

2018

DECLARATION SHEET

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

Signed .....Deborah Alame-Jones.....

Date ..... 17th May 2018.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Sources are acknowledged by giving explicit references in the body of the text. A bibliography is appended.

Signed ..... Deborah Alame-Jones.....

Date ..... 17th May 2018.....

STATEMENT 2

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

Signed ..... Deborah Alame-Jones.....

Date ..... 17th May 2018.....

STATEMENT 3

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

Signed ..... Deborah Alame-Jones.....

Date ..... 17th May 2018.....

ACKNOWLEDGEMENTS:

Firstly, I would like to gratefully acknowledge the guidance, support and encouragement of my doctoral supervisor, Dr David Morgans, who took over the role of supervisory advisor midway through my candidature and to whom I am eternally indebted.

I would also like to thank Professor Kelvin Donne for believing in the value of my proposal and my previous supervisor, Dr. Barry Ip, for his cheerful presence and invaluable comments whilst I was enrolled at the University of Wales, Trinity St David's Swansea Campus. This thesis would have remained a dream had it not been for the encouragement and support of my colleagues and friends there, and for this I owe them my deepest gratitude.

I also wish to thank the members of the Post Graduate Research team for their invaluable advice and support, and for their swift responses to my many questions

Lastly, I would like to thank my family for their unconditional support and patience, and especially my husband Christopher for his tolerance of my almost complete absence of any free time, particularly during the latter stages of this work.

Abstract

The purpose of this thesis is to demonstrate the contemporary relevance of Kant's transcendental psychology against the orthodoxy of the dominant analytic school of philosophy, in an aim to salvage it from criticisms that resulted in the widespread view that Kant had little to say about the mind that was correct or useful. Historically, this had led to the near exclusion of Kant's views of the mind from mainstream philosophical debate; those who acknowledged intellectually the psychological import of the work deemed as having transgressed the bounds of proper philosophy. It is argued that this was, and still is, an unfortunate and narrow view, since an interpretation which fully embraces the transcendental aspect can provide invaluable insights and direction for contemporary research in cognitive science and cognitive neuroscience. A major focus of this work is to provide a rigorous conceptual analysis of the modern problem of consciousness and to show that every approach has become a response, positive or negative, to the Cartesian distinction between body and mind. Today, more than three hundred years after Descartes' philosophical dualism, this powerful and persuasive argument still continues to hold fast. Cognitive neuroscientists have amassed a deep and detailed understanding of how our brains process information from the external world, but the question of how this information is transformed into conscious experience is deemed an unsolved problem. It is proposed that, although Kant never uses the concept of consciousness in the now dominant sense of phenomenal *qualia*, his theory of the transcendental subject is a valuable tool in unravelling the philosophical complexities that are commonplace in current theories.

Contents

1. Introduction

The purpose of this thesis is to demonstrate the contemporary relevance of Kant's theory of mind, his "transcendental psychology" to cognitive science and cognitive neuroscience and also to the modern problem of consciousness, arguably the central issue in contemporary philosophy of mind today. Although written over two hundred years ago, his critical philosophical writings indicate that he had profound and original insights about mental functioning and cognition that have considerable significance. He held ideas about the nature of mental processing, the unity this requires, and also of consciousness and self-consciousness; some of which have already been assimilated into the cognitive sciences whereas others have only quite recently started to be appreciated. This thesis examines the relevance of Kant's work in this regard in order to bring out the particular influences that these ideas have had as well as what can still be learned from them.

One idea that has had a particularly strong influence is his central method of transcendental argument, which has become a major methodological tool, and is known more readily as "inference to the best explanation". This is the view that if we are to study the mind, we must infer the conditions necessary for experience by means of the postulation of unobservable cognitive mechanisms which explain outwardly observable behaviour. Kant, in his major work *The Critique of Pure Reason*, had delineated in the abstract these cognitive conditions: It is, for example, a methodological requirement for cognition that what he termed "the manifold" of sense data form a temporal series, have spatial contiguity, and causal connectedness. These are necessary conditions for all cognitive pursuits without which nothing is experienced. In short, his examination of the necessary prerequisites for cognition has the structure of a transcendental argument from which the justification of our abstract cognitive apparatus can be inferred (See Lyre, 2006).

As Lyre states "the reconstruction of transcendental arguments as inferences to the best explanation heavily undermines Kant's own far more rigorous understanding of his enterprise" (Lyre, 2006 p. 493). Nevertheless, the inference to the best explanation lies at the core of his critical enquiry and as such is regarded as providing the necessary link between two disparate realms, of outwardly observable

behaviour and meaningful discourse and occult psychological or mental correlates or antecedents. Regardless of its non-empirical Kantian roots, contemporary cognitive scientists, in a similar vein, also try to justify their assertions by showing that they provide the necessary presuppositions for the possibility of meaningful discourse about certain aspects of the mind and of observable behaviour, presuppositions which cannot be denied without contradiction. In other words, in order to give epistemic justification to their claims, they also try to show what must be the case in any system in order for a particular phenomenon to occur, and to do so without appealing to psychological "facts", but rather by appealing to certain "conceptual prerequisites". Kant, in *The Critique of Pure Reason*, identifies a number of these conceptual prerequisites, in terms of *a priori* cognitive functions that are necessary for the possibility of cognition. These include certain "faculties", the "pure forms of intuition", space and time, and the twelve "categories" of the understanding, which are the most fundamental concepts of thought, without which nothing could be intelligible (A93–94/B126).

Kant distinguishes between "transcendental" or "pure" and "empirical" aspects of human cognition (A20/B34, A34/B60, A51/B75, B81-82) and describes transcendental functions as configuring the empirical faculties of human consciousness by providing the structure they need to become faculties of cognition. The empirical faculties concern our receptivity to the action of affecting objects on our minds (objects are given through sensibility in "intuitions" (A19/B33)) and are *a posteriori*; the transcendental features are *a priori*, since they cannot be derived from sensation, but represent the ways in which sensation is necessarily ordered in experience. According to Kant, the transcendental features are part of, or derived from "spontaneous" cognitive capacities, whereas the empirical features are directly related to, or derived from, what is given in sensibility. Thus, Kant offers an account of how we are *passive* in our perceptual experience of the world whilst *active* in how we conceptualise it, through what he calls the spontaneity of "judgment" in perceptual experience. Kant maintains that our understanding of the world has its foundations not merely in "intuitions", but in both intuitions and the *a priori* concepts which order it. As he put it, famously, "Thoughts without content are empty, intuitions without concepts are blind" (A51/B75).

It has been noted by several philosophers viz: Wilfrid Sellars (1974), Jay Rosenberg (1986), Ralf Meerbote (1989), Thomas Powell (1990), Patricia Kitcher (1993, 2006) Andrew Brook (1994, 2004, 2008) that Kant's ideas bear a remarkable similarity to contemporary functionalist explanation in cognitive science, and each have explored different functionalist interpretations of Kant. It was first argued by Wilfrid Sellars that this is Kant's "unknowability thesis" we "do not know [mental] processes save as processes which embody these functions" (Sellars, 1974, pp. 62-90). Andrew Brook, in a similar vein, claims that this is a part of his doctrine of *noumenalism* about the mind, that Kant was the first to articulate the methodological insight into the relation of "concepts" and "percepts", and that his general picture of the mind was as "a system of concept using functions" (Brook, 1994, p. 14). Similarly, Patricia Kitcher describes Kant's theory of mind as an account of the representational content of judgments "like that defended by contemporary functionalists", an abstract theory, and a product of an essentially interconnected causal process, of the combination of separate mental states into other mental states (Kitcher, 1993, p. 111)[1]. Thus, there has been recognition by scholars of the common ground that exists between contemporary functionalist explanation in cognitive science and Kant's own project. The claim is that the Kantian notion that sense experience has to be conceptually ordered if there is to be the possibility of coherent thought is alive and well in the fundamental tenet and methodology of cognitive science and cognitive neuroscience.

Functionalism, as a theory of how to model the mind is *the* most influential contemporary empirical theory, and has given rise to a huge amount of research claiming to be able, eventually, to explain much about the mind that presently remains beyond our understanding. There are two serious research programs into the nature of cognition, the classical "language of thought" paradigm, introduced by Jerry Fodor in his 1975 book *The Language of Thought* and the later "connectionist" research program which resulted from two large volumes of work completed in 1986 by David E. Rumelhart, James L. McClelland and colleagues in the Parallel Distributed Pressing (PDP) research group. Both assume that the scientific basis of cognition is computational. The theory comes in several other forms, but the central idea is that what matters is how the mind works at an abstract level rather than how it is constituted in a physical system. According to functionalism, the essential nature

of mental states such as desires, pains, itches, "inner feelings", the sensation of green, etc., what philosophers call *qualia*, does not reside in any material that they may be found in, but rather in the *function* that each performs. The phenomenal, experiential or the oft-mentioned "what it's likeness" [2] aspect of cognition is disregarded, as in all reductive, materialist accounts. Mental states are constituted by their functional role within a cognitive system at a theoretical level where they are identified by their role within the whole system rather than on any material or introspected or felt natures. Kant's view of the mind can also be regarded as a system of functions (as aforementioned, he calls them "faculties") for applying "concepts" to "percepts". In the first *Critique*, he explicitly rejects the Cartesian conception of the mind as a "substance", particularly in the Paralogisms (a list of logical fallacies that arise when reason erroneously tries to operate beyond the limits of possible experience) where he argues that at this theoretical level we cannot know the nature of any substrate of this system which is the mind. Mental states can be sufficiently explained without taking into account any underlying medium, physical or otherwise. All we can know is what the mind must be like in terms of its organisation and function, i.e. the necessary prerequisites of cognition at an abstract or theoretical level. No conclusions can be drawn about the subject's underlying nature. For example, in the third Paralogism, Kant corrects the Cartesian view of personal identity by stating that the "consciousness of myself at different times" is only a "formal condition" for personality or personal identity (A363). It is not a type of "substance", as Descartes had claimed. In other words, the unity of an individual is merely that of a single, thinking, temporally successive subject that, in the jargon of computer science, is *multiply realisable* over personally discontinuous psychological states. This is akin to the fundamental tenet of functionalism "the mind is what the mind does" as well as the negative dictum "function does not determine form".

A further Kantian idea, noted by Andrew Brooks and Patricia Kitcher, is the notion of synthesis, which is nowadays termed "the binding problem". This is the problem of accounting for the ways in which the separate features of objects, such as colour, shape, boundary, and texture are integrated to produce a recognisable whole. Ann Treisman, emeritus professor of psychology at Princeton University, developed a solution to this called the Feature Integration Theory (1980, 1996, 2003, 2005,

2006) which is akin to Kant's theory in quite significant ways and is regarded as one of the most influential psychological models of human visual attention in the field, forming the basis for thousands of experiments in cognitive psychology, cognitive science and cognitive neuroscience. According to the Feature Integration Theory, the mind must synthesise or bind representations of the world into single objects by means of a three-stage processing mirroring Kant's own notions of apprehension, reproduction and recognition in concepts (see Brook, 2004).

Until fairly recently, cognitive science had been more concerned with mental content and the processing of content rather than on phenomenology or "consciousness" *per se*; for example, Treisman's binding problem is concerned with our awareness of objects, how object recognition proceeds in three stages, but not in consciousness itself. So whereas there had been a lot of progress in investigating empirically those aspects of human behaviour that is ordinarily *linked* to consciousness, such as attention, memory, the ability to discriminate, categorise, and react to environmental stimuli, etc., consciousness *itself* was left out of the picture as it does not fit neatly into empirical science. It had been ignored because it had appeared so clearly impossible to say anything constructive about it within the materialist presuppositions of cognitive science. Proper respect for the objective, empirical nature of science seemed to require the denial of the very existence of consciousness. During the last twenty or so years, however, there have been renewed attempts to account for the "mystery" of consciousness within a scientific paradigm. But this has given rise to the "hard problem of consciousness", viewed by many as "the last great mystery for science" (Chalmers, 1995). In fact, the study of consciousness, once considered taboo, since it was too abstract, too subjective or too difficult to study scientifically, re-emerged as one of the hottest new fields in neuroscience, particularly in neurobiology; it has become a much discussed topic in philosophy also, and has resulted in the publication of hundreds of thousands of papers, articles and books. Indeed, the "hard" problem of consciousness has perplexed philosophers and scientists ever since: Where does consciousness come from? It seems to "exist", at least in our own case. We cannot deny or doubt that we are conscious: when we introspect, there is an ever-changing flux of sensations, thoughts, memories and feelings that comprise our world. But, the question posed is, "how does all this arise from the nerve cells, chemistry and electricity in the brain?"

Nowadays, a common view among proponents is that everything to do with the mind, including consciousness, will eventually turn out to be explicable in terms of neurophysiology, or that it may even lie within the domain of physics (Roger Penrose & Stuart Hameroff, 2011; Henry Stapp, 2009, 2014). It is also an area of intense philosophical debate with opinions as polarised as the disciplines themselves. Zombie thought experiments are offered, i.e. there could be beings that are behaviourally, cognitively, and even physically identical to us, yet lacking conscious experience, there is "no one at home" (Chalmers, 1995). Or inverted spectrum thought experiments; where it is imagined that people have radically different conscious experiences of colour; for example, they might see red where everyone else sees green but, due to their linguistic training in the use of colour words, they react to coloured objects and even process information about colour in exactly the same way as everyone else. These thought experiments are intended to show that mental representations can have functionality as representations without the input of consciousness, which means that there is something crucial left out, viz., first-person phenomenal experience. Theories attempting to account for this deficit are almost too numerous to mention. They range from attention theories to global work space theories, to recent neuroscientific programs aimed at identifying certain regions and systems in the brain most closely associated with consciousness of various kinds, the so-called NCCs, short for the "neural correlates of consciousness".

Francis Crick and Christof Koch were the first to explicitly connect neural synchronisation with a theory of binding and consciousness, their hypothesis being that binding involves an attentional mechanism that temporarily binds the relevant neurons together. In their seminal paper "Towards a neurobiological theory of consciousness" (1990) they hypothesised that neurones generate consciousness through coherent semi-oscillations spiking at a frequency within the 40-70Hz range. The vast plethora of books, papers and research articles produced in its wake, means that looking for the NCCs is now the dominant scientific method of investigating consciousness, (see Crick and Koch, 1995, 1998; Chalmers, 2000; Metzinger, 2000, 2003, 2011; Koch 2004; Block, 2005; Bayne, 2007, 2010; Tononi and Koch, 2000, 2008, 2015; Hohwy, 2007; 2009; Kiverstein, 2009. Oizumi, Albantakis, Tononi 2014, Hohwy & Bayne, 2015). A fairly recent theory is the Integrated Information Theory or ITT, developed by Giulio Tononi (Tononi, 2004, 2008, 2012, 2015) at the

University of Wisconsin, that assigns to the brain, or any complex system, a number denoted by the Greek letter Φ that tells you how integrated a system is, or how much more the system is than the union of its parts. Any system with integrated information different from zero has consciousness. Information theorists measure the amount of information in a computer file in bits, and, according to Tononi, we could, in theory, also measure consciousness in bits, by means of a consciousness metre. When we are wide awake, for example, our consciousness should contain more bits than when we are asleep.

No matter the scientific model of the mind, it seems that the charge can always be made that the model is studying function, information or at most mere correlates, that it is not uncovering the nature of consciousness itself. Yet, others claim that function is *all* that is required to understand mind, and that we can dispense with consciousness altogether. This area of intense philosophical disputation with its deeply polarised opinions shows little sign of abating - the problem of consciousness is firmly entrenched in the scientific research arena as the vast amount of literature on the topic shows. Yet it seems that the more we study the brain and its functions and processes, the less we understand the reasons for its connection to the conscious mind, as the multifarious attempts to close the so-called "explanatory gap" (Levine, 1983, 2001) attests to. This fact by itself would seem to *strongly* indicate that somewhere along the line there has been a fundamental error or false assumption. However, it is a testimony to the power of the grip of old ideas on the minds of scientists and philosophers alike that this has been paid scant attention to in the literature, and is perhaps it is also where scientists' "lay" intuitions have been the source of the confusion. We tend to indulge our culturally acquired, socially upheld habit of treating commonly used words as descriptive of something that exists quite independent of what we say, that they *refer* to something, i.e. they are somehow "objectified" and this can cause conceptual problems, particularly so with such a slippery term as "consciousness". The word is a murky one. What exactly is it? What does it refer to; does it pick out any properties? Does it reside in the brain?

Philosophers and non-philosophers alike differ in their intuitions concerning what consciousness is. Given the slipperiness of the concept one would expect that advocates of the existence of a "hard problem" of consciousness would have considerable arguments backing up their use of the term. Often, however, the nature

of problem is treated as self-evident and not argued for at all, and the existence of consciousness is posited as a brute fact. This has led to some quite radical speculative metaphysics of the kind Kant warns us against, such as that of pan-protopsychism (Chalmers, 2003, 2013) and epiphenomenal property dualism (Chalmers, 1996 and early Frank Jackson, 1982). Kant's transcendental method, however, allowed him to analyse the metaphysical requirements of cognition without venturing into such speculative and ungrounded metaphysics. According to Kant, since the mind itself plays an active role in constituting the very features of phenomenal experience and therefore of empirical knowledge, it cannot itself be a known "object" of that knowledge, as some kind of "thing" but is known *a priori*.

Put another way, if the "transcendental unity of apperception" is a necessary condition for experience, it cannot then be given *in* experience. Such a mind does not belong to the world and cannot be studied empirically, at least not as something that resides "in the head". In relation to this, because many if not most aspects of neural dynamics, structure, and function can be modelled computationally, the question has been posed of whether there could be an equivalent of a conscious mind *in silico*? Could a machine have a mind in the same way that humans do and feel how things are? As Susan Stuart notes in a paper on the subject "[I]n the area of machine consciousness there have been and are efforts to create consciousness artificially", citing Cotterill, J. 1995, 1998; Haikonen, 2003, 2007; Aleksander and Dunmall 2003; Sloman, 2004, 2005; Aleksander, 2005; Holland and Knight, 2006 and Chella and Manzotti, 2007. She also notes "that a similar amount of effort has gone in to demonstrating the infeasibility of the whole enterprise" citing Dreyfus, 1972, 1979, 1992, 1999; Searle 1980, Harnad, 2003; Sternberg, 2007 (Stuart, 2010, p. 37).

Although artificial models of consciousness have replicated certain functions of the brain, and may have applications for neurological research, there has not been nor is there likely to be in the foreseeable future, a phenomenally conscious machine, although some AI scientists are optimistic about its possibility, even stating that "machine consciousness will almost certainly be achieved perhaps as soon as in the next two decades" (Reggia, 2013, p. 129). It has also been speculated, by futurists, that the enterprise of developing an understanding of machine consciousness could enable us to upload a conscious human mind onto a machine, thereby prolonging infinitely a human's life (Goertzel and Ikle, 2012). However, it is a matter of dispute

whether there is any *sense* in which artificial, computationally based constructs could ever be the loci of meaning and of phenomenal consciousness. One of the many problems is that, although there may be certain behaviours associated with consciousness, there is unlikely to be a way in which objective third-person tests could ever have access to first-person phenomenological experience, whether in a human or machine. This is related to the "other minds" problem and also to the question of whether objective, scientific theorising about consciousness, where we view the world as we normally do, as a mind-independent "real world" comprising the totality of objects, can ever adequately conceptualise what it is to have meaningful conscious experience within its categorical framework.[3]

This brings us to one of the main arguments of this thesis, namely, that the scientific/ materialistic, neurobiological/functional paradigm that lies at the root of cognitive science, combined with a fair deal of over-intellectualising has led to the creation of the hard problem of consciousness, which is continually reproduced by those who adhere to it, and made harder by attempts to solve it. It is suggested that this is also because the computer model of the mind has held cognitive scientists captive, and has metaphorically structured debates and discussions such that it hides the very peculiar nature of the question being asked. According to the Turing-Church principle that lies at the foundation of traditional cognitive science, we live in a Cartesian mechanical universe, and our minds are replicable within reasonable accuracy on a universal Turing machine; we are simply Cartesian machines, with no place for the mind. The problem that has resulted from this is how to account for mind?

It is suggested that this is a problematic understanding of human nature. We are not simply machines, as Kant, not only in the *Critique of Pure Reason*, but also the *Critique of the Power of Judgement* and other writings before and afterwards, is at pains to point out. Although he may not have used the concept of consciousness in the now dominant sense of phenomenal *qualia*, his theory of the transcendental subject is a valuable tool in addressing the philosophical complexities that have arisen in the field. His is a theory of active agency, involving "judgment", recognition and awareness of objective content as well as of the "apperceiving" subject, which is quite radical. Kant was interested, in particular, with the need for acts of synthesis or binding and the kind of unity required for such acts and in this he

has much to offer contemporary debate. His "transcendental psychology" is an account of the faculties that are required for a mind to be an *agent* or *subject* of cognition as much as it is an account of the conditions that are required universally and necessarily for something to be an *object* of cognition, as in traditional accounts. Kant stresses this need. Not only does he stress this but he also stresses the unity of the subject and of the subject's experiences. Kant thought that human self-consciousness can be regarded from two perspectives: as the "logical subject" of thought or "transcendental subject" and as the "object of inner sense", the empirical self. He writes of "transcendental apperception", or "the transcendental unity of self-consciousness" (B132) as the fundamental condition for the cognition of objects in the phenomenal world. The fact that we can make a judgment at all presupposes this unity of consciousness in a subject who is synthesising or combining representations or percepts according to the categories of experience, which are the *a priori* rules of the understanding. Empirical apperception, on the other hand, refers to consciousness of the particular contents of the subject's own mental states (A107). When we talk about the transcendental subject, or the "transcendental unity of self consciousness", however, such consciousness is not of a kind of intelligible object, but of oneself as "subject" through "spontaneous" acts of synthesis. When one is aware of oneself this way, one is aware of one's mind "as it is", by being a spontaneous synthesising agent. Self-consciousness as inner sense or receptive consciousness of what we passively "undergo" (as we are affected by the play of our own thought) is differentiated from consciousness of our activity, i.e. of what we are "doing" in synthesising or unifying our experience (1798, Ak. vii, p. 161).[4] The latter is not consciousness of the self as object, of the ordinary, changing self, but as "*subject*'" through which experience is unified. Kant's point is that we must be able to have some kind of cognitive access to ourselves as a condition of our ability to perceive objects; that is, as a condition of our capacity to "apply the categories" to "the manifold of intuition" which is independent of and logically prior to such ability.[5] For Kant it is essential for the whole idea of the spontaneous activity of synthesis that experience is not something passive that happens to us. Perceptual data has to be ordered and classified and this is something we *do.* Kant held the view the "manifold" must conform to the rules of the human mind, and that it is the knowing subject who, through the spontaneous activity of the mind creates order within

nature. Importantly, this introduced the picture of a human mind as an active originator of experience rather than just a passive recipient of perception.

So here we come to the heart of the matter, what does all this mean for cognitive science? Kant, it would seem, had recognised over two hundred years before the cognitive revolution of the twentieth century that the word *consciousness* has a number of different connotations, ranging from awareness of one's perceptions and sensations, what he would call "empirical apperception", to "transcendental apperception", the perception of oneself as an agent endowed with intentionality and free will. Viewed this way Kant's theory is amenable to the idea that empirical awareness of what we "undergo"[6] is workable under a materialistic neurobiological/ functional descriptive system because the problem can be reduced to questions about how cognitive processes are organised. However, in contrast, the latter connotation, of oneself as agent, transcends such a description. In relation to the binding problem mentioned above, for example, perceptions are bound together, whether they are conscious perceptions or not. As was mentioned, the processes that bind the percepts occur independently of consciousness, are subconscious processes, so this aspect of unity is not one that a theory of consciousness *per se* has to explain. The questions that remain to be explained and that do not fit easily into the classical scientific paradigm are: Why is it that my conscious perceptions belong, as it were, to me? What is the nature of this "I" that the perceptions belong to? And the most problematic: How can a physical brain made of matter give rise to conscious experiences or ineffable *qualia*?

During the last twenty years or so, the study of consciousness, eliminated from the field due to the rise of behaviourism and the overarching difficulty of accounting for it within a materialist science, has again become part of current research agenda, and as mentioned, a vast body of literature has been written. However, Kant's insights into the synthesis and combination of mental content that are required for this unity, as well as the kind of unity this entails, have been under-appreciated. This thesis is an attempt to bring to the fore Kant's original contributions, especially in relation to "the hard problem of consciousness", the problem of integrating consciousness into our conception of nature; the problem of why, besides the "intentional states" and "informational processes" that are said to be the underlying mechanisms responsible for cognition, there should be actual conscious events and

feelings, and why there *seems* to be an unbridgeable chasm between the mental realm and the physical medium in which it is instantiated. There appears to be a gap in our understanding. This is considered by some to be the most important scientific question to be solved in this present age. As John Searle has put the matter:

> The most important scientific discovery of the present era will come when someone - or some group - discovers the answer to the following question: How exactly do neurobiological processes in the brain cause consciousness? (Searle, 1993, 2008, p. 61).

Kant's epistemological solutions, systematically inferred via his central methodological innovation, his method of transcendental argument, are of contemporary significance to this debate. According to Kant, although we are prone to separate objective spatial experience and inner subjectivity, rendering the former "the world of things", the other the world of consciousness, these are not two different realities that might somehow come together, but at a fundamental level they cannot be defined separately from each other.

As Daniel Dennett notes, cognitive scientists are in the grip of delusions and confusions about human cognition and that cognitive science is "a land of plenty for philosophers" since so many of their questions are "ill thought out" (Dennett, 2009, p. 232). [7] If Kant were alive today he would surely agree, for he can be viewed as not only the "intellectual godfather of contemporary cognitive science" (Brook, 1994, p. 12; 2004, p.1) but also a fellow worker in the field. However, he held a fundamentally different view of human nature in stark contrast to the reductionist, deterministic, and mechanistic picture that is painted by cognitive scientist today. He was a revolutionary, with novel and compelling ideas about the mind, human freedom, and the place of mankind in nature, and of the irreducible but also non-dualistic mindedness of embodied creatures, whose mental properties are as basic in nature as biological properties, and metaphysically continuous with them. It therefore behoves us to give Kant his due and recognise his profound insights into the nature of human cognition.

1.1. Overview

This thesis represents a challenge to the orthodoxy of the dominant school of analytic philosophy which has coloured Kantian exegesis, as well as much of philosophy of mind, throughout most of the last century up until the present day. Although there is no one single defining feature of analytic philosophy, and it can perhaps be best understood as a tradition linked together by ties of influence and family resemblances (see Glock, 2008, p. 204), it commonly involves the application of certain logical techniques with the aim of attaining conceptual clarity. Thus, there are many variations of analytic philosophy, but as a general picture, all adherents to the tradition tend to hold that conceptual analysis is a central part, if not the only part, of philosophy. Initially, analytic philosophy involved a turn towards linguistic analysis as the subject matter of philosophy that embodied an accompanying methodological turn towards the clarification of the meaning of statements and concepts by breaking them down into parts so as to reveal sharply and perspicuously what can and cannot be said. Originating in Cambridge in the late 1890s as a revolt by Bertrand Russell and G.E. Moore against the neo-Hegelian absolute idealism that had dominated philosophy during the latter part of the nineteenth century, it is still regarded as the predominant philosophical tradition in the English-speaking world, and over the last two decades its influence has been steadily growing. As Michael Beaney, Professor of Philosophy at the University of York and editor of *The Oxford Handbook of the History of Analytic Philosophy*, writes in his introduction to the volume:

> Analytic philosophy is now generally seen as the dominant philosophical tradition in the English-speaking world, and has been so from at least the middle of the last century. Over the last two decades its influence has also been steadily growing in the non-English-speaking world (Beaney, 2013).

Analytic philosophy is also often combined with the belief that philosophy itself should be consistent with the successes of modern science, and is linked historically to logical positivism and logical empiricism and the idea that mathematical logic and observational evidence is indispensable for knowledge of the world. In early formulations the well-known "verification principle" determined that a proposition is meaningful only if there is a finite procedure for conclusively determining whether it

is true or false. It, therefore, originally rejected outright a transcendental realm of being or metaphysics, which was regarded as having led to much fruitless speculation, and was strongly committed to the view that any and all genuine questions of truth or existence belonged totally within the realm of science. However, although proponents originally viewed all metaphysical claims as meaningless, they later made efforts to re-construe them as significant assertions concerning language.

The year 1959 signalled a return to metaphysic when Peter Strawson, wrote *Individuals: An Essay in Descriptive Metaphysics.* Firmly rooted in the analytic tradition, he introduced his "descriptive metaphysics", which he regarded as fruitful in revealing the structure of our conceptual scheme, the fundamental aim of which was to clarify fundamental conceptual frameworks. (It was called "descriptive" as opposed to "revisionary metaphysics", the aim of the latter being to revise our ordinary way of thinking and our ordinary conceptual scheme in order to provide an intellectually and morally preferred picture of the world). [8] In his later seminal book, *The Bounds of Sense,* Strawson uses this to develop "austere" versions of six Kantian themes: objectivity, space, the unity of space and time, substance and causation (Strawson, 1966, p. 24) which he claims lie at the core of what he calls "Kantian descriptive metaphysics" as opposed to the "revisionary" metaphysics of his transcendental idealism, which he rejects. The *Critique of Pure Reason* was thus retranslated into an analysis of the concepts of the possibility of experience that hinges on the Transcendental Deduction, which for Strawson is an analytic argument aimed at proving the "objectivity thesis" that experience necessarily involves knowledge of objects.[9] Strawson's book became extremely influential and gave rise to many others based on the same interpretive model, thereby sparking a trend whereby philosophers came to regard the subjective aspect of the *Critique* as incoherent and as founded on a conceptual confusion, viz: that of *psychologism*. For Strawson, *The Bounds of Sense* presents a philosophical analysis of the *Critique* in which he roundly charges Kant with the excesses of "transcendental psychology". The claim was that it is possible to formulate a version of the argument of the Transcendental Deduction that abstracts entirely from the subjective aspect, and that considers only the objective aspect. In Strawson's view Kant's attempts to interpret the transcendental conditions of empirical knowledge as elements in our subjectivity

does not add anything to "descriptive metaphysics" and can be dismissed. "Psychologism" is a tricky concept to define, but at its simplest it is a form of philosophical fallacy that attempts to reduce diverse forms of knowledge, including concepts, principles of logic and mathematics, to psychology. It is a derogatory term, a kind of blanket condemnation capable of being used for significantly different types of argument. Kant's version of psychologism was deemed to be the illicit explanatory reduction of the necessary, *a priori*, and universal subject-matter of logic to the merely contingent, *a posteriori*, and relativised subject-matter of empirical psychology.

Thus it was that the analytic tradition, dominating as it did, much of Western academia, eventually led, through the publishing of powerful and influential studies of Kant, to the vilification and subsequent demise of his "transcendental psychology", the subjective part of the *Critique*. Although Kant's efforts to curtail speculative metaphysics and place rigid boundaries around the domain of philosophical inquiry were welcomed by the analytic movement, his apparent transgressions of those boundaries into psychology were not, and coloured Kantian exegesis, so that talk of the subjective side became philosophical taboo.[10] However, the foundational premise of this thesis; in fact, the basis on which it is written, is that this is a mistaken and narrow view. Whereas the transcendental aspect has, in general, been dismissed by philosophers of the analytic persuasion who have tended to reject such readings as unworthy of serious philosophical analysis, this thesis breaks with that tradition. Rather, it shares the viewpoint of Norman Kemp Smith, (who published, in 1929, the original standard English version of the text), and who complained "No interpretation which ignores or underestimates the psychological or subjective aspect of [Kant's] teaching can be admitted as adequate" (Kemp Smith, 1962, p. 51).

As mentioned in the introduction, Kant was particularly concerned with the need for acts of synthesis (or combination) and with the kind of unity required for such acts and his transcendental analysis of the prerequisites of our cognition gave rise to many original philosophical insights pertinent to several ongoing debates in contemporary philosophy of mind. Although Kant's transcendental psychology remains taboo for some mainstream contemporary analytic philosophers, it is contended that perhaps they need to get over this unreasonable prejudice. As Robert

Hanna so cogently reminds us "[t]he Kant we study nowadays is manifestly a Kant who has been reworked and represented to us by those who participated directly in the analytic tradition's long and winding struggle with the first *Critique"* (Hanna, 2004, p. 5), a tradition that "has now reached a stage of crisis", and is "speeding towards a crash", the origins of which can be traced back to analytic philosophy's rejection of Kant "via its intimate but stormy relationship with logical positivism" (ibid., p.11) [11] Analytic philosophy has not been without its detractors. Famously, Quine has argued that there is good reason to doubt that there is a special truth attached to analytic truth; in *Two Dogmas of Empiricism* (1951), he demonstrates, in two important theses, that there is no such thing. Steven Schiffer, in *Remnants of Meaning* (1987) also questions whether the semantic project that lies at the heart of contemporary analytic philosophy is not itself incoherent and impossible.[12] It is therefore argued that we should think beyond the rigid narrowness of technical proficiency of the analytic tradition in interpreting Kant, since there is a more profitable interpretation of the text, revealing invaluable insights that, when brought to the fore, can contribute much to current debates about the mind and cognition. It also has the potential to counter an increasingly emerging stark picture of humanity that threatens to undermine the sense of our own freedom and agency. The reductionist, deterministic, and mechanistic picture of human nature, in which a person's mental activities boil down to a multi-realisable functional system or as "entirely due to the behaviour of nerve cells, glial cells and the atoms, ions and molecules that make up and influence them" (Francis Crick, 1994), can be supplemented by a new, potentially liberating one, which provides room for freedom, morality and which reaffirms our understanding of what it is to be human - as individual, self-legislating or autonomous, intentional agents.

Although the original grand mission of analytic philosophy, that of laying out clearly and sharply what can and cannot be said, has faded over time, the basic idea of using precise language and logic to delineate philosophical problems remains. Pertinent to the topic of this thesis is that recent decades have seen the growth and flourishing of boldly *speculative* metaphysics within the analytic tradition and there is a strong trend today within philosophy of mind. Specifically, David Chalmers adopts the analytic framework using a form of modal logic borrowed from Saul Kripke, a sophisticated semantic theory in the philosophy of language, which he

regards as useful for clarifying the mind-body relationship, and uses it to embellish a Cartesian argument in an attempt to create something close to a proof of dualism. Chalmers' modal framework of primary/secondary intension is combined with a quite esoteric 2-dimensional semantic theory of possible worlds scenarios. There are necessary truths that apply to all possible worlds. However, his arguments share the tendency with the tradition of analytic philosophy of providing a rigid framework, fixing the meaning of words and providing precise definitions. The framework and the assumptions are set out from the start and given the premisses the conclusions automatically follow. In the tradition of analytic philosophy "intuitions" are taken as evidence for philosophical conclusions. All is carried out from the armchair, through gut intuition and *a priori* philosophising. At the other extreme, Daniel Dennett, also an analytic philosopher, is disparaging of "intuitions", denies the relevance of armchair theorising, speculative metaphysics, and rejects dualism, i.e. that there is something "extra" that needs explaining over and above function, and takes on a Humean position, as a type of scientifically-oriented empiricist, for whom the self does not, in some sense, "exist". For Dennett there is no "hard problem" of consciousness: the fact we think so is a consequence of misdirection by philosophers such as Chalmers and others as well as cognitive illusions created within our own brains; he claims the brain's computational circuitry fools us into thinking we know more than we do. Dennett's main idea is that computers can help realise one of the goals of analytic philosophy, namely, to produce unambiguous statements about behaviour, and dismisses reference to "mind" as meaningless. Anything meaningful we can say about mind or consciousness can be explained purely in computer functionalist or biological terms.[13] This has resulted in an ongoing and apparently tireless game of what Dennett has called "burden tennis" (Dennett, 1993) where the field of play is conceptual space and each side claims that the ball is in the other's court.

Other philosophers are also increasingly disparaging of analytic philosophy's obsession with intuitions, possible worlds, and two-dimensional semantic theory, which has been referred to, disparagingly, as "the industry of modal intuition mongering"[14] (Mandik and Weisberg, 2008, p. 20). Bruce Wilshire in a book on the subject declares: "Analytic philosophers are self assuredly smug in that they claim to know the real problems and the proper methods for investigating them, which

amounts to scientism, i.e. uncritical acceptance, amounting to worship, of the methods and outlooks of science". They "divide the emotive from the cognitive, and the moral from the factual," and it is their "endemic weakness" to make "overly simple and rigid distinctions" (Wilshire, 2002, p. 8).[15] Even Brian Lieter, long time defender of the analytic tradition writes in his Leiter Report: [16]

> Analytic philosophers generally become unbearably trite and superficial once they venture beyond the technical problems and methods to which their specialized training best suits them, and try to assume the mantle of "public intellectual" so often associated with figures on the Continent. The best analytic philosophers are usually very smart (clever, quick, analytically acute), but less often deep (Leiter, 2001).

Accordingly, Chapter 2 presents a sustained defense of Kant's transcendental psychology against its many detractors within the analytic tradition, and is concerned with rescuing it from the charge that it is unworthy of serious attention. Transcendental psychology is Kant's theory of mind, his critical analysis of the necessary *a priori* faculties required for cognition. Much of it is to be found in the so-called "subjective" part of the Transcendental Deduction, the second chapter of Transcendental Analytic section of the *Critique*, where Kant brings both the perceiver and subjective experience into accounts of the world. Kant's investigation of the psychological prerequisites of cognition, although deemed secondary to his main purpose, which was to justify our conviction that physics, like mathematics, is a body of necessary and universal truths (B19-21), was truly a remarkable feat in its own right. What is more, it has stood the test of time; it furnished him with several insights concerning cognition and mental functioning which were not only important in his own era, but which are still relevant today. He held that all contents of cognition are determined by the activities of a set of primitive, *a priori* and universal and *spontaneous* cognitive capacities, also known as "cognitive faculties" (*Erkenntnisvermögen*) which order experience. Kant was by no means an empirical psychologist; his cognitive theory was motivated by epistemological and metaphysical concerns. He distinguishes this from introspection-based empirical psychology and also from rational psychology (see Hatfield, 1992).[17] Nevertheless, his transcendental psychology is a kind of psychology, which, when rightly understood, can contribute much in terms of conceptual clarity and direction for

contemporary research in the cognitive sciences and the neurosciences, where some of the most puzzling philosophical problems have arisen and remain.

Moreover, considered as a general movement, analytic philosophy has had an uneasy relationship with historically oriented philosophy, and it has tended to proceed without recourse to the past. Philosophers of the analytic tradition have tended to think that conceptual analysis is a central part (if not the only part) of philosophy and whilst it may be true that it can do some good philosophical work, it should be realised that philosophy is not without a history; it is a historical movement as well as one concerned with more technical problems of logic and epistemology. Glock poses the question of what attitude philosophers should take towards the history of philosophy, the history of ideas and history in general, and his answer is what he calls "weak historicism" where the study of the past is seen as useful without being indispensable. Other philosophers argue that there should be a rather stronger form of historicism than the one he recommends (Williams, B. 2002, p.173; Alvarez, M. 2011, pp. 95-102)[18]. For Karl Ameriks, too, the question of the role of the past is no idle matter; but stands as the central problem that philosophy as a whole must answer (Ameriks, K. 2006). Philosophy is not simply a method of thought, but requires a sense of the distinctive issues that have developed over time and cannot be said to truly understand the problems it sets itself without an appreciation of the historical context in which these problems evolved.

Consequently, Chapter 3 turns to the problems in philosophy that were a legacy from Kant's predecessors, Rene Descartes and David Hume. The purpose of this analysis of the philosophy of Kant's forbears is to bring to light the particular problems that were bequeathed by them to philosophy, which Kant addresses through his "Copernican Revolution" (Bxvi), as there are parallels with the problems confronting him, the steps he took to resolve them and conceptual problems in the philosophy of cognitive science today. Some of these puzzles are, for example: How does a mind a *res cogitans* act on matter, a *res extensa*, or in contemporary terms, how do mental states causally interact with physical states of the brain? How do we account for the unity of consciousness; where is the "I" that Descartes says is so immediately known? What sense is there to freedom of the will in the face of a deterministic and mechanistic world? This is related to the notion of "causal closure in the brain" (Kim, 1993).[19] Thus, the analysis of the philosophy of Kant's forebears

is provided by way of a critique of the scientific materialism that lies at the foundation of cognitive science; since, it is suggested, the latter has its origins in a fundamental misconception or error of thought, a legacy from the former. This is the often implicit, unrecognised assumption of the Cartesian/Humean view of the mind in which consciousness is conceived of as a kind of entity (or *quasi* entity), which either "contains" experiences (Descartes) or which "consists" of them (Hume). It is proposed that although no theorist explicitly propounds this view, a residual alliance to this way of thinking about the mind instils confusion in many of the current approaches in the science of the mind. Adopting a historical perspective is particularly called for because contemporary problems in the field are the result of the lasting influence of this largely forgotten or neglected philosophical heritage. It is argued that Kant was the first to deal rigorously with this model or picture of the mind but his solutions have been broadly overlooked and this is a situation that should be rectified.

As discussed earlier, the main difficulty with functionalism in cognitive science is that it does not give sufficient justice to qualitative phenomena or *qualia* and omits several crucial elements that are necessary for cognition. This is because our perceptions and feelings have a qualitative character to them - there is something "it is like" to be in those states or, stated differently, they are *phenomenally conscious* to the subjects who undergo them. But this subjective realm is not accounted for in a functionalist/materialist explanation and is what ultimately leads to "the hard problem of consciousness" (Chalmers, 1995, 1996) since it opens up an "explanatory gap" (Levine, 1983), the problem of how "physical processes in the brain" give rise to subjective, phenomenal, experience; it involves the inner aspect of thought and perception: the way things feel for the subject. In his seminal paper "*Facing up to the Problem of Consciousness*" (1996), Chalmers resurrects this ancient puzzle of philosophical perplexity that was brought into sharp focus by Descartes. In Descartes' view the mental is absolutely distinct from physical processes; the two take place in different and distinct "substances". Today there is a new form of dualism, "property dualism" where the mind is viewed as existing in parallel with the body, but in a "dimension" that is separate from the material world. Chalmers has, in effect, merely substituted an explanatory dualism for Descartes' original substance dualism. Since in his view, a neuro-reductionist/functionalist explanation of mind is

rendered "scientific" at the cost of removing from it its most basic and fundamental characteristic: consciousness, or phenomenal, first-person experience, and this is what eventually leads to the view that consciousness is merely epiphenomenological, i.e. an accompanying event to cognition that lies outside the chain of physical causation and something which has to be *explained*. Chalmers writes:

> Experience is the most central and manifest aspect of our mental lives, and indeed is perhaps the key explanandum in the science of the mind. Because of this status as an explanandum, experience cannot be discarded like the vital spirit when a new theory comes along (Chalmers, 1995, p. 206).

It is contended that Chalmers, at least in some of his thinking, is the modern counterpart of Descartes, and espouses an updated form of dualism, namely "property dualism", the view that there are two metaphysically distinct kinds of properties in the world, mental and physical, whereas Daniel Dennett takes on the role of Hume, as a type of scientifically-oriented empiricist, for whom the self does not, in some sense, "exist". Since the cognitive revolution, philosophy of mind has become one of the main concerns of analytic philosophy. Both Dennett and Chalmers are analytic/ functionalist philosophers of mind. Dennett, however, denies there is a hard problem, and regards the self as nothing beyond the various "subagencies and processes" (Dennett, 1998, p. 105) in the nervous system that compose us and thinks of *qualia* as the part of an experience left over once all the objective parts are eliminated, and which are illusory. In particular, he denies there a medium of consciousness or as he likes to call it, "the *Me*dium", which is why *qualia*, "conceived of as states of this imaginary medium, do not exist" (Dennett, 2015, p. 2). Thus, the Cartesian and Humean approaches to cognition continue to hold sway today; functionalist approaches to cognition are, more or less, updated Hume, and Chalmers' "hard problem" is Cartesian; in fact it is a reworking of Descartes' argument for substance dualism in the *Sixth Meditation* (1641) where he notes that since he can "clearly and distinctly" conceive of himself existing apart from his body (and vice versa), and since the ability to clearly and distinctly conceive of things as existing apart guarantees that they are in fact distinct, he is *in fact* distinct from his body.

Sections 5 and 6 introduce the idea that Kant may have been the first to describe the peculiar logical semantics of self-reference, and how this philosophical analysis

of the nature of reference to self has profound contemporary relevance. For example, leading contemporary theories of consciousness, "Higher Order Theories" (HOTs) construe a mental state as self-aware, and hence conscious, in virtue of being an *object* of a numerically distinct second-order state. According to HOT theories, a phenomenally conscious experience of red would be based on the content of a mental state red with a direct relationship to the meta-thought red (Rosenthal,1997, 2005). That is, the higher level meta-thoughts or representations are distinct representations, the latter being phenomenally conscious in virtue of the former which represents it. This, however, does nothing to account for the distinctive feature of phenomenal awareness the "for-me-ness" of the experience. To be something it is like requires "for me" to "grasp" it in consciousness. For Kant, conscious awareness of an object has an implicitly dual nature, such that to be conscious of an object is also to be aware that one is in that very state. But this awareness is not, itself, a separable feature of the first order state. One has explained consciousness precisely when one has explained this dual feature. Such consciousness is more aptly described as a mode of being in a contentful state directed at the world, rather than it being so in virtue of a relationship to a numerically distinct second order state. Kant terms this "transcendental apperception" and it is linked to his idea of "transcendental designation" and the peculiar logical semantics of self-reference.

Chapter 4 presents an analysis of the extent to which Kant's transcendental psychology can be considered compatible with modern functionalist theories. As mentioned, several philosophers have brought attention to the incipient functionalism that can be discerned in the *Critique*, following the lead of Wilfrid Sellars in 1970. Ralf Meerbote, Patricia Kitcher, C. Thomas Powell, and Andrew Brook belong to this group. According to Brook, for example, the Paralogisms (crucial chapters in the Dialectic section of the *Critique*, where Kant criticises faulty arguments about the mind made by the rationalists (A298/B354)) [20] can be construed as one long argument that how the mind functions tells us nothing of its nature, which, he claims, is akin to the negative doctrine of contemporary cognitive science encapsulated in the dictum "function does not determine form" (Brook, 1994, 2004). He refers to Kant's famous tenet, "thoughts without content are empty, intuitions without concepts are blind" (A51/B76), which he says encapsulates the necessary

cognitive complementary and semantic interdependence of intuitions (percepts), which derive from experience and concepts, which come from the understanding. This relationship between percepts and concepts, he argues, has become as central within contemporary cognitive science as it was vital to Kant's purpose in the *Critique.* Patricia Kitcher also claims Kant is very much a forerunner of the functionalist program in cognitive science to describe the mechanisms that underlie cognition and states that "the easiest way to think about syntheses may be to regard them as processes that realize (mathematical) function. Given a set of input states a synthesis produces a certain output state" (Kitcher, 1993. p. 74). Thus, several contemporary scholars see Kant as giving abstract descriptions of cognitive mechanisms akin to functionalism in cognitive science. The general claim is that the unknowability of "things in themselves" which entails neutrality concerning the underlying composition of the mind means that he would have had to allow that multiple realisability is at least *open* to intellectual possibility.

However, Section 4.3 presents the argument that whereas Kant's *a priori* analysis of mentality bears a strong resemblance to functionalist theories, and can be considered an early form, it also diverges from and transcends it in several crucial respects. Not only does his analysis transcend functionalism, but valuable insights are *obscured* by such an interpretation. This is an important caveat against regarding Kant as a kind of "functionalist *avant le mot"* as Brook describes him (see Brook, 1994, p.13). Although Kant's theory is conformable to the idea that empirical awareness is workable under a functional descriptive system, his work concerning the mind has much more to offer, and reading him as such obscures the many positive contributions that can be made. For, as aforementioned, functionalism eventually gives rise to a seemingly unsurpassable problem which has become the focus of unending debate, the "hard problem of consciousness". According to David Chalmers, everything about human cognition apart from for the fact of *qualia* or subjective phenomenal experiences can be or will one day in the future be explained in reductive (computational or neural) terms, what he terms "the easy problems". But this leaves the problem of explaining subjective, phenomenal experiences themselves, the "what it's likeness" of experience. This is deemed by many to be the pre-eminent philosophical problem of today, the ignorance of which may be "the

largest outstanding obstacle [to] a scientific understanding of the universe" (Chalmers, 1996, p. xi).

Chapter 5 fleshes out the modern hard problem of consciousness and describes the historical and philosophical background from which it emerged. It further discusses how Chalmers reverts the philosophy of mind back to Cartesianism since he views the mind as neither physical nor material, but a fundamental "property", which gives rise to his "property dualism". This is because, according to Chalmers, there must be an "extra ingredient" in any explanation of phenomenal consciousness that goes beyond descriptions of functions and physical processes. Chalmers' analysis exemplifies a trend in philosophy of mind that puts phenomenal consciousness or "what it feels like to be a cognitive agent" at the centre of our understanding of mind. However, this presents us with the same problem as Cartesian dualism: the question of how non-material or non-physical mental properties interact with matter. Chalmers, as a functionalist, virtually equates the terms "functionalism" and "physicalism" in his ontology of mind. Everything apart from consciousness is physical, yet consciousness is taken to be a natural phenomenon, falling under natural laws. Thus, the hard problem that arose from functionalism is concerned with how processes in the brain are supposed to *give rise to* subjective experience, and is the view that consciousness is something extra which is somehow produced by neural states beyond the functional cognitive processes realised in the brain. This notion, that there is somehow an "explanatory gap" between the physical and the mental that needs to be closed, is the leading challenge to materialist views about the mind, and is an updated version of the Cartesian mind-brain problem, a resurrection of a conceptually flawed understanding. The central mystery about a hard problem of consciousness supposedly in need of explanation is also a reaction to and an artefact produced by adherence to the functionalist orthodoxy in which consciousness is reducible to or explicable by a set of functional cognitive processes realised in the brain. Such a view or picture *creates* an artificial explanatory gap between function and phenomenology in the first place. Kant's insights indicate that accepting that there really is a "hard problem" of consciousness, in the way that is stated, is a philosophical mistake.

Section 5.4 contrasts proponents of a hard problem of consciousness with the "no-self theorists", eliminative materialists, analytic reductionist/ functionalists for

whom the self/consciousness is an illusion. Daniel Dennett regards himself as the modern counterpart of Hume, as a type of scientifically-oriented empiricist, for whom the self does not, in some sense, "exist"; there are only various "subsystems" and "processes". Similarly, in his magnum opus *Being No One,* Thomas Metzinger's approach is based on a teleo-functionalist and naturalist view of consciousness. His thesis is that what we think of as the self is nothing beyond a special kind of dynamic representational content. He claims that "no such things exist in the world: nobody ever had or was a self " (Metzinger, 2003, p.1) - what we take as a self is no more than an appearance produced by the operations of a complicated information processing system that simulates and represents aspects of the system's states to itself. This follows from a specific neurobiologically grounded theory of consciousness, which Metzinger terms the Self-Model Theory of Subjectivity or SMT. The self and the world it perceives are illusions; the self is a "virtual self-model" (ibid., p. 544), and neuro-physiological processes are all that really exist. Metzinger claims that a disembodied but appropriately stimulated brain in a vat could, phenomenologically, enjoy exactly the same kind of conscious experience as an embodied one. The body can be separated off from brain processes which are metaphysically conceptualised as the minimal constitutive supervenience locii of experience, a set of minimally sufficient neural correlates. [21]

It is argued that Kant's depth of insight into the sources of our cognition can address the impasse between those who claim there is a "hard problem" to be addressed by a science of consciousness (Chalmers) and reductive or eliminative materialists who deny there is anything of the sort - consciousness is an "illusion" (Dennett, Metzinger). Although his theory can be construed as a set of abstract functional constraints valid to all cognisors, in keeping with functionalism, it can also be construed as a more "global" realisation in time and space of a complicated system of mental as well as *bodily* operations. Of particular significance in this regard, and something which cannot be emphasised enough, is that Kant had an ontological commitment that functionalism overlooks, in the sense that there is a positive metaphysical thesis deeply connected to Kant's theory of mind. For Kant, space and time are the "forms" of our intuition or sensibility and they are also constitutive of objects for us, the world of experience we inhabit must necessarily be a world of material objects. And as a member of that world, at least insofar as there

are bodies in that world that interact with and sense material objects, the "knowing self of apperception" must have a material body too. That is to say, the possibility of knowing objects in the empirical world necessitates that I have a material body in it.

Consequently, Chapter 6 presents an analysis of enactive, embodied, cognition which replaces the representational/functional model with agential activity and emphasises the role of the body and its place in the environment in creating cognition, arguing that even the most abstract of concepts are rooted in characteristics of our bodies and in our embodied interactions with the environment. Lived experience is at the heart of this enterprise, an experiential starting point that lies in stark contrast to traditional functionalist accounts that posit internal processes or representations as a starting point for explaining cognition and frame questions accordingly. As the originator of the theory stated: "Lived experience is where we start from and where we all must link back to, like a guiding thread" (Varela, 1996, 1999). Varela's conception of enactive embodied cognition is based on his work with biologist Maturana and the development of autopoiesis theory (Varela, Maturana and Uribe, 1974; Maturana and Varela, 1980, 1992). Within this quite recently emerged and developing embodied perspective, cognition appears as a dynamical process of real time variables with the capacity for self-organisation, not as a syntactic set of rules defining combinations of symbols that structure representational machinery, nor involving "neural representations in the brain" as on traditional accounts. Cognition is embodied when it is deeply dependent upon features of the physical body of an agent. Moreover, the mind is not "in the head" since its roots are in the body as a whole and also in the extended environment where the organism finds itself. This means that the constitution of a mind is always concurrent with the extended presence of other minds in a social network and the world. One of the key thoughts here is that if we want to understand consciousness it is not enough to simply look at the brain, and a set of functions or representational mental processes. We need to look to the embodied, situated life.

This goes beyond both functionalism and also the "consciousness in the brain" or "brain bound" paradigm (Thompson and Cosmelli, 2011). The biological substrate of consciousness is the whole organism in its dynamic interaction with the environment, not the brain taken in isolation from the bodily situatedness in which it finds itself. According to enactivism, we also need to take into consideration the fact

that living beings are autonomous agents that actively generate their own identities through *sense-making* rather than being the passive recipients of sensation. Moreover, as agents enact or engage with the world, they generate or bring forth their own cognitive domains. It is not enough to simply look at abstract information processing models of the mind, whether computational or neurological; rather, what is of importance to cognition is the embodied, situated life. The idea is that the situatedness in which a body finds itself is a necessary condition for cognition - the hypothesis that the biological machinery of consciousness will most likely turn out to be brain activity coupled to a body in interaction with its environment. Enactivism is pursued in two different styles: *sensorimotor* enactivism (Alva Noë et al) and autopoietic enactivism (Evan Thompson et al).[22] Both types take it to be a fundamental commitment that cognising agents are to be viewed as situated in an irreducibly meaningful world, where meaning is "enacted" or constituted through the tightly-coupled dynamic relationship between an agent's brain, body, and environment.

This chapter focuses on the recognition that a great deal is implicit in Kant's notion of ordering and unifying that is only recently in the process of being rediscovered in this newly emerging theory of embodiment. Kant's work, particularly his later work, goes beyond the abstract ordering and unifying of conceptual machinery that is put forward in the first *Critique* and considers the body. In his later unfinished project, the "Transition from the Metaphysical Foundations of Natural Sciences to Physics" to be found in the *Opus postumum*, published after his death, Kant theorises the dual emergence of natural mechanisms and organismic life (including mind) alike from a single ontologically neutral but dynamic material substrate, the dynamic aether, the so-called Aether Deduction (*OP* XXII: pp. 206-233, and 241). This all pervasive and self moving aether, which can be considered a *vis vivifica* or life-force is sufficient to explain the phenomenon of all organic life (ibid., p. 219). This consideration of organic life has been termed his organicism, which adds "meat" to the conclusions of the first *Critique*, with the thesis that the *transcendental* aspect of cognition is to be understood in terms of the model of biological epigenesis, a theory of biological formation, which lies at the heart of Kant's conception of reason. In fact, it has been convincingly argued that the idea of epigenesis guided and underpinned his critical philosophy, and in particular, the

relationship between reason and the categories of the understanding (Mensch, 2013).[23] It is the very fact of embodiment and orientation in the world that makes it possible to experience objects in the first place, and it is suggested that since embodiment can neither be reduced to the forms of intuition nor to a merely empirical fact, Kant must have conceived of the spatially oriented human body as a "transcendental ground for our cognition" in its own right. That is, it develops the thesis that for Kant, as with modern enactive theories, mind was explanatorily and ontologically continuous with life, in the sense that whatever is metaphysically required for a human mind is also ontologically present in its organic life.

Today, it is believed by many scientists that consciousness, including phenomenal consciousness itself, will sometime in the future, turn out to be completely explicable through the natural sciences, in terms of mechanisms, mathematics, computational logic and all that the scientific conception of the world includes, and that the task for a supposed "science of consciousness" is to try to explain the first-personal conception of the world completely in terms of it. It is also often simply taken for granted that it is only a matter of time before satisfactory machine models of human intelligence are produced. Cognitive scientists as well as philosophers refer confidently to "the mechanisms of cognition": the point can be brought home by the way that one leading eliminative materialist summarises this sentiment, stating that to abandon mechanism is tantamount to embracing magic, "since it cannot be magic, there must be mechanisms" (Churchland, 1986, p. 461). Kant's critical philosophy can contribute important and ingenious insights to the debate, one of the most significant being that mechanistic, computational, materialistic, natural science is *not*, to borrow Wilfrid Sellars' famous phrase, "the measure of all things". For if scientific or reductive naturalism is true, then indeed, in the words of Robert Hanna *"we are nothing but naturally mechanized puppets epiphenomenally dreaming that we are real persons*" (Hanna, 2006b, p. 436).

It is noteworthy that when we think of ourselves as embodied agents, we have a very different sense of "self-consciousness" than when we think of ourselves as simply introspectors. If we take this sense of bodily agency as the primary form of self-consciousness, then we are able to see self-consciousness not as a clear and isolated imaging or objectifying of a self, but as a lived sense of practical engagement. This entails that there is an intrinsic connection between the

cognitive/affective psychological life of the mind and the biological life of the body. Kant's thesis on embodiment has been termed "transcendental embodiment" which, for Angelina Nuzzo, is "the unifying thread of Kant's epistemology, moral philosophy, aesthetics and teleology of living nature" (see Nuzzo, 2008, pp. 8, 9). Other philosophers have expressed similar views (Steven Palmquist, 2013; Matthew Rukgaber, 2009; Susan Stuart, 2007, 2007a, 2008). Here the body is not simply the locus of the empirical senses, but rather the reference point of the formal sense for spatial orientation. This suggests that the body for Kant is the "transcendental ground" for our cognition, the locus or site for our "sensibility", or so it will be argued.

The next section provides some initial comments that will lead the reader into the main body of the thesis.

1.2. Kant and Cognitive Science – initial comments

Cognitive science is defined as "the interdisciplinary study of mind and intelligence embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology" (Thagard, 1996, 2014). The predominant approach in the field is that of functionalism, its central hypothesis being that thinking can best be understood in terms both of representational structures in the mind and of computational procedures that operate on those structures. The functionalist paradigm has been the prevailing model of the mind for around sixty years, its emergence coinciding with the rise of computing machines that were developed during the 1950s and 1960s, and which were inspired by Alan Turing's earlier work on machine tables. It was between 1960 and 1967 that Hilary Putnam developed his famous *theory of functionalism*, based on the idea that mental states are to be identified with respect to the causal or functional role they mediate between sensations and behaviour. As is well known, computational states are defined, not in terms of specific hardware configurations, but in terms of their relations to inputs, outputs, and their relationships to other computational states. Thus, in the case of functionalism of the mind, to talk of mentality is merely to talk of material systems

at a "higher level" (i.e. beyond biology or any kind of physicality); proponents hold that the way to model the mind is through its functions, what it does and can do. To quote arch-functionalist Daniel Dennett: "Functionalism is the idea that handsome is as handsome does, that matter matters only because of what matter can do" (Dennett, 2005, p. 17). This means that as far as functionalism is concerned the mind could be "copper, soul, or cheese" (Putnam, 1975); there is no need for any special mental stuff or underlying stratum (such as a brain) in order to account for particular psychological states. Another way of putting this is through the fundamental tenet of functionalism, "function does not determine form". Mental functioning can be realised, in principle, in objects of many different forms, what is termed "the multi-realisability thesis".

The multi-realisability thesis is the metaphysical claim that mental processes are the operations themselves and are not to be identified with anything material that realises them. Cognitive science has mushroomed rapidly on the basis of this paradigm, resulting in detailed theories of cognitive processes including perception, attention, memory, language and decision-making. The functionalist model is the prevailing model today within both cognitive science and cognitive neuroscience, where it has expanded beyond formalist cognitive psychology to include neural models. What underlies the technical jargon of both is that mental states are, in fact, or should in principle be, defined on the basis of causal connectedness and relation to other states, and to stimuli and responses between them. As was mentioned earlier, Kant's view of the mind can be regarded as compatible with contemporary functionalist theories in the sense that his views were also centred on how the mind works at an abstract level, rather than on how it might be physically/materially constituted or on its introspective contents. Kant was concerned with the "functions" or conditions needed for functions to work. Unlike his predecessors, he did not ask how we come to have knowledge of the world. He starts with the fact that we do have knowledge, i.e. have the ability to make conceptual judgements, and asks how this can be the case. He is uninterested in what knowledge we can derive from experience but in the *a priori* rules and conditions that govern our understanding. Using his method of transcendental argument or enquiry, it was his purpose to discover from our experience and judgements what the necessary features of these must be in all cases. He concludes several things, viz., that our knowledge is the

result of "acts of synthesis", that judgements have the particular content they do have in virtue not of their immediate causal relationships to objects, but only through their dependence on intuitions (A68/B93); that in order that representations of objects be anything to anyone they must belong with others "to one consciousness"(A116) i.e. they must be synthesised or combined with others into one unified representation, and that this, in turn, requires the application of concepts, what he terms "the categories of the understanding" which order and unify experience. His emphasis throughout is on the workings of the mind. Kant's functionalism is of a very general kind, however. Unlike contemporary cognitive scientists he was uninterested in specifics. Nevertheless, it is reasonable to state that this aspect of his theory can be viewed as an early form of functionalism, what we might term a *proto*-functionalism.

Kant's question was how is knowledge possible? Or to put it in the language of the *Critiqu*e: "What are the *a priori* conditions upon which the possibility of experience rests?" (A96). As stated earlier, several scholars have noted that Kant's views are compatible with contemporary functionalist theories, in the sense that they too allow consciousness to be characterised at a high level of abstraction that would allow instantiation into any number of physically realisable systems. Wilfrid Sellars,[24] oft credited with originating the theory of functionalism in the philosophy of mind, referred to Kant's "revolutionary move", which was "to see the categories as concepts of functional roles in mental activity", and claimed that for Kant "we do not know these [mental] processes *save as processes which embody these functions*" (Sellars, 1974, pp. 66-68). Similarly, Ralf Merboote claims that "Kant's transcendental psychology, often maligned, is a cognitive psychology. More specifically, it is a faculty psychology which speaks of capacities and abilities of various sorts which are needed for empirical cognition" (Meerbote, 1989, p. 161). In fact, a vocal minority of philosophers have been at the centre of attempts to salvage Kant's transcendental psychology from the criticisms of more positivist minded interpreters, and to revive it for its insights to cognitive science. Andrew Brook, in *Kant and the Mind*, advocates this construal and maintains that Kant's insistence on the unknowability of the mind implies a broad agreement with functionalism that: "(i) mental functioning could be realised in principle in objects of many different forms; and, (ii) we know too little about the form or structure of the mind at present

to say anything useful at this level in any case, except that mental functions will never be straight-forwardly mapped onto any forms that may be associated with them, whatever these forms might be like" (Brook, 1994, p. 13). He claims that Kant accepted a variant of both these positions and that he not only accepted the notion that function does not dictate form, but "accepted a very strong version" of it (Brook, 1994, p.14).

Thus, it has been recognised that Kant's insistence on the unknowability of the noumenal mind implies a broad agreement with the main functionalist claims. Functionalist readings of Kant emphasise the fact that he frequently treats concepts, both the *a priori* categories and ordinary empirical concepts, as functions that make it possible to transform the content of sensory experience into judgments. In the *Critique* Kant claims that we can know nothing about the "substrate" that underlies our mental functioning (A350). It follows that if the mind is unknowable, then it could take different forms. We cannot determine form from function; indeed, we cannot determine from function something even as basic as whether the mind is simple or complex (A353). As he writes:

> Through this I or he or it (the thing) which thinks, nothing further is represented than a transcendental subject of the thoughts = *X*. It is known only through the thoughts which are its predicates, and of it, apart from them, we cannot have any concept whatsoever, but can only revolve in a perpetual circle, since any judgment upon it has always already made use of its representation (A346/B404).

As with modern functionalist theories Kant's "transcendental" model can be viewed as centred on how the mind works at an abstract level, without reference to its material constituents. He organises the mind into certain "faculties", i.e. sensibility, (*Sinnlichkeit*), or the capacity for spatial and temporal representation (A22/B36), understanding, (*Verstand),* or the capacity for thinking or conceptualising (A51/B75), imagination (*Einbildungskraft*), or the capacity for the connecting of elements by forming an image (A120) and reason (*Vernunft*) or the capacity for logical inference and practical decision making.[25] Each of these faculties can be described, in functionalist terminology, as responsible for different phases of the constructive process that takes raw sensory experience as *input* and produces thoughts or judgments as *outputs*. A functionalist interpretation of Kant, therefore,

would focus on the theory of "synthesis" that contains Kant's division of mental labour into these more elementary tasks that are necessary for cognition, where he systematises the mind into functional "modes" responsible for different stages of the cognitive process that receive the disorganised influx of sensory experience as input, and which produce thoughts or "judgments" as output. Kant characterises synthesis as the activity by which the understanding runs through and "gathers together" the elements given by sense experience in order to form concepts, judgments, and ultimately, for any cognition to take place at all (A77-8/B102-3).

Interestingly, there are two broadly different kinds of synthesis in the *Critique*: the first is the synthesis of the various elements of experience, i.e. colours, edges, textures, shapes etc. into representations of single objects. The second kind of synthesis is the question of how the various represented objects must be bound together into a single representation of a world. This is akin to a key question in cognitive science of how the unity of perception is brought about by cognitive and neural mechanisms in the brain, the so-called "binding problem" (Treisman and Gelade, 1980). Discrete sensory features have to be bound together to form coherent perceptual objects, and these objects have to be bound to a common spatial framework so as to appear as parts of one globally unified perceptual world. Although there are only two broadly different *kinds* of binding, it is thought to occur at virtually all levels of perceptual processing, and thought, by some, to be a crucial event for unified phenomenal consciousness itself (Crick and Koch, 1994). In fact, this question, concerning perceptual binding and how it is linked to subjective phenomenal experience is regarded as one of the foremost problems in the scientific study of consciousness today: how does the human mind synthesises its modal and sub modal processes to generate a unity of conscious experience? (Zmigrod & Hommel, 2011, 2013). The binding problem is, at root, the problem of how the unity of conscious perception is brought about by the distributed activities of the central nervous system (Revonsuo & Newman, 1999).

At this stage it is necessary to explain how "cognitive neuroscience" is connected to "cognitive science". Neuroscience was first on the scene, the term "neuroscience" emerging as a label for the interdisciplinary study of nervous systems during the 1960s; the title "cognitive science" was not adopted until the mid 1970s, as the label for interdisciplinary studies of cognition. Until the 1980s, there was very

little interaction between the two disciplines. This lack of interaction was bolstered by the view of certain early functionalists, e.g. Hilary Putnam (1967) and Jerry Fodor (1974) that, since cognition could be multiply realised in many different neural as well as non-neural substrates, nothing essential to cognition could be learned by studying the brain, and all that really mattered was function. However, due to the development of new and much more powerful tools for studying brain activity, such as PET (positron emission tomography) and fMRI, (functional magnetic resonance imaging) this dogma was questioned and "cognitive neuroscience" was born. Harvard psychologist George Miller and neurobiologist Michael Gazzaniga coined the term "cognitive neuroscience" specifically for the study of *brain* implementation of cognitive functioning, the goal of which was now to address the biological foundations of human cognition.[26] Cognitive neuroscientists are able to study cognition in the brain by means of new imaging technology which enables them to see how behaviour and cognition, as studied by cognitive scientists, is expressed in functions in the brain, as studied by neuroscientists. What matters in terms of this thesis is that both are forms of functionalism, where cognition is necessarily reductionist; mind is akin to some form of computational mechanism, i.e. the mind is the "software" and the brain the "hardware" or "neural mechanisms". Cognitive science and cognitive neuroscience are guided by the paradigm of information-processing systems. In fact, computational cognitive neuroscience encourages reductionism by taking the mind to be software running on the brain. The object of study of cognitive science and cognitive neuroscience is the same, except that each focuses on different "levels" of explanation. A crucial component of this vision is that states of the system carry information about or "represent" aspects of the external world. The basic tenet is that all cognitive functions are at bottom a set of rules for handling symbolic entities that represent items of the world.

A consequence of computationalist/ functionalist, and also "neuro-reductionist" approaches is that they generally lead to a strange eliminativism, that is, to the paradoxical elimination of consciousness as the domain of subjective experience during the very process of explanation. From this point of view, the mind, although still thought of as "in the head", is reduced to representational mechanics, defined as a system of inputs and outputs. The problem that eventually arose from this was the question of how to account for consciousness, defined as subjective, first-person

phenomenal experience. As aforementioned, David Chalmers coined the phrase that describes it -"the hard problem of consciousness", and thanks mainly to him, the study of consciousness is now one of most vibrant fields in neuroscience, particularly in neurobiology. Frank Jackson (1982) and Joseph Levine (1983) had already claimed that there is an "explanatory gap" between the mental and the physical, and that there are reasons of principle why phenomenal consciousness cannot be reductively explained; a purely objective account of phenomenal conscious experience in terms of functions and processes cannot give an understanding of "what it is like" to have that experience. This is now regarded as the leading challenge to functionalist/materialist/ neurobiological views about the mind; an explanatory gap seems to open up since nothing in the functional or physical correlates of mental states explains why this state subjectively *feels* a certain way; it does not explain "what it's like" to have the experience. The claim is that all reductive strategies to explain how something feels a certain way seem to leave a gap in the explanation, in that, strictly speaking, such explanations cannot really be *understood.* This is an updating of the Cartesian notion that every property is either a mental kind or a physical kind, and where the extensions of those kinds are mutually exclusive. This resurrection of the Cartesian mind-brain problem, and the proposal of the existence of an unbridgeable gap between the physical world and the realm of phenomenal consciousness, is seen as a challenging argument against reductionist science. Three hundred years since the original Cartesian model initiated an earlier problematic dichotomy between mind and matter, Chalmers puts this ancient tangle of philosophical and scientific perplexity into sharp focus for modern researchers. In his seminal book, *Facing up to the Problem of Consciousness,* he frames the situation thus:

> I think it is widely acknowledged that consciousness is the biggest obstacle to a reductionist program in neuroscience, and I think most people in the field, both in science and in philosophy, agree that so far neuroscience has at most addressed the easy problems of consciousness and not the hard problem (Chalmers, 1996, p. xi).

"Easy problems" are so called because the specification of a mechanism that can perform the function is all that is required to solve it. That is, their proposed solutions, regardless of how complex or poorly understood they may be, can be

entirely consistent with the modern materialistic conception of natural phenomena. However, claims Chalmers, the "hard problem" is distinct from this, for even when the performance of all the relevant functions is explained, the problem of subjective phenomenal experience will still persist. This has become the most persistent philosophical question in cognitive science and cognitive neuroscience since the 1990's; whether and how subjective mental properties or *qualia* can be explained in terms of physical properties of the brain, or matter. This "hard problem of consciousness" which is linked to the "mind-body problem" is taken to be virtually *the* pre-eminent philosophical problem of today.

The principle tenet of a "science of consciousness" is that there is "something it is like" to have an experience, that consciousness is *essentially* characterised by reference to there being "something it is like" to be in certain mental state. To those who take there to be a real problem, everything about human cognition apart from for the fact of *qualia* can be or will one day in the future be explained in reductive (computational or neural) terms, the "easy" problems. But this leaves the *hard* problem of explaining the subjective phenomenal experience, or "what it's likeness" of experience. Others claim that function is all that is required to understand mind; anything meaningful we can say about mind or consciousness can be explained entirely in computer functionalist terms. In fact, every approach to the "problem of consciousness" has become a response, positive or negative, to the Cartesian distinction between body and mind. Today, centuries after Descartes philosophical dualism, the powerful and persuasive argument still continues to hold fast. Many theories of consciousness have emerged: with operations in the global workspace (Baars, 1988; 1996, Dehaene & Naccache, 2001), or with competition for action control (Shallice, 1988), or with informational content (Chalmers, 1996; Tye, 1995, 2005; Dretske, 1995; Tononi, 2004, 2008, 2012) or with higher-order thought (Armstrong, 1968; Rosenthal, 1997, 2005; Lycan, 1996; Carruthers, 2006; Byrne, 1997, 2004). Also, motivated by the notorious interpretation problems and the "measurement problem" in modern physics, there have been numerous attempts to modify or "complete" quantum mechanics through a theory of human consciousness by means of a combination with gravitation theory (Roger Penrose/Stuart Hameroff's Orch OR model, 1996, 2011, 2014), with quantum information theory

(Caves, Fuchs & Schack, 2007) or with psychology and brain science (Stapp, 2009, 2014).

It is proposed that paying attention to the insights of Kant would show that accepting that there actually is a problem of consciousness, in the way that is stated, is a philosophical mistake; it is not only insoluble as its stands, but that all efforts to solve it intrinsically risk both perpetuating its pre-eminence and further guaranteeing its insolubility. There is no solution to the hard problem of a consciousness, nor a way of getting rid of the problem, at least in not the way that it is currently presented. In fact, it guarantees, in a particularly powerful way, the perpetuation of the problem, of keeping it alive indefinitely. This is why the hard problem of consciousness, as it is presented in much of the literature, seems as intractable as ever.

Recent empirical work in cognitive neuroscience claims to be able to support the continuing endeavour to find consciousness in the brain through discovering the so-called "neural correlates of consciousness" or NCCs (Koch, 2004, Lamy, D., Salti, M., Bar-Haim, Y., 2009) i.e. the neural representational systems the activation of which is sufficient to bring about the occurrence of a specific conscious experience when certain specific, identifiable neural background conditions are in place.[27] The first main task for the neuroscience of consciousness, according to proponents of this view, is to find these neural correlates, specifically the minimal neural correlates for the phenomenal contents of consciousness. It had been discovered, in some early work in neuroscience on epileptic patients, that direct stimulation to the cortex in conscious subjects had certain highly individualised effects; it had been shown that stimulation of certain areas bring about experiences with a very particular phenomenology (Penfield, 1954, 1958). The fundamental idea of finding the NCCs is that this can be applied more generally; and once more is known about the brain, scientists will be able to discover precisely how consciousness comes about; i.e. there will be an isomorphism between phenomenal experience and these NCCs. This endeavour to find the neural correlate of consciousness has become a sustained, intense focal point for scientific research on consciousness. In fact, Crick and Koch, on presenting their conception of an NCC in 1998, stated that "[w]henever some information is represented in the NCC it is represented in consciousness" (Crick and

Koch, 1998, p. 98). That is, the main purpose of a science of consciousness is to uncover the neural representational systems whose contents systematically match the contents of consciousness. The aim is that this will eventually lead to a scientific theory of that will explain how consciousness relates to the brain.

This is an area of intense philosophical disputation with deeply polarised opinions, and it shows little sign of abating - the problem of consciousness is firmly entrenched in the scientific research arena as the vast amount of literature on the topic attests to, with others claiming that function is all that is required to understand mind, and that we can dispense with consciousness altogether, that anything meaningful that can be said about mind or consciousness can be explained purely in computer functionalist terms. It is suggested that this understanding of cognition has been heavily and sometimes unwittingly influenced by the Cartesian intuitions about inner and outer, the ontological divide between the mental and physical, which is reinforced by ordinary, everyday mentalistic discourse. It is also a reaction to, and an artefact produced by, adherence to the functionalist orthodoxy in which consciousness is reducible to or explicable by a set of functional cognitive processes realised in the brain. On Kant's view, mistaken beliefs about the mind and consciousness arise from reification of first person phenomenal experience because certain philosophers (Descartes) have projected the particularity of private, first person experience onto a third person entity called "the mind" or "thinking substance". [28] Chalmers reinvents that dichotomy by claiming that science needs to give an account of how certain neural processes *give rise* to qualia described as something extra on top of brain functionality. This picture reasserts the Cartesian dichotomy between the mental and the physical/functional domain - consciousness is still some sort of non-extended, non-spatial property that eludes current scientific explanation. This dualism is denied by others who take the Humean position that there is nothing to explain over and above function, and has resulted in the endless intellectual parlour game that Dennett terms "burden tennis" where the field of play is conceptual space and each side claims that the ball is in the other's court. In fact, the consciousness debates have provoked more angst than most as opposing sides tend to find the others positions not simply wrong but manifestly preposterous. Due to its perceived intractability, others, for example, Colin McGinn, have claimed that

we are “cognitivily closed” or constitutionally incapable of ever solving the hard problem (McGinn, 1989).

Recently, a novel view of the mind has emerged as an alternative to representational theories of cognition, that of embodied or enactive cognition, first proposed most explicitly by Varela, Thompson and Rosch (1991). This is a gradual move away from the still widely-held functionalist approach to cognition that presents the manipulation of inner symbols or mental representations as the mark of the mental. From this perspective the mind appears as a dynamical process, and not a merely a syntactic one; a dynamic of real time variables with the capacity for self-organisation. As noted above, cognitive science encourages reductionism by taking the mind to be software running on the brain, but according to embodied cognitive science the object of study is not mental representation, but the non-linearly coupled body-environment system. Most importantly, the mind is not “in the head” (Alva (Noë, 2009) since its roots are in the body as a whole and also in the extended environment in which the organism finds itself. To the extent that embodied cognitive scientists do study brains, they study them only as parts of behaving animals in information-rich environments. This means that the constitution of a mind is always concurrent with the extended presence of the environment and of other beings in the world. This idea has strong roots in Kant. In fact, what is present in Kant, also finds a convergent development from current philosophy of biology and the scientific notion of autopoiesis. On this view, instead of either trying to solve the hard problem of consciousness, (Cartesian) or describing consciousness as an illusion, (Humean) as is commonplace in reductionist, functionalist accounts, a new understanding of a form of immanent teleological “presence” involving truly biological features is now on the horizon, inevitably intertwined with the self-establishment of a conscious identity which is at the same time the living process, a “teleological circle”.

It is argued that Kant was committed to an active, sensorimotor view of consciousness, realising the spontaneity of rational agent through acts of “synthesis” or binding in an embodied biological system. His views, to be found in his later work, the *Critique of the Power of Judgement*, depict the human mind as metaphysically continuous with biological life, and display his recognition that rational human agents are necessarily also rational human living organisms, i.e.

biological animals capable of intentionality whose rational mindedness and rational directedness towards objects in the world, other real persons, and themselves, is fully continuous with this (see Hanna R and Maiese M, 2009). Moreover, in his last philosophical writings, the *Opus Postumum*, Kant refers to the work undertaken in the *Critique of Pure Reason*. Without invalidating the *a priori* categories that had been the possibility of all knowledge, he finds an entirely new foundation for them, the lived body. The moving forces of matter, the prime subject of natural science, are not deduced from the *a priori* categories of reason but themselves are a basic *experience* underlying all *a priori* categories.

In the *Critique of Judgement* Kant classifies various biological theories on offer and endorses epigenesis (*CPJ* 5: 422- 424); he also extends this theory analogically to his theory of cognitive innateness and the "epigenesis of pure reason" (B167). In fact, the concept of epigenesis lies at the heart of his conception of reason. Jennifer Mensch (2013) has written extensively about this, and the use of the biological term in Kant's work, drawing on a substantial body of philosophical and scientific works, including his published writings, correspondence and Nachlass, to substantiate her claim that Kant's engagement with the life sciences shaped his philosophical development even prior to the publication of the *Critique of Pure Reason*. She notes that even before his critical period, in "The Only Possible Argument in Support of a Demonstration of the Existence of God," 1763, Kant explicitly rejects the preformationist conception of biological generation and embryogenesis, according to which biological individuals and their complex structures are "pre-formed" (and require only the mechanical addition of bulk through nutrition in order to develop) in favour of the epigenetic view, whereby the basic forms or structures of creatures themselves are emergently generated by the spontaneous operations of a "generative force, and two kinds of internal constraints on the generative force, "germs" and "dispositions" which determine the outcome of the developmental process. Furthermore, she points out that in 1771, at the very beginning of his Critical period, Kant wrote: "Crusius explains the real principle of reason on the basis of [preformationism], *Locke* on the basis of [physical influx] like [*Aristotle*]; *Plato* and *Malebranche*, from [intellectual intuition]; we, on the basis of *epigenesis* from the use of the natural laws of reason" (Immanuel Kant, *Kants gesammelte Schriften*, 29 vols. (Berlin: Walter De Gruyter, 1902–), vol. 17, p. 492; quoted in Mensch, *Kant's*

Organicism, p. 83). According to Mensch "[t]his (...) allowed Kant to think of reason as a creature of its own making, as something self-born yet containing germs and predispositions for the possibility of its completion within an organic system that had been generated by itself. This model of epigenesis "allowed the openness of reason's possibilities to be maintained" (Mensch, 2013, p. 153). [29]

Thus, although Kant is often conceived as having offered little attention to the fact that we experience the world in and through our bodies, there is evidence that not only does he, throughout his career and in works published before and after the Critiques, reflect constantly upon the fact that human life is embodied, but the *Critique of Pure Reason* itself may be read as a critical reflection aimed at exploring some significant philosophical implications of this fact. In his later work Kant extends such notion to non-human animals, i.e. living beings in general. In the spirit of the modern perspective of enactivism, his thesis is that each organism brings forth his own embodiment to face the world, thus creating its own *umwelt* or consciousness. This is a further example of how Kant's insights continue to be of relevance in current debates in cognitive science and neuroscience about the mind and its place in nature.

The following chapter begins the work of defending Kant's transcendental psychology against a purely analytic reading. This is necessary because the foundational premise of this thesis is that a purely analytic interpretation is mistaken and narrow; and that there is much need for an understanding which fully embraces the "subjective" aspect of the *Critique*. Taking seriously the transcendental aspect is truly warranted because it can be made fruitful in addressing several conceptual puzzles and problematic ideas which have arisen in the cognitive sciences due to the subliminal grip of Cartesian and Humean assumptions about mental life in general which continue to hold sway today. This is related to the fact that analytic philosophy has moved on from its original concerns in the philosophy of logic and mathematics, and currently "possible world semantics" is regarded as useful in addressing the mind-body problem. As stated earlier, a 2D possible worlds semantic framework has been used to promote a type of conceivability argument in connection with metaphysical possibility in relation to this (Chalmers, 1996). The core notion is that for phenomenal concepts, conceivability implies possibility, and that whatever is logically possible is also metaphysically possible. The argument is also based on the

Cartesian-like "intuition" of distinctness between consciousness or *qualia* and physicalism, and on inference from an epistemic gap between them to an ontological one, that there really *is* a gap, and not just a gap in our understanding. A "hard problem of consciousness" thus arises because consciousness or *qualia* cannot be given a functional analysis, but is conceived as something "extra" which accompanies or is produced by neural states, and which lies beyond any functional organisation of the brain. This "conceivability argument" as it is termed is, in truth, a rehashing of the Cartesian real distinction between mind and body in the *Sixth Meditation* retrofitted for compatibility with contemporary modal logic by Chalmers (and indeed Kripke before him [30]) and amounts to new form of dualism (property dualism) which substantialises or objectivises mind as well as matter. This is the kind of problematic understanding of mind that Kant addresses in his analysis of the philosophy of his forebears. In light of this, the following chapter begins the defense of Kant's transcendental psychology against its many detractors within the analytic tradition, and is concerned with rescuing it from the charge that it is unworthy of serious attention. Transcendental psychology is Kant's theory of mind, his critical analysis of the necessary *a priori* faculties required for cognition that, when rightly understood, can assist considerably in providing conceptual clarity and direction for contemporary research in the cognitive sciences and the neurosciences, where some of the most puzzling, and ancient, philosophical problems have resurfaced and remain.

2. In Defence of Kant's Transcendental Psychology.

Kant's *Critique of Pure Reason* is universally regarded as a watershed in the history of philosophy, since it radically transformed the nature of Western thought. One of the main reasons for this was that it changed the philosophical conception of a human being from that of a passive *spectator* of the natural world, who, in a sense, exists apart from it, to that of an autonomous *agent* who has a necessary part to play in the construction of that very world. In this first *Critique* Kant sought to explain the objectivity of our understanding by describing the operation of certain human cognitive activities; that is, he sought to explain objectivity by appealing to features of the mind. Philosophers of the dominant positivist analytic school of philosophy, however, claimed that it is possible to formulate a version of the argument that abstracts entirely from the subjective aspect, and considered only the objective aspect. Analytic philosophy split from Kantian thought and from continental philosophy through the logical positivists' association with Frege, who was instrumental in putting a rigorous logic at the heart of philosophy. He was influential in the philosophy of mathematics, logic and language and held two fundamental tenets: a) that the basis for mathematics could be securely derived from logic and b) that a rigorous analysis of the underlying logic of sentences or propositions would enable us to judge their truth-value. This resulted in a turn towards language as the subject matter of philosophy that involves an accompanying methodological turn towards linguistic analysis. This manner of doing philosophy has dominated academic departments in various regions, most notably Great Britain and the United States, since the early twentieth century.

Initially, the logical empiricists of the Vienna Circle strongly rejected the aspirations of metaphysics, and more or less restricted philosophy to what they called "the logic of scientific language", using the principle of verification as the key to the notion of linguistic meaning and invoking verifiability as a criterion of meaningfulness. In the United Kingdom, where Logical Positivism was the product of the analytic tradition of the Vienna Circle, Russell, and A.J. Ayer constructed various theories of knowledge and methods of logical analysis aimed at making philosophy purely scientific, using the tools of scientific testing and procedure to

avoid the unprofitable web of speculative metaphysics. As a result of the influence of the new trend, logic and psychology went their separate ways; the mind, which was rightly considered central to Kant's account of objectivity, was expunged from it. The mind simply had no place in philosophical accounts of the objectivity of knowledge. On their positivist list of priorities was the elimination of traditional metaphysics and the main complaint of analytic philosophers concerning the *Critique* was that Kant's transcendental psychology was incoherent and founded on a conceptual confusion, that of *psychologism*, i.e. the illicit explanatory reduction of logic to empirical psychology. Ever since the onset of analytic philosophy, "psychological" readings of the *Critique* have been discouraged; those who have acknowledged intellectually the psychological import of the work have been charged with lapsing from philosophical sanity. This being so, the Transcendental Deduction, the second chapter of Transcendental Analytic section of the *Critique*, where Kant investigates the psychological prerequisites of cognition, was retranslated by various analytic philosophers into different kinds of analytic argument. Although differing in detail, what united them was that each of them sought to prove the anti-sceptical conclusion that empirical objects have a necessary categorical structure, whilst simultaneously avoiding any unnecessary and "irrelevant" reference to the subjective or to psychology. For example, Peter Strawson's famous Objectivity Thesis states that "for a series of diverse experiences to belong to a single consciousness it is necessary that they should be so connected as to constitute a temporally extended experience of a unified objective world" (Strawson, 1966, p. 97).

The purpose of this chapter is to show that this is a mistaken and narrow view. Indeed, contra the charge of irrelevance, questions concerning the status of Kant's transcendental psychological claims lie at the *heart* of the critical project. For the *Critique*, according to Kant, is *precisely* the abstract investigation of human cognitive capabilities in order to determine the scope of objective human knowledge. This was his "Copernican Revolution" (Bxvi) that turned the focus of philosophy away from metaphysical speculation on the nature of reality to a critical examination of the nature of the thinking and perceiving mind. An analytic interpretation that retranslates the work simply in terms of the objective conditions for the justification of empirical knowledge claims, (the objective side) is inadequate; there needs to be recognition of a further, equally important dimension, that of Kant's detailed

examination of the mental capacities that make cognition possible, (the subjective side). Kant's transcendental psychology, despite his claim that he deemed it secondary to his main concerns, which was to justify our conviction that physics, like mathematics, is a body of necessary and universal truths, is invaluable, not only in terms of the insights it can contribute to contemporary cognitive science but also within the discipline of philosophy itself. Kant was by no means an empirical psychologist; his cognitive theory was motivated by epistemological and metaphysical concerns. Nevertheless, his transcendental psychology is a kind of psychology, which, when properly understood, can provide insights and direction for contemporary research.

2.1. The Subjective and Objective Deductions

Analytic philosophy as an academic discipline or tradition finds its origins in the work of Gottlob Frege (1848-1925), Bertrand Russell (1872-1970), G. E. Moore (1873-1958), and Ludwig Wittgenstein (1889-1951) and has developed into a complex movement (or set of interconnected sub-traditions) that dominate academic philosophy today - although, as mentioned earlier, there is no single defining feature of analytic philosophy, and it can be perhaps best understood "as a tradition tied together both by ties of influence and by family resemblances" (Glock, 2008, p. 204).[31] During the 1960s a trend was initiated within this broadly defined analytic tradition that coloured Kantian exegesis, so that talk of the subjective side of the *Critique* became philosophically questionable. Kant had claimed there could be no perception of reality unmediated by human conceptualisation; that knowledge requires the application of categories which mould and shape experience into coherent form. All knowledge is necessary relative to a conceptual scheme, beyond which knowledge is impossible. Whereas Kant had located this in the mind, the positivists saw it as embodied in the language of logic or science. As already stated, the main complaint about the *Critique* is that Kant's transcendental psychology was founded on a conceptual confusion, that of *psychologism* which is related to his idea of the transcendental. Kant uses the label *transcendental* for "all knowledge which is occupied not so much with objects as with the mode of our knowledge of objects in so

far as this mode of knowledge is to be possible *a priori*" (B 25/A12). In the Transcendental Deduction he distinguishes between the "objective" and the "subjective" deductions (Axvii). The objective and subjective deductions present arguments about the *transcendental* structure of the representations of objects and the *transcendental* structure of a mind that has those representations, respectively. All commentators agree that the purpose of the *Critique* is to investigate the *a priori* conditions that are necessary for the possibility of knowledge, and also that this involves the conceptual and other cognitive conditions of having representations of object. However, analytic philosophers denied that this necessarily entails an analysis of what the mind, the "subjective sources" of understanding (A97), must, as a consequence, be like. It is suggested that this widespread contempt for this latter "psychological" or "subjective" aspect of the work, by philosophers of the dominant analytic traditional persuasion represents a blinkered view. As stated in the introduction, Norman Kemp Smith, the translator of the *Critique* into its original English Standard version, complained in 1918, "No interpretation which ignores or underestimates the psychological or subjective aspect of [Kant's] teaching can be admitted as adequate" (Kemp Smith,1962, p. 51). In fact, the Subjective Deduction is every bit as essential to the goal of the Deduction as the Objective Deduction. [32] Without the arguments put forward in the Subjective Deduction, the whole of the Deduction will not work. Kant was interested, in particular, with the abstract nature of mental processing, requiring the need for acts of synthesis and the kind of unity required for such acts, and interpreted properly, his work contains a wealth of insights that were not only important in his own time but which continue to have relevance today. The following section presents a short history of the debate.

2. 2. The History of the Debate

Even from a perfunctory reading of the *Critique*, one would be hard-pressed to deny its psychological content. Kant talks about psychological matters regularly, particularly in the Transcendental Aesthetic and the Transcendental Deduction, which contain many discussions concerning psychological processes and powers. He also discusses psychology in the Paralogisms (chapters where he discusses faulty arguments about the mind mounted by his predecessors, notably Descartes). Yet not

only in these parts of the *Critique* do we find psychological language. For Kant adopts an implicitly psychological vocabulary even in the most philosophical parts of his work. In fact, so pervasive are Kant's "psychological" discussions, that, as several recent Kant scholars have pointed out, if we fail consider this aspect of the work then there is not really much remaining to discuss. As one prominent Kant scholar has expressed it, "if interpreters (…) excise or ignore all the discussions of cognitive processes and powers then they will have very little left to read" (Kitcher, 1993. p. 4). Nevertheless, contempt for this psychological aspect of the *Critique* has resulted in most commentators either ignoring or dismissing such readings as beyond the pale of serious philosophical analysis. Moreover, this interpretative model has been so dominant that, as Kitcher notes, "even recent philosophy of mind has been anti-psychologistic" (ibid., p. 8.). This is not a new phenomenon: despite the prominent and extensive use of psychological language in the *Critique* there has, historically, been a misunderstanding of its role. In fact, from the time of Kant himself, the role of psychology in the *Critique* has been much discussed and opinions have varied. Some have considered that the book is primarily a work in psychology, although assessments about its precise nature and its propriety for Kant's purposes have differed. There are those that are happy to find a full-blown empirical psychology in the work (even though Kant did not recognise this). However, most have contended that the psychological content is a demerit, and evidence that Kant was in the grip of a deep-seated conceptual confusion. In fact, from Karl Leonard Reinhold, (1758-1823), who popularised his views, through until Henry Allison (Allison, 1996), and Paul Guyer (1987, 1989) the view has often been expressed that Kant's "psychological" subject matter is incoherent, and that he had erred in casting his arguments in this form. Scholars in general have tended to steer clear of this aspect of the *Critique*, regarding those who choose not to do so as having lapsed from philosophical sanity and as having committed the error of "psychologism". But what exactly is "psychologism"? Psychologism is a derogatory term, a kind of blanket condemnation capable of being used for significantly different types of argument. It is for this reason, not a simple concept to define. However, in its central usage it is the logical fallacy of confusing the normative with the factual or, to put it in more perspicuous terms, in Kant's case, it is the explanatory reduction of the necessary, *a priori*, and universal subject-matter of

logic to the contingent, *a posteriori*, and relativised subject-matter of empirical psychology. Kant, along with those who dare to take his psychology seriously, in being charged with "psychologism" are regarded as having been guilty of transgressing the proper bounds of philosophy.

Gottlob Frege and Edmund Husserl were jointly responsible for instigating this strong anti-psychological stance within philosophy, and initiated a trend which became a dominant influence on its subsequent development. In *The Foundations of Arithmetic*, Frege writes, as a first principle of his method, that is "crucial always to separate sharply the logical from the psychological, the subjective from the subjective sources of understanding" (Frege, G., 1884, p. xxii). Frege was concerned to provide a philosophical defence of the claims of mathematics to be a system of objective knowledge. He was against, on the one hand, mathematicians who treat knowledge as a matter of relating ideas, and on the other, those "conventionalists" that claim that they have the right to create new mathematical concepts. Frege's conception of mathematical knowledge is of an objectively existing reality which exists or, rather, "subsists" independently of psychology or human cognitive capacities. Language serves to represent this domain by means of sentences which say either truly or falsely how things are. The meanings of words cannot, therefore, be a function of how things are "in the mind", i.e. of ideas and thought processes, but rather, their meanings are grounded in the contribution they make to fixing the truth-conditions of those sentences in which they occur. In this way Frege turned the focus of philosophical attention away from "thought processes" to language and its function in the communication of knowledge. The language he is referring to is not actually existing natural language, but the ideal logical language which links human beings to the structures of "objective reality". Although he only attacks the use of psychological concepts in the philosophy of mathematics and logic, this conception of the principles to be used when conducting philosophical investigations became a dominant influence on the subsequent development of philosophy.[33] For his views influenced Russell, Wittgenstein and Carnap, the three founders of analytic philosophy, who were instrumental in establishing logic and Fregean philosophy of language as the paradigm of proper philosophy, thereby perpetuating this way of thinking. Through them the practice of psychologism became a serious philosophical

faux pas, one which drew scorn and pity, anyone making any significant association of logic with psychology regarded as almost beyond redemption.

A contributing factor leading to the demise of Kant's psychology was that interest in Frege was revived in 1950 when prominent philosopher J.L. Austin published an English translation of the *Die Grundlagen der Arithmetik*. This led to a new generation of philosophers who were influenced by Frege. Peter Strawson was one such scholar, and in *The Bounds of Sense*, in which he presents a philosophical analysis of the *Critique*, he roundly charges Kant with the excesses of "transcendental psychology". This book became extremely influential and gave rise to many others based on the same interpretive model. Although Strawson admits of the *Critique* that "the idiom of the book is throughout a psychological idiom" (Strawson, 1966 p. 19), he refuses to read the book psychologically and either ignores Kant's frequent references to psychological processes or dismisses them as irrelevant, reprimanding him for not practising serious philosophy, writing of Kant's transcendental psychology that "there is no doubt that this doctrine is incoherent in itself and masks rather than explains the real nature of Kant's enquiry" (ibid., p. 16). Moreover, he writes of synthesis that "the whole theory (...) like any essay in transcendental psychology, is exposed to the *ad hominem* objection that we can claim no empirical knowledge of its truth; for this would be to claim empirical knowledge of the occurrence of that which is held to be the antecedent condition of empirical knowledge" and that "the entire theory is best regarded as one of the aberrations into which Kant's explanatory model inevitably led him" (ibid., p. 32). The propositions of the doctrine of synthesis "belong neither to empirical psychology nor to an analytic philosophy of mind" (...) "they belong to the *imaginary subject of transcendental psychology*" (ibid., p. 97) [my italics]. Kant's attempt to interpret the transcendental conditions of empirical knowledge as elements of our subjectivity does not contribute anything to the project of descriptive metaphysics with which he is engaged. *The Bounds of Sense* spawned what has been termed "analytic Kantianism", a descriptive metaphysics that is taken to be a good model of what Kant was really up to, and which claims that therefore his philosophy should be reconstructed accordingly.

Thus it was that powerful intellectual forces both within and without Kantian scholarship resulted in the exclusion of Kant's transcendental psychology from

mainstream philosophical debate. Strawson had the strong conviction that the analytic approach in philosophy was the most fruitful one, and all of his work exemplified it. For him, the *Critique* is best understood as a purely analytic argument and he viewed his task as one of disengaging the analytic from the psychological side with a view to bypassing the psychological altogether. He supposes that what he is dismissing is mere idiom on Kant's part and that the true insights of the work lie in an "analytical argument which is in fact independent of [the doctrine of the faculties]" (ibid., p.16). [34] In his attempt to distance the apperception thesis from undesirable psychological connotations Strawson reconstructs it into the analytic argument that in order to have self-consciousness, subjects must be able to ascribe mental states to themselves, in what has become known as the "self-ascription thesis". He writes:

> Unity of consciousness to which a series of experiences belong implies...the possibility of self-ascription … [that is] the possibility of consciousness, on the part of the subject, of the numerical identity of that to which those different experiences are by him ascribed (ibid., p. 98).

Strawson's aim is to "disentangle" the logical or analytic from the psychological aspects of the *Critique* in order to present it as an "analytic argument" (ibid., p.16). As Patricia Kitcher (1993, 2006) has noted, he does not so much "disentangle", however, as raise psychological arguments only to dismiss them as useless. Examples of other philosophers with similar views abound. Jonathan Bennett, for instance, in *Kant's Analytic*, published at the same time as *The Bounds of Sense*, also shied away from a psychological reading of the *Critique*. Instead of charging Kant with "psychologism", however, he courteously attributes to Kant a Wittgensteinian view, rather than one concerned with "psychological" issues such as synthesis. Moreover, Henry Allison and Paul Guyer, although offering dramatically opposed interpretations of the *Critique*, at least agree on the issue of Kant's psychology, and that is, to avoid it at all costs. Other commentators also argue along similar lines. T.E.Wilkerson writes that "self-consciousness is more happily described as the ability to identify one's own experiences as one's own" (Wilkerson, 1976. p. 52). In fact, Wilkerson really combines a self-ascription reading with a specifically "logical" one. He writes "the unity of self-consciousness is (…) a *formal* unity consisting in the *formal* fact that experiences are mine" (Wilkerson, op. cit., p 52) [my italics]. In

these ways the doctrine of apperception is reconstructed in terms of the logic of self-ascription and "psychology" is avoided.

The transcendental unity of apperception poses a substantial problem for such interpretations, since it is the necessary attribute of a "thinker", and a "thinker" is a being capable of cognitive experience, who is self-conscious and who has a mental life. Nevertheless, so concerned are scholars of the analytic tradition to downplay the subjective side of the deduction that they have argued that it is either a non-psychological or an innocuously psychological claim about thinkers. However, in neglecting this aspect of the *Critique* and adhering to an analytic interpretation, they lose the richness of Kant's analyses into various aspects of human cognition. They also miss out on the opportunity to contribute to contemporary philosophical and scientific debate about the mind. During the latter part of the last century the face of psychology has changed. Behaviourism has declined. Cognitive psychology and cognitive science are now areas of rapid growth and exciting discoveries. It would be a shame if, through uncritical adherence to the analytic tradition, the opportunity to participate in philosophical discussions concerning these discoveries were lost. As mentioned earlier, Kant was by no means an empirical psychologist; his cognitive theory was motivated by epistemological and metaphysical concerns. Nevertheless it is a different *kind* of psychology and understood properly it can provide insights and direction for contemporary research.

Thus far this chapter has examined how certain interpretations of the *Critique* have, in the past, led to the charge of the fallacy of "psychologism" of which analytic philosophers have been wary. This is, in part, due to the Frege's anti-psychologism, which influenced the rest of philosophy. The core meaning of psychologism is that some important aspect of the realm of normative logic relies upon or is constituted by facts about human cognition; or, in simpler terms, it is the explanatory reduction of logic to empirical psychology. Kant scholars believed that the less they said about Kant's "transcendental psychology" the better. In referring to it they tended to highlight Kant's own negative comments about the value of psychology whilst ignoring the many references that are made about psychological processes.

Psychology may very well be inappropriate in logic; however, it is somewhat extreme to banish it from the rest of philosophy. It is true that Kant analyses the

"concept" of objective experience and the "logic" of various knowledge claims yet also refers to psychological features and processes. This does not mean that Kant's use of psychological language reveals a deep seated confusion on his part. His project was to determine how it is that we come to have certain types of knowledge or of how we come to reason and know. In so doing he naturally considers the cognitive apparatus that humans standardly possess. Kant did not feel the need that more recent scholars have to separate strictly "logically necessary" from "causal" or psychological conditions. Although there is a tendency among scholars to place studies under the heading of one discipline or other, there is no reason why they cannot belong to both. In other words, there is no need to be confined to "logically necessary conditions" or to "psychology". For there is an alternative possibility, which construes Kant as examining both the necessary conditions of the mind's operations and the actual psychology involved in them. This type of account neither limits itself to "logically necessary condition" nor commits the "fallacy" of psychologism. Although analytic commentators have taken Kant's notion of the transcendental or logical to exclude the empirical, so that if we take A as being logically necessary or a transcendent feature of B, A cannot be an empirical fact about B, Kant did not feel the need of those philosophers to separate the necessary from causal and other empirical conditions, for he felt that his enterprise concerned both. The purpose of the next section is to rethink some of the major areas of contention.

2.3. Rethinking the Transcendental Deduction

The central and most important section of the Critique is the Transcendental Deduction of the Pure Categories of Experience. The "transcendental unity of apperception" lies at the heart of the Deduction as Kant's principle doctrine concerning the necessary requirements for cognitive experience. As Kant defines it:

> [It] is that self-consciousness which, while generating the representation "I think" (a representation which must be capable of accompanying all other representations, and which in all consciousness is one and the same), cannot itself be accompanied by any further representation (B 132).

A transcendental deduction in general is a justification of the employment of *a priori* concepts that render experience intelligible for a subject. This being the case the Deduction would appear to be a doctrine the subject matter of which is concerned, in some sense, with human psychology. However, Kant was very clear that his philosophical aim was to be distinguished from that of empirical psychology. For example, he states that empirical psychology, what he termed "the physiology of inner sense" (A347/B119), a psychology based on introspective observation, could never gain "knowledge (...) of any object" (A381) i.e. determine any underlying structure in the flux of inner sense. He also declares, in a famous passage in the *Metaphysical Foundations of Natural Science*, that empirical psychology can never achieve the status of a science and should be banished from metaphysics (AAIV, p. 471). Yet, problematically, he also made extensive use of psychological vocabulary, even in the self-proclaimed philosophical parts of the book. This is what has led to the dispute among scholars as to the extent to which Kant used psychological grounds as a basis for his arguments.

That Kant made extensive use of psychological language is impossible to deny. Throughout the *Critique* he made frequent appeals to psychological processes, in particular in discussing his doctrine of synthesis which he writes about as if it were a causal process in the mind. As he puts it, synthesis requires that the manifold "be gone through in a certain way, taken up and connected". It is "the act of putting different representations together, and of grasping what is manifold in them in one [act of] knowledge" (A77/B103). He also appeals to introspection, often appearing to say that it is the having of direct knowledge of a self as the subject of the synthetic activities that underlies the unity of apperception. Although he is clear that the only *knowledge* we have of ourselves is as "appearance" through empirical apperception, we, nevertheless, *in* transcendental apperception, are able to have "consciousness" of ourselves as the locus the activity of synthesis. This he calls a "thought". As he writes:

> In the transcendental synthesis of the manifold of representations in general, and therefore in the synthetic original unity of apperception, I am conscious of myself, not as I appear to myself, nor as I am in myself, but only *that* I am. This *representation* is a *thought*, not an *intuition.* Now, in order to *know* ourselves, there is required in addition to the act of thought, which brings the manifold of

> every possible intuition to the unity of apperception, a determinate mode of intuition whereby this manifold is given…The *consciousness* of self is thus very far from being a knowledge of the self...(B157/8) [my italics].

Kant also refers to the "I think" as "the logical subject of thought" (A350), a "formal unity of consciousness" (A105) or a "formal proposition of apperception" (A354). Again, at A398 he writes that "the 'I' in 'I think' is only the formal condition, namely the logical unity of every thought". However, what he meant by "formal" or "logical" is not entirely perspicuous. The approaches of Strawson and other Kantian analytic philosophers make sense only on a sharp distinction between what Kant meant by "logical" and what we now mean by "psychological", a presupposition which does not bear much scrutiny. In the Transcendental Deduction Kant seeks to prove that the pure concepts of the understanding, the twelve categories, identified earlier in the Metaphysical Deduction, are objectively valid for empirical objects, and this is the problem of whether it is legitimate to use categories like *cause-effect* to think about empirical objects, and the solution Kant gives is that the categories are the very conditions of the possibility of experiencing objects at all. We cannot experience objects unless we experience them as falling under these *a priori* principles. He also discusses the necessary interdependence of object consciousness and the transcendental unity of apperception, which is the ultimate condition of the possibility of experience. There could be no experience of objects at all if the manifold of representations were not synthesised or combined by the mind through the unity of apperception (A108/135). This means that there can be no experience of objects without experiencing them as falling under the categories and no self-consciousness without experiencing objects as falling under the categories. The "transcendental unity of apperception" is the point where the self and the world come together.

The transcendental unity of apperception is, therefore, both the name for a faculty of synthesis and also the for what he referred to as the "I think", consciousness of oneself as subject. So, when Kant says that "I" of "I think" is the "logical subject of thought" (A350) or a "formal subject of thought" (A105) or a "formal proposition of apperception" (A354) what does he mean by this? As discussed above, Wilkerson explains it by saying that that the unity of consciousness is a "formal" unity "consisting simply in the formal fact that experiences are mine"

(Wilkerson, op. cit., p 52). However, it is not at all clear what he can mean by this. He cannot mean that it is a fact of formal logic that mental states belong to a thinking subject. For Kant logic investigates the rules governing understanding in general, whereas Transcendental Logic "concerns itself with the laws of understanding and reason solely insofar as they relate *a priori* to objects" (A57/B82). Kant held that logic is part of what we would nowadays call "psychology". The following passage from the *Anthropology* illustrates how Kant saw the contrasting domains of logic and psychology. It shows that we would nowadays include both his logic and his psychology as being in the domain of psychology.

> The lower cognitive power is characterised by the passivity of the inner sense of sensations; the higher by the spontaneity of apperception - that is of pure consciousness of the activity that constitutes thinking - and belongs to logic (a system of rules of the understanding) just as the former belongs to psychology (to a sum total of all inner perceptions under laws of nature) and establishes inner experience (An Ak VII, pp.140-41).

In other words, there is no sharp distinction between what we mean by psychology and what Kant means by logic. Logic, for Kant, is the abstract study of the mind. As Robert Pippin, who prefers a metaphysically neutral rather than a Strawsonian logical analysis, points out, the contrast between logical and psychological was hardly as clear-cut for Kant as it supposedly is for us, (Pippin, 1987, 2014, see also Hatfield. G, 1992). He did not understand logic as rules for well formed formulae and rules for truth preserving inferences, as in Porte Royal logic. For Kant logic sets out the rules that constitute thinking *as such*, and so its scope is far wider. Kant clearly holds that his "logic" of the mind is part of what we would now consider psychology - his theory of logic is an abstract theory of thinking. [35] There is no need for Strawsonian concerns about the naturalistic fallacy of confusing the normative with the factual, or the logical with the empirical. For Kant, to explore the "logic" of experience is simply to explore how the mind works. In fact, Kant went further than claiming that the abstract study of the mind is part of logic. He also thought that "general logic", which determines the basic structure of the way we think, in that it specifies the forms of judgement on which the categories are based, is part of the structure of the mind. In this case A being necessary for B is not incompatible with A being the cause of B. Traditional logic and "transcendental logic" are not to be

considered separate domains, but are better characterised in terms of the particular aspect of understanding at issue. Whilst traditional logic investigates the *form* of understanding, "transcendental logic" is concerned with *content*. Kant claims that the traditional logical forms and the transcendental logical categories completely coincide. He writes: "In the metaphysical deduction the a priori origin of the categories in general has been proved through their *complete agreement* with the general logical functions of thought" (B159) [my italics].

Kant deduces the conceptual structure of experience from traditional logic, which is based on the Aristotelian syllogistic system. In the Introduction section of the Transcendental Logic he writes that traditional logic treats thinking in complete abstraction from all content. He then takes the core of traditional logic as providing a "general logic" in that it concerns itself with just those general conditions that are necessary for any thinking to take place at all. General logic concerns these rules "without any regard to the difference in the objects to which the understanding may be directed" (A52/B76) and merely "treats of the *form* of thought in general" (A55/B79). In other words, thinking, being exclusively discursive, can provide itself with no content at all, which can be given only through sense experience, or "experience in general", i.e. the range of our experiences independently of particular content. Kant's transcendental proofs are then meant to show that experience would be impossible but for the *a priori* origins of certain features of cognition. These are not innate features but epistemological in nature. Kant writes in the Introduction: "In what follows (...) we shall understand by *a priori* knowledge, not knowledge independent of this or that experience, but knowledge absolutely independent of all experience" (B2-3). *A priori* knowledge is knowledge that is established independently of experience, but it is nevertheless tied to our cognitive capacities and is what is common to all experience. Thus, the generality of logic is a normative generality: logic is general in the sense that it provides constitutive norms for thought *as such* independently of subject matter. This is contrasted with "applied logic" which is "a representation of the understanding under "accidental, subjective conditions", i.e. empirical and subject to causal laws.

In order to further illustrate that what Kant means by logic is quite different to what it means in contemporary usage, it will be helpful to examine what he says in his lectures on Logic, *the Jäsche Logic*, since most of the subjects found in the

Critique can be traced back there, which is hardly surprising since Kant read it for forty years during which time he was also developing his system. The Transcendental Doctrine of Method, the second part of the *Critique* in which Kant explains the purpose of the work and provides the framework for understanding it, is of particular relevance here, for it can readily be seen to be a distillation of the Logic, which means that it is a fundamental and indispensable background to any proper understanding of the *Critique*, in the sense that it provides the original format of Kant's architectonic system of which the methodology of the *Critique* is a specific case. That is, it exemplifies the pure method of enquiry of which the *Critique* is merely (although this is surely not the right word!) an example. In the *Jäsche Logic* lectures Kant presents an introduction to logic which is founded on "the objective unity of the manifold in cognition", i.e. a logic based on transcendental apperception. Kant begins these lectures by declaring that "the understanding is bound in its acts to rules we can investigate" (*Jäsche Logic*, intro, p. 13). However, it is interesting to note that the topics of these lectures include consciousness, empirical concepts arising from sense experience, comparison and abstraction, which are in the domain of what we would consider psychology. It is important to realise that Kant, along with others of his time, did not regard these topics as merely providing psychological backing for the principles of formal logic. If we take Antoine Arnauld, for example, he describes logic as "reflecting on these natural operations" (i.e. conceiving, judging, reasoning, ordering) which helps us to reason better to correct defects in the mind's operation and claimed, further, that "we become better aware of the operations of the mind by reflecting on its operations" (Antoine Arnauld, trans.1964, pp. 29-30). Kant's meaning of "logic" is similar. He takes the understanding itself to be a logical or epistemic power. Therefore, when he came to write the *Critique* it is reasonable to presume that in his quest to determine what our cognitive make-up had to be like in order for us to be capable of knowing what we do, he again took the understanding to be a "logical" power and his aim was to study that power itself. The following remarks illustrate how confusion leading to different readings of the *Critique* might arise:

> As pure logic (...) has nothing to do with empirical principles, and does not, as has sometimes been supposed, borrow anything from psychology, which therefore has no influence whatever on the canon of the understanding. (...) What I call applied logic (…) is a representation of the understanding and of the rules

> of its necessary employment *in concreto*, that is under the accidental subjective conditions which may hinder or help its application and which are given only empirically. It treats of attention, its impediments and consequences, of the source of error, of the state of doubt, hesitation and conviction, etc. Pure general logic stands to it in the same relation as pure ethics, which contains only the necessary moral laws of a free will in general, stands to the doctrine of the virtues strictly so called - the doctrine which considers these laws under the limitations of the feelings, inclinations and passions to which men are more or less subject (A54-5 / B78-9).

According to the anti-psychologism lobby, Kant, in contrasting "applied logic" with "pure logic" is contrasting it with something merely "logical" or "conceptual" i.e. empty of empirical meaning, and nothing to do with real minds, i.e. as something analogous to the way that the "necessary moral laws" of "pure ethics" are non-empirical laws to be considered independently of particular human actions. However, for Kant, whereas "applied logic" is "a representation of the understanding under "accidental, subjective conditions", i.e. empirical and subject to causal laws, "pure logic" is a representation of the understanding as *abstracted* from the effect of these conditions. He is talking about the general features of a mind capable of cognitive experience. In this case the representation in question is not "non-empirical" and purely "formal" or "conceptual", but is in fact what is common to all experience, experience *in general*. This is also Kant's reasoning when he talks of the "moral laws of a free will in general". Again, he is not talking about non-empirical laws in the sense that we could not observe them in particular acts of virtue, for indeed they what is are common to all virtues. Although these general features are certainly more than merely empirical, being also necessary, and are also incapable of being established by empirical means, they can still be considered empirical in the sense specified. Thus, the initial remark "pure logic (…) has nothing to do with empirical principles and does not borrow anything from psychology" is not to be understood in the way that proponents of anti-psychologism understand it. It cannot be used to support analytic or "logical" interpretations. For it is, despite protests to the contrary, compatible with empiricism in the sense specified. For Kant this means that although pure logic cannot be shaped by experience, it still specifies the general structure of experience. Proponents of anti-psychologism would no doubt object that all this makes Kant's psychology too empirical to count as philosophy. However, as

was mentioned earlier, Kant, in the first *Critique,* is not interested in the actual physical embodiments of mental processes or faculties, but in the very general requirements of a mind capable of performing various cognitive tasks and in this his work is different from empirical psychology and is, in fact, very much centred in epistemology. That is, in investigating our cognitive faculties, the forms of sensitive intuition, the categories and transcendental synthesis, Kant is seeking conditions for knowledge the subject matter of which is epistemic as opposed to psychological. At any rate, there are other reasons for resisting an analytic interpretation of the *Critique*. As mentioned earlier, Kant did not feel the need to separate causal from necessary conditions, for he felt his enquiry concerned both.

Moreover, if we go one step further and cast doubt on whether there even is a special kind of truth attached to analytic truth, such interpretations lose their attraction. This century Quine has done just this. He has argued that there is good reason to doubt that there is a special truth attached to analytic truth. In fact, in *Two Dogmas of Empiricism* (1951) he demonstrates, in two important theses, that there is no such thing. If this position is granted then, far from an analytic interpretation rescuing Kant from the charge of psychologism, the opposite is true. Arguments of the Critique are diminished if they are cast in analytic form. More recently, Robert Hanna has also argued that, despite the fact that logical psychologism is false, there is an essential link between logic and psychology, in the sense that logic is intrinsically psychological and human psychology intrinsically logical, and defends the Kantian thesis that logic is intrinsically psychological in a way which does not make it vulnerable to the charge of psychologism. The reason for this is that logic is cognitively constituted by all rational animals, who possess an *a prior* cognitive "logic faculty" (Hanna, 2007). His logical cognitivism says both (i) that there is an essential connection between the logical and the psychological, and (ii) that logic is not explanatorily reducible to empirical psychology. The latter claim is an anti-psychologistic, and more generally an anti-naturalistic claim, in the manner of Quine. In other words, as Sher remarks, "logic is globally explained and globally justified by our cognitive constitution" (Sher, 2006).[36]

As mentioned previously, Hanna is sympathetic to Jennifer Mensch's Kantian "organicism" (Mensch, 2013) for whom "not only Kant's pure general and transcendental logic, but also the underlying conception of rational systematicity that

guides his dialectical logic, are all grounded on a deeper "organic logic" of teleological development and conceptual integration"(Hanna, 2014, p. 8). This hinges on the claim that the biological theory of epigenesis guided and underpinned Kant's critical philosophy, in particular, the relationship between reason and the categories of the understanding.[37] In *Kant's Organicism* Mensch writes that Kant "was prepared to borrow freely from the models and vocabulary of the embryological debates underway. Indeed (...) it was these models that would eventually help Kant discover the origin of knowledge itself" (Mensch, 2013. p. 53). She argues that the significance of epigenesis for Kant goes beyond the claim in the Transcendental Deduction that reason is self-generating, and the categories self-thought, and that epigenesis implies a "transcendental affinity" within cognition itself "an affinity which grants unity to "the experience of nature's coherence" (ibid., p. 134). Hanna describes it thus: "According to this "organic logic," holistic purposive schemes in nature, theory, and morality are all immediately grasped by what Kant in the third *Critique* calls an "intuitive understanding" " (op. cit., pp. 8-9) and what this means, in the words of Thomas Nagel, is that "rational intelligibility is at the root of the natural order" (op.cit., p. 11) This issue will be explored in Chapter 6.

2.4. How Kant Does not Undercut his own Psychological Claims.

Scholars of the analytic philosophical persuasion, who dismiss "psychological" readings of the *Critique* support their arguments by emphasising the fact that Kant himself had reservations about psychology and that consequently any material which is "psychological' in character should be dismissed as irrelevant or else translated into a non-psychological or innocuously psychological claim about thinkers. They point out that Kant dedicates a whole chapter to defeating the pretensions of the rational psychologists and also to declaring that empirical psychology can never attain to the level of a proper natural science, for its domain is introspective observation, the contents of "inner sense" and is therefore not quantitative, or at least the contents of inner sense can be quantified only in one dimension, which means that no informative mathematical model of them is possible. (For Kant science

"proper" must be based on the model of mathematics, since this is the only method in which it is possible for a science to achieve the necessity and certainty required). Moreover, they claim, he seems to undercut his own psychological claims. For instance, he says of the subjective aspect of the Transcendental Deduction, what he terms "the investigation of the understanding itself", that it is "hypothetical", i.e. inessential to his main goal (Axvii), although it is omitted that Kant also mention that it is of "great importance". They also allude to the fact that he reminds logicians, seemingly in anticipation of Frege himself, that "pure logic has nothing to do with empirical principles, and does not, as has sometimes been supposed, borrow anything from psychology…" (A54/B78). In addition to this they point out that Kant explicitly contrasts his doctrine with the psychology of Locke (A86/B118-19), preferring to call his area of study "transcendental logic". They claim that although the *Critique* contains many references to psychological features and processes it also contains several criticisms of psychology and negative comments about its significance. These are emphasised in order to back up the claim that Kant's transcendental psychology should be disregarded.

The following is the passage in the Preface to the First Edition where Kant points out that his quest for the limits of human reason has both an "objective" and "subjective" aspect. The objective side...

> ...refers to the objects of pure understanding, and is intended to expound and render intelligible the objective validity of its *a priori* concepts. It is therefore essential to my purposes. The other seeks to investigate the pure understanding itself, its possibility and the cognitive faculties upon which it rests; and so deals with it in its subjective aspect. Although this latter exposition is of great importance for my chief purpose, it does not form an essential part of it. For the chief question is always simply this: - what and how much can the understanding and reason know apart from experience? not: how is the faculty of thought itself possible? (Axvi – xvii).

This is an oft-quoted passage from Kant where he is at pains to point out that the emphasis of his investigation is not psychological but epistemic in that it seeks to determine the objective conditions and constrains on knowledge. The "subjective" side is of secondary importance to this goal. This declaration has been regarded by many Kant scholars as a retractory statement, and as evidence that Kant himself saw, at least vaguely, that he had made a mistake in casting his arguments in

psychological form. This is then regarded as justifying the view that Kant's discussions of psychology are not worthy of serious attention at all. After all, does not Kant say, in the very next sentence, that the subjective side of the deduction is "somewhat hypothetical in character"? They also point to the fact that Kant himself had clear reservations about psychology. For example, he states that empirical psychology, what he termed "the physiology of inner sense" (A347/B119), a psychology based on introspective observation, gains knowledge "not of any object", i.e. can never determine any underlying structure in the flux of inner sense (A381). He also declares, in a famous passage in the *Metaphysical Foundations of Natural Science*, that it can never achieve the status of a science and should be banished from metaphysics (AAIV pp.471). Furthermore, he devotes a whole chapter, the Paralogisms of Pure Reason, to deflating rigorously the pretensions of rational psychology. These negative comments are emphasised in order to plead the case for the impropriety of psychology for Kant's purposes. The claim is that his reservations about it indicate that he himself came to doubt his own.

At first sight there does seem to be ambivalence and confusion, even on Kant's own part, about the status of his cognitive claims. For on the one hand, his arguments undeniably contain cognitive subject matter, yet on the other he explicitly denies the relevance of psychology for his purposes, thus seeming to undercut own psychological claims. However, the crucial point of the matter is that Kant's use of psychological concepts is different from that of empirical and rational psychology. The *Critique* contains its own psychology divorced from empirical and rational psychology. To repeat, the cognitive subject matter Kant refers to is not strictly "psychological". He merely borrows modes of explanation from psychology in order to construct his explanations. He does not appeal to psychological argumentation in order to do so. Kant's arguments are neither directed at the phenomena of "inner sense", as are those of empiricist psychology, nor the "soul" as "substance", as are those of rationalist psychology. However, they still constitute a psychology of sorts. His arguments are "transcendental". Transcendental psychology does not belong in the domain of psychology proper, neither rational nor empirical. In fact, one of Kant's main aims in the Paralogisms chapter is to distinguish his own transcendental enquiry from that of rational psychology. For, as he sees it the claims of rational psychology are transcendent, i.e. beyond what possible experience could

verify. Therefore they are unwarranted. Transcendent is the term used by Kant to describe those principles which "profess to pass beyond" the limits of experience as opposed to those principles "whose application is confined entirely within the limits of possible experience" (A296/B352). (Such claims are also founded on faulty syllogistic reasoning and are fallacious, as will be discussed in Chapter 3.2.). He is also adamant that empirical psychology is irrelevant to his own project. He sets his project apart from it at the beginning of the Deduction where he insists that accounts of the empirical origins of concepts and beliefs are without philosophical interest. Empirical psychology, for Kant, is incapable of serving the purposes of the Deduction, since it is of no use in dealing with philosophical questions about the "validity" of cognitive claims. Kant used an analogy from law to describe the project that lay before him. In a court of law, a distinction is made between "questions of fact" and "questions of right". For questions of fact, the court appeals to testimony and material evidence, both of which are appeals to experience; with regards to questions of right, however, the court appeals to rules of law. Establishing, on the basis of rules, whether an action is right is called, by Kant, a "deduction." Thus the Deduction addresses the questions of "*quid juris"* the "question of right" as opposed to the "*quid facti"*, "question of fact". Using the legal terminology of his day, Kant asks by what right we apply the categories in an *a priori* manner, i.e. he asks a justificatory question about the right to apply the categories independently of experience. He is not concerned with empirical questions of fact which essentially appeal to experience, as these lie outside this domain. They are inadequate to deal with the problem Kant is addressing. Thus, Kant clearly distinguishes his psychology from that of his contemporaries. His aim is not to study the mind as it is in itself. In investigating "the cognitive faculties", he is seeking to determine *a priori* conditions for knowledge, an enquiry into the logical, conceptual and justificatory order of thought, not an investigation into actual thought processes. The result is an abstract description of the capacities a mind must have in order to have knowledge of the world, not speculation about actual psychological processes, as Kant is uninterested in the actual physical instantiation of mental or psychological processes. His aim is to show that the application of the categories to all objects of experience is justified. This entails a conceptual exploration into the prerequisites of cognition in order to

determine the necessary and universal elements that must be presupposed in all cases.

Yet, although Kant's psychology differs from the mainstream psychology of his day, it is still a psychology of sorts; it is the psychology of the thinking or knowing self, the "I" of "transcendental apperception". Kant uses the language of psychology regularly, for instance, in his discussion of cognitive faculties, powers and activities, such as our capacity to synthesise mental states and the unity that is required for this. This is particularly so in the chapter on the Deduction of the categories. He also makes frequent negative attacks on the value of both empirical and rational psychology. However, we can reconcile the apparent discrepancy and ambivalence by recognising that Kant clearly distinguishes the aims and methods of his transcendental enquiry from the mainstream psychology of his day. It consists of abstract descriptions of features that a mind must have to be capable of certain types of knowledge, not explanations of how it has them. Kant's overall aim in the Deduction is with determining the conditions and constraints on knowledge and with the justification of the objective validity of our knowledge claims, particularly claims to synthetic *a priori* knowledge. However, this entails a transcendental investigation into the necessary conditions of having experiences structured the way they are, a crucial one being the capacity to "synthetically bring into being a determinate combination of the given manifold" (B137), i.e. to unify representations and relate them to an object. We cannot study the connections among introspective contents of inner sense, but we can study what the mind must be like and able to do to have them.

Thus, Kant's negative comments about the value of psychology throughout the *Critique* leave his transcendental psychology intact. His reservations about it do not indicate that he came to doubt his own. Rather his problem is how to express novel ideas in a language insufficient to express them. Kant scholars have tended to give prominence to Kant's negative comments about the value of psychology in order to plead the case that he saw, at least vaguely, that he had erred in casting his arguments in psychological form. However, Kant makes it clear that he regards his discussion of psychology as playing a significant role. Although he says that he regards the subjective side as "not an essential part" of his enterprise, he also declares in the same sentence that it is of "great importance". Furthermore, in the

next sentence, immediately after saying that the subjective side is "somewhat hypothetical in character", he states that he will "show elsewhere, it is not really so" (Axvii).

What then, are we to make of the "subjective" or psychological aspects of the Critique? The crucial point is that the objective and subjective sides of the *Critique* are really part of the same argument, the latter being part of the former. The "objective" side was Kant's main concern, being a means to an understanding of how physics can be synthetic *a priori* knowledge. The "subjective" side enters the picture because one of Kant's strategies for dealing with the objectivity of our knowledge claims was to examine what a mind must be like to have them. Kant felt that since the categories must be applied there will always be a subjective element. He tells us at the beginning of the Deduction that in order to show how the categories could "relate to objects", i.e. be applied, he will examine the "subjective sources which form the *a priori* foundation of the possibility of experience" (A97). The main requirements of a mind capable of objective knowledge, he argues are its abilities to synthesise diverse states and the unity required to use them. Objective experience requires not only the capacity to unify representations and relate them in a determinate way to an object, but also the condition that they are brought together in one mind. As he writes:

> [A]ll unification of representations demands unity of consciousness in the synthesis of them. Consequently it is the unity of consciousness that alone constitutes the relation of representations to an object, and therefore their objective validity and the fact that they are modes of knowledge and upon it therefore rests the very possibility of the understanding (B137).

Kant's theoretical analysis of cognition is that the synthesis that produces the unity of apperception also produces all representations of objects. He also claims that "the synthetic unity of apperception is the highest point to which we must ascribe all employment of the understanding" (B134 note), and that "the principle of the synthetic unity is the supreme principle of all employment of the understanding" (B136, title). That is, although the aim of the objective deduction is to establish the legitimacy of the categories by showing that we could not have objective knowledge without them; this leads Kant into a discussion of the capacities of a mind capable of using them. Kant's ideas about synthesis and unity are an important part of the

deduction of the categories. Just as our knowledge requires the conceptual framework provided by the categories and the forms of Space and Time, it also requires synthesis and a unified mind. It is in this sense that Kant's considers his discussions of "psychology" as secondary to his main concern. Such discussions, although a by-product of his epistemology, are still an important part of his overall enquiry. So, although Kant's emphasis is on the objective side, i.e. the conditions of representations having objects, this entailed, absolutely, the subjective side, i.e. a deduction of the subject's transcendental nature. In fact, his conceptual analysis of human mental capacities is crucial to his transcendental enquiry. In Kant's opinion only an analysis of our cognitive experience can justify our epistemic powers. For example, that we can determine that one of our mental states follows another in time shows that we have a fundamental concept of cause. We cannot account for this as an acquisition of causality through the senses, for as Hume would agree it never was "in the senses". But this suggests that there must be something "in the mind" that supplies it. Even though Kant's primary interest may not have been in his "transcendental psychology" i.e. the abstract necessary structure of the mind, but in the objective conditions of knowledge, he nevertheless clearly acknowledges that such discussions are an important part of his enquiry and this suggests that he deemed it worthy of serious attention. In fact, his insights about human cognition and the self were extraordinary. Most importantly, they were not only remarkable in his own time but continue to have particular relevance today.

Fortunately this is now being recognised: philosophers have become increasingly critical of this tendency of past scholars to re-describe the *Critique* and squeeze it into the mould of post-Fregean philosophy and insist Kant's transcendental psychology can be an enlightening approach. [38] That Kant made frequent appeals to psychological processes is hard to deny, yet in actual fact he explicitly rejects psychologism (A52-54/B76-79). Also in the *Jäsche Logic*, he writes:

> This science of the necessary laws of the understanding and of reason in general, or what is one and the same, of the mere form of thought as such, is what we call *logic*…. Some logicians, to be sure, do presuppose *psychological* principles in logic. But to bring such principles into logic is just as absurd as to derive morals from life. If we were to take principles from psychology, i.e., from observations

> concerning our understanding, we would merely see *how* thinking does take place and *how* it is under various subjective obstacles and conditions; this would lead then to cognition of merely *contingent* laws. In logic, however, the question is not about *contingent* but about *necessary* rules; not how we do think, but how we ought to think. The rules of logic must thus be derived not from the *contingent* but from the *necessary* use of the understanding, which one finds in oneself apart from all psychology (*Jäsche Logic*, 9: 13-14).

He is not interested in the psychological manifestations of particular mental processes *per se*; his purpose is, rather, to explore the necessary conditions required for cognition, the generic conditions for the intentionality of thinking about objects in general.[39] In this sense his work contains a "psychological" element, which cannot be neglected. Kant's was interested, in particular, with the need for acts of synthesis and the kind of unity required for such acts and in this he has much of value to offer. Ignoring or dismissing this aspect of the *Critique* is a philosophical mistake, for in so doing the opportunity to contribute to current debates is wasted; not only within philosophy of mind, but also within cognitive science and cognitive neuroscience, where a picture of humanity is emerging that threatens to undermine the sense of our own freedom and agency. It therefore behoves us to give Kant his due and recognise his profound insights into the nature of human cognition.

In concluding this section, considered as a general movement, analytic philosophy has had an uneasy relationship with historically oriented philosophy, and it has tended to proceed without recourse to the past. As mentioned in Chapter 1, it has been frequently criticised for lack of historical awareness, ranging from what Hans Johann Glock calls "historiophobia" to "anachronism" (Glock, 2008, p. 90). According to Glock, this charge of historical neglect comes from several directions, and even "unites its two main rivals within contemporary Western philosophy, continental and traditionalist philosophy". Moreover, "this criticism is also shared by some who by common consent are analytic philosophers themselves" (ibid., p. 89). The reason for this was that the availability of a new tool, modern logic, meant that historically inclined philosophy could be dispensed with; indeed all earlier work could be dismissed in the name of a radical new beginning. There was also the perhaps rather arrogant tendency to regard philosophy proper as not even beginning until analytic philosophy came on the scene. Frege, the father of analytic philosophy

had himself distinguished between the analysis of thought, concepts and inferences and "either the study the history of our knowledge of concepts or of the history of our meaning of words" (Frege, 1884, p. vii). He regarded the former as the task of the philosopher, whilst discarding the latter as superfluous. Not only is analytic philosophy generally viewed as having a tendency to ignore the past, but also to distort it. Whenever analytic philosophers have engaged with past philosophers, they have tended to offer "rational reconstructions", ignoring or dismissing the parts which didn't fit in, as with Peter Strawson's analysis, among others. However, rational reconstruction of this kind, often leads to restricted focus, a narrow view and can also twist the views of the earlier works so much that they bear little resemblance to the original. Proponents of the analytic tradition may well think that conceptual analysis is a central part of (if not the only part of) philosophy, and whilst it may be true that it can do some good philosophical work, analytic philosophers should realise that philosophy is not without a history; philosophy is a historical movement as well as one concerned with more technical problems of logic and epistemology. As Robert Hanna has so cogently stated, "We cannot do philosophy in the present without implicitly adopting an understanding of philosophy's past and also that we cannot *properly* do philosophy in the present without making this implicit historical understanding explicit" (Hanna, 2004, p. viii). Richard Rorty has also accused analytic philosophy of being an attempt to escape from history (1979, pp. 8-9) and Bruce Wilshire has criticised its "radically ahistorical and modern progressive point of view" (Wilshire, B. 2002, p. 4). Moreover, Bernard Williams, one of the most respected and venerated practitioners of the analytic tradition, has urged analytic philosophers to adopt a more genetic and historical perspective (Williams, B. 2002). It is doubtful whether one can really do philosophy without engaging, not at a merely superfluous, but a deep level with at least some parts of the history of the subject. Philosophical thinking does not occur in a vacuum and more or less deliberately and explicitly, and with various degrees of awareness, we are always reacting to the problems, positions, concepts and arguments of our predecessors. Philosophers of the past, both recent and remote still have much to offer, and the better we know and understand their concepts the better off we will be in our own analyses of philosophical problems. This perspective is particularly called for in the philosophy of cognitive science, since, it is argued, present

difficulties in the field are a result of lasting influence of a largely forgotten or neglected philosophical heritage. A considerable number of philosophical writings, especially those connected to a science of consciousness, rest on fundamental conceptual confusions as a result of this philosophical neglect. As Charles Taylor, professor emeritus at Mc Gill University reminds us: "Philosophy and the history of philosophy are one. You cannot do the first without also doing the second. Otherwise put, it is essential to an adequate understanding of certain problems, questions, issues, that one understand them genetically" (Taylor, 1984, p. 17). The tendency to ignore or to show a disregard for historical issues is also prevalent within the sciences in general, where so often scientists' lay intuitions prevail. Perhaps this lack interest in the history of philosophy is due to a perceived irrelevance to the subject matter at hand, which then renders them unaware of the philosophical assumptions that might underlie a particular theory. With this in mind, the following chapter examines the fundamental problems in the history of philosophy that Kant was reacting against, with a view to articulating his radical views against them, in order to be able to assess the similarities to, and parallels with, contemporary issues within cognitive science. The essential focus of this thesis is that Kant was the first to deal rigorously with a certain model or picture of the mind and that this has implications for contemporary discussions of functionalism, the ensuing "hard" problem of consciousness and the problem of indexical self-reference.

3. A Historical Perspective

The previous chapter was concerned with defending Kant's transcendental psychology from dismissal by the mainstream analytic philosophical movement where it is commonly regarded as being unworthy of attention. Since the aim of this thesis is to examine the contemporary relevance of Kant's theory of mind, such a task was necessary to counter this claim and to eliminate any doubts there may be that it is worthy of such serious consideration. The current chapter turns to the problems in philosophy that were a legacy from his predecessors, notably Descartes and Hume, the purpose of which is to bring into focus the fundamental problems in philosophy that Kant addresses. For mainstream cognitive science has retained a large part of its Cartesian and empiricist legacy which has contributed to the perpetuation of a problematic picture of human mentality whilst also significantly shaping the contemporary understanding of human nature. It has been suggested that a certain lack of awareness of the history of philosophy might be to blame, rendering cognitive scientists the passive or unknowing captives of the Cartesian/Humean framework. Science relies on multiple layers of hidden assumptions, and scientists and practitioners, although they may be specialists in their field, are often unaware of the subliminal ways in which these assumptions guide their thinking, direct their explanations and eventually mould their understanding of the phenomena of study.

The word "consciousness" arrived rather late in the philosophical arena. The sages of ancient Greece had no such term. The word *Sunoida*, like its Latin equivalent *conscio*, meant the same as "know together with" or "having joint or common knowledge with another" (*con*-"together" and *scio* "to know"). The expression "conscious" was introduced into philosophy by Descartes (*Principles of Philosophy* (1640)[40]; who used the terms *conscientia*, *conscius*, and *conscio* to signify a form of knowledge, i.e. the direct knowledge of what is passing in our minds. What we are consciously aware of are "thoughts", and these include ordinary thinking, sensing or perceiving, understanding, wanting and imagining. Fifty years later John Locke, in his *Essay Concerning Human Understanding* (1690), defines consciousness as "the perception of what passes in a man's own mind", which we access by means of "inner sense" (*Essay*, II-i-19). For Descartes we do not

immediately perceive the external world. Perception takes place in the *soul* (nowadays, the brain) only and is caused by the movements caused in the brain (soul) by vibrations affecting the sense organs. Empiricists Locke, Berkeley and Hume adopted this view as more or less self-evident (see Locke, 1690/1959, Bk II, Ch. IX, §3; Berkeley, 1710/1975, Pt I, §1; Hume, 1748/1962, Bk I, Pt IV). For both Descartes and the empiricists, perception or cognition pertains to "ideas", and not directly to external objects. Ideas are modes of thought that, Descartes writes in the *Third Meditation*, "represent" or "present" or "exhibit" (he uses these words interchangeably) objects to the mind (see Cottingham, J., Stoothoff, S., Murdoch, D. 1992, pp. 88-9). Descartes provides multiple definitions of the term idea, which he divided into six distinct kinds of entities. Locke defines idea as "that term which, I think, serves best to *stand for* whatsoever is the object of the understanding when a man thinks"[my italics]. Hume agrees with Locke but limits ideas to mental reconstructions of perceptions caused by "impressions" of the senses. This notion, of ideas representing or standing for something in the mind was adopted, again uncritically, by scientists at the start of the mid 20th century cognitive revolution, where it was combined with ideas of modern computer technology, and where these ideas were translated into the mental representations of cognitive science, and conceived of as "mental causes", the interactions of which are said to constitute mental processes that constitute a mind. According to this picture, the human mind just *is* a complex system constituted by mental representations, operating on or computing information, whether in abstract workings or functions of the mind or later, in cognitive neuroscience, in the "wetware" of the brain.

Thus, when Descartes invested the Western world with an immaterial *res cogitans* and a material *res extensa*, he initiated a host of seemingly insurmountable philosophical and scientific problems that have plagued philosophers and scientists since. For the contemporary "hard problem of consciousness" is the legacy of this world view; it is, in effect, an updating of the same conceptual model, and gives rise to similar problems, one of which is of explaining how an immaterial mind (now a material brain) can give rise to ineffable, first person, phenomenal consciousness; the puzzle of how to fit a subjective conscious mind into the realm of objective empirical science. Descartes thought of the body as a machine and the mind as an immaterial entity or substance which animates it. This led to the problem of interaction, of how

an immaterial mind could cause changes to the physical body and how a physical body could cause changes to the mind. Although many have doubted his theory of mental animation, the idea of the body as machine has been retained and with it the same question of interaction - how such a machine could produce consciousness, subjectivity and phenomenal experience in general. [41]

Hence the scientific materialism which lies at the root of cognitive science and cognitive neuroscience has its origins in this fundamental misconception or error of thought, the often implicit, unrecognised assumption of the Cartesian/Humean view of the mind, in which phenomenal consciousness is conceived of as a kind of entity, (or *quasi* entity) that either "contains" (Descartes) experiences or which "consists" of them (Hume's bundle of experiences). Mind thus conceived is a kind of inner repository where informational content originating from the external world can be transferred and manipulated. There is also the Cartesian idea of the body as a machine separated from the non-mechanistic, immaterial entity, the mind. Cognitive scientist and cognitive neuroscientists who are interested in the mind, have a set of questions that they want to understand: what is thought, what is consciousness, and what is cognition? How can the electrical firing of millions of neurones in the brain create conscious experience of the world? Human consciousness seems to be unified, some things seem to be in awareness and others not. They seem to be presented before the mind as a stream of consciousness within a theatre, there is continuity over time, and they are experienced by the same subject. Scientists and philosophers attempt to come to some agreement about these issues. Some scientists claim that now there are new technologies of brain science, there is no longer a need to pay attention to what philosophy has to say about these questions, and dismiss it as irrelevant. Prominent examples of these kinds of scientific thinkers are molecular biologist turned neuroscientist, Francis Crick and neuroscientist Christof Koch, (currently chief scientific officer of the Allen Institute for Brain Sciences) whose combined research aims at discovering the neural correlates of experience (NCCs) in order to find the difference between neural activity that produces consciousness and that which does not with the aim to get nearer to what consciousness *is*. As Crick put it "No longer need one spend time attempting … to endure the tedium of philosophers perpetually disagreeing with each other. Consciousness is now largely a scientific problem" (Crick, 1996, p. 486). However, *crucially*, most of what

empirical science has to say about consciousness, language, memory, perception and emotion *is* the expression of a particular philosophy that is unquestioned, the quasi Cartesian/Humean view that the mind is something inner; (the brain/ mental processes) disconnected from other people and from the outer world. Also there is the Cartesian manner of thinking about consciousness as a stream of consciousness in the "theatre" (Dennett, 1991) of the mind. A successful science of consciousness must be able to give an account of the contents, the stream, the unity and the continuity. As was mentioned above, there is the notion that the external world gives rise to a particular state in the brain, which will be subjectively experienced as a conscious representation or model of the external world. This internal representation, whatever it amounts to, is the principal subject matter of an emerging "science of consciousness", the salient point being that there is "something it is like" to have an experience, that consciousness is *essentially* characterised by reference to there being "something it is like" to be in certain mental state.

It is suggested that the "hard problem of consciousness", along with the related quest to discover the NCCs, is based on a confused understanding of mentality, and far from the charge that cognitive science and neuroscience has no need for philosophy, it needs its guidance now more than ever (see Bennett and Hacker 2003; Thagard, 2009; Hacker, 2012).[42] Whilst conceding that cognitive scientists and neuroscientists have made significant discoveries concerning the workings of the brain and cognition, these discoveries have been obscured by their presentation within an incoherent conceptual framework that, despite its professed materialism, is fundamentally Cartesian/Humean in the sense mentioned above. Although some contemporary philosophers do recognise and criticise these hidden assumptions underpinning much of cognitive science and neuroscience, it has gone largely unrecognised that the radical critique of this model or understanding of the mind began with Kant, whose powerfully innovative transcendental inquiry into the necessary *a priori* conditions for cognition was not only a major philosophical breakthrough in his own time, but still surpasses contemporary anti-Cartesian efforts today. It is important to recognise the innovation of his thought and give his profound philosophical insights into the nature of human mindedness their just due. With this in mind, the following section provides an analysis of the fundamental

problems in the history of philosophy that Kant was reacting against, with a view to articulating perspicuously his radical views against them.

3.1. Fundamental Problems in Philosophy

Beginning with his Inaugural Dissertation (1770) on the difference between right and left-handedness and spatial orientation, Kant produced one of the most comprehensive and influential writings in the history of philosophy. His central thesis, that the possibility of human knowledge presupposes the active participation of the mind, is to be found in the *Critique of Pure Reason*, which is universally held to be a watershed in metaphysics and epistemology. His "Copernican Revolution", as he proudly termed it, is a reversal of approach to the thinkers before him. Kant claimed that just as Copernicus had reversed the assumptions of classical astronomy that the sun moved around the earth, his project set out to turn around the assumptions of his predecessors, which was that "representations" must conform to an object independent of the mind in order to constitute knowledge. For Kant, the converse was true; any possible object of experience must conform to *a priori* conditions of knowledge. It must reflect the constitution of the cognitive faculties brought to experience by the subject, his "transcendental psychology".

Transcendental psychology is Kant's theory of mind, the study of those faculties that are *a priori* required for cognition. What this means is that cognition is taken as fundamental and then it is worked out through *critique*, a process of critical analysis, what makes that possible. Prior to this, theories of mind had centred on the epistemological question of whether knowledge was a product of senses (broadly, Hume and the Empiricists) or a product of pure thought (broadly, Descartes and the Rationalists). Empiricism is the theory that experience rather than reason is the source of knowledge; Rationalism, that it is reason alone that is the source. Kant's question, in a nutshell, was "What are the *a priori* conditions of the possibility of experience"? He overturns the relationship between knowing subject and experienced object, arguing that the properties that we can assign to the object are nothing but the very preconditions for knowing the object itself: in other words, the world is populated by objects only inasmuch as they fit our predetermined sensory and cognitive apparatus. Although Kant shared with his forebears an interest in the

foundations of knowledge, he thought that something was lacking with the enterprises of them both. He offered a solution to this deficiency that combined elements of each position yet passes beyond them. For unlike his predecessors, it is the "representation" itself that makes the object possible rather than the object that makes the representation possible. This introduced the human mind as an active originator of experience rather than just a passive recipient of perception. Kant was concerned with the epistemological issue of how it is that our cognitive claims are vindicated. As discussed in the previous chapter, this was articulated through his famous *quid juris* question (A84/B116), which asks by what rights or legitimacy do we apply concepts, which are not acquired from experience, to the contents of experience? (A85/B117). This philosophical quest led him to draw the limits of human reason through his emphasis on the "transcendental" and "necessary conditions" for all cognitive pursuits. Any knowledge claim which went beyond these conditions is deemed illegitimate.

Despite his sense of the limits of human reason, however, Kant's insights were seminal, and transformed the nature of Western thought. Although, like philosophers before him, he sought to explain the possibility of new scientific knowledge and the possibility of human freedom, his solutions to such problems were radically different to those of his predecessors. He thought that what was wrong with previous philosophical enquiry was that philosophers had not seen the necessity of a very general inquiry into what makes sense. Those prior to him had attempted to explain our knowledge of the world in terms of a passive reception of an external reality or truth, an objectively perfect realm that was only confusedly apprehended by the senses. Kant redefined the concepts of "truth"' and "reality" and with them the task of philosophy. The purpose of philosophy is not to arrive, by a process of intricate inference, at knowledge of an objective truth not directly accessible to experience. Indeed, he took it for granted that we have objective knowledge and that sensible experience is a foundation of such knowledge. What interested him was how the various types of knowledge hung together; and in order to come up with a solution to this he set out to define the very conditions that make cognitive experience itself possible. Thus, Kant's main concern, unlike those thinkers before him, was not with what knowledge we can derive from experience, but in human understanding itself and whatever conceptual principles or rules there may be that govern that

understanding. That is, his purpose was not to discover how we infer reality from experience but to examine the necessary conditions involved in us "having cognitive experience" or knowing anything at all. This examination was revolutionary, and it transformed the Western conception of a human being from that of a "spectator" of the natural world, who, in a sense, exists apart from it to that of an "agent" whose spontaneous activity is necessary to its very creation. For the subject of experience, although not entirely producing its own experience, contributes so much to it that without this input no experience is possible. As he wrote in the preface to the Second edition of the *Critique*:

> Hitherto it has been assumed that all our knowledge must conform to objects. But all attempts to extend our knowledge of objects by establishing something in regard to them a priori, by means of concepts, have, on this assumption, ended in failure. We must therefore make trial whether we may not have more success in the tasks of metaphysics, if we suppose that objects must conform to our knowledge (Bxvi).

To begin with the Cartesian picture: Descartes had set out on a quest for absolute certainty. His purpose was to place knowledge on a sure foundation by "out-doubting the sceptics"; to show that all our knowledge is based on rational foundations and that these involve self evident truths from which everything we know can be deduced. His method was one of universal or hyperbolic doubt. Thus, in Meditations I and II he writes that he is going to reject as false all that he has been taught about the world as well as his own experience:

> I shall go on setting aside everything which might, in the slightest degree be supposed to be doubtful, just as if I had found out that it was completely false; and shall continue to follow in this path until I find something which is certain, or at least, if I am unable to do anything else, until I have learned that it is certain that there is nothing in the world that is certain (*Meditations 1*).

However, although his method was to doubt "everything (...) which might in the slightest degree be supposed to be doubtful", his purpose was to find an absolutely secure foundation, an Archimedean point for human knowledge and something that we could be absolutely certain of and which could never be false or doubtful. This was to be "so certain and of such evidence that not even the most extravagant of suppositions entertained by the sceptics would be capable of shaking it" *(Discourse on Method, Part IV*). After casting doubt on sense experience, scientific information

and mathematics, Descartes found the certainty that he sought in the Cogito. As soon as one tries to conceive any condition under which "I think therefore I am'" could be false one is reassured of its very truth. There can be absolutely no dispute about the existence of the Cogito, for his existence as a thinking thing is presupposed in the very act of doubting. As he writes in the Discourse:

> I observed that, whilst I thus wished to think that all was false, it was absolutely necessary that I, who thus thought should be somewhat; and as I observed that this truth , *I think, hence I am (cogito ergo sum)* was so certain and of such evidence that no ground of doubt, however extravagant could be alleged by the sceptics, capable of shaking it, I concluded that I might, without scruple, accept it as the first principle of the philosophy of which I was in search (*Discourse on Method, Part IV*).

Moreover, the nature of the self is revealed to us in an immediate way. For Descartes then asks "What am I?" and concludes, from the fact that he can "clearly and distinctly" conceive it, that he is a "thinking thing" or "substance". From this one truth Descartes claims that we can discover a criterion about all truths, namely clarity and distinction. On the basis of "clear and distinct ideas", the truth of which is guaranteed by a perfect, benevolent and undeceiving God, we are able to have absolute certainty about the external world. So, according to Descartes we can have absolutely certain knowledge of ourselves and certain aspects of the external world, for from an analysis of our rational faculties and by employing particular procedures of reason alone we can make judgements that cannot possibly be wrong. In fact, we are compelled to make them; for what is "clearly and distinctly" conceived imposes themselves on us in such a way that we must accept them as true.

Let us compare this with the Humean position. Hume's epistemological concern was also with "knowledge" but he rejects the Cartesian methodology of universal doubt and in so doing is forced into adopting an attenuated scepticism of his own. For Hume denies that there is direct access to a continuous self, the primary and certain foundation on which Descartes builds his edifice of knowledge. On the contrary, there is no such self and no such certainty. We are doomed to be cut off from knowledge both of the external world and of the self. We have no rational foundation for knowledge of the external world, for as he puts it:

> The mind has never anything present to it but the perceptions, and cannot possibly reach any experience of their connexion with objects. The supposition of such a connexion is, therefore without any foundation in reasoning (*An Enquiry Concerning Human Understanding, section VII, part 1*).

Furthermore, philosophy teaches us that:

> Nothing can ever be present to the mind but an image or perception, and that the senses are the only inlets through which these images are conveyed, without being able to produce any immediate intercourse between the mind and the object. The table, which we see, seems to diminish as we move farther from it; but the real table which exists independent of us, suffers no alteration; it was therefore nothing but its image, which was present to the mind (ibid. §12. 8).

So, according to Hume's empiricism, just as we have no rational foundation for an external world that is separate from us, neither do we have certain introspective awareness of a self. Hume says that when he introspects he cannot perceive a self at all. As he writes in the *Treatise of Human Nature*, when he turns his reflection on himself, he can never perceive a self without one or more perceptions; nor "can perceive anything but the perceptions":

> I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can perceive anything but the perception (*Treatise Book I, Section IV, Ch. vi*).

For Hume, the self or mind is nothing but a "bundle of perceptions" just as external objects are "bundles of perceptions" and it is incoherent to suppose that it is a substantial unity whose nature could be grasped by any kind of "rational intuition" of the sort that Descartes claims.

Now let us consider the Cartesian and Humean positions together in order to see more perspicuously how Kant supplements their accounts: The aspect of the self which the rationalists emphasise is being a subject aware of its own existence, its true nature being comprehensible only in consciousness. For Descartes, the one thing we could be absolutely certain of was the self and its existence. As he says in the *Discourse on Method,* although he could doubt everything in the world, the one thing he could be certain of was that the "I" that thinks should be "somewhat" (*Part IV)*. Thus, Descartes' epistemology starts out from a picture of the self as a solitary mind

having certain knowledge only of its immediate apprehension. From this direct awareness, what he terms his Archimedean point, it is able (with some help from a benevolent God) to find itself in a world grasped both as external and real. Against this Hume makes the point that in introspection we are not aware of a "subject" at all but an "object" like any other. For Hume this object is a mere "bundle of impressions" just as what we are immediately aware of in experience are not external objects but "bundles of impressions", which in turn give rise to ideas on the basis of which, in combination with the mental habit or custom and the "association of ideas", we derive our knowledge of the phenomenal world. For Hume, our natural instinctive belief that there is a real world external to our senses cannot be rationally justified. Since all we are aware of are "perceptions" or "images", we cannot know that they correspond to anything in the external world. The same applies to our existence as persistent unitary selves. We have no experience apart from the "bundle"; therefore we cannot infer such entities.

Kant considers Descartes' reasoning "transcendent" and incapable of being verified in experience, whereas Hume's account is reductivist and lacking in explanatory power. He also felt that it undermined both scientific thinking and the moral claims of human reason. Something deeper than mental habit or custom and the association of ideas is necessary, for one cannot attempt to infer reality from experience when experience itself stands in need of explanation. He therefore uses a reversal of approach by removing the primacy of our sensorially derived experience and rendering it peripheral. Rather than attempting to infer the existence of the empirical world from direct apprehension of his own existence, as had Descartes, or from our perceptions, as had Hume, Kant takes the self-evident truths of empirical experience and deduces a metaphysical subject who necessarily exists as their counterpart. For Kant if the phenomenal unity of experience is to take place, subjective unity must first be presupposed.

Although Kant agrees with Hume that when we introspect we find no unitary self or ego, but at most a bundle of impressions, he argues that a unitary active self must, nonetheless be postulated in thought. For Kant, the elements in Hume's bundle of perceptions has a unity, a unity which results from the amalgamating activity of a subject, a thing (an x) of which we can say nothing but which plays a certain role in our cognitive abilities and which necessarily exists and provides the explanatory

power missing in the Humean account. This Kant termed the "transcendental subject". Both Descartes and Hume held "representationalist" theories of perception and their associated "problem of the external world" and this has been bequeathed to contemporary cognitive science and neuroscience where they have lodged again in a wide variety of sense data theories and reductionist programmes. Kant was the first to provide a radical critique of this understanding of the mind.

Thus, although Kant thought that there were serious errors in the views of his predecessors, he also thought they were partially correct. He agrees with Descartes that the actuality of self awareness is indubitably present and also with Hume that sensibility provides us with a bundle or "manifold" of impressions. However he combines both to conclude that we have to connect these impressions in a determinate way and thus impose a necessary order on them through which the objective world is rendered intelligible. According to Kant, there are twelve logical functions of judgment necessarily presupposed in the human mind, the "categories of the understanding", which can only be used in legitimate cognitive judgments if and when they are used to identify particular objects or events in space and time. Moreover spatio-temporal designation is essential to the presentation of objects in the world that are experienced, and also to our cognitive reference to them. The twelve categories, identified in the Metaphysical Deduction section of the *Critique*, are objectively valid for empirical objects because they are the very conditions of the possibility of experiencing objects at all. We cannot experience objects unless we experience them as falling under these *a priori* principles. There is, therefore, a necessary *a priori* interdependence of object consciousness and the transcendental unity of apperception, the ultimate condition of the possibility of experience. There could be no experience of objects at all if the manifold of representations were not synthesised or combined by the mind through the unity of apperception (A108/135). He also points out that an important source of confusion inherent in the Cartesian picture is that it overestimates its epistemological position with respect to the self by claiming that introspection presents us with epistemic contact with "a thinking substance". Rationalists freely used *a priori* concepts in metaphysics without asking: How can *a priori* concepts be referred to any particulars about which we purport to make metaphysical claims? All that could be said of the introspective experience is

that there is thinking going on, not that there is a metaphysical subject that can be known. As he writes:

> [I]t is evident that rational psychology owes its origin simply to mis-understanding. The unity of consciousness (...) is here mistaken for an intuition of subject as *object* and the category of substance is then applied to it (B421-2) [my italics].

As Kant diagnoses it, we do not and cannot have any such intuition of a self. Our knowledge of the world and also our awareness of ourselves has a different foundation. It is founded on the "transcendental" ground (*grund*) of the unity of consciousness, what Kant terms the "transcendental unity of apperception". He also describes it as a "pure unchanging consciousness", a unity which is a necessary condition of objective experience and objective cognition or knowledge, since without it no object could be thought.

Descartes had stressed that the self is a subject of experience. For Hume any self that we encounter is an "object", a "bundle of impressions" just as other objects are "bundles of impressions". Kant's critical philosophy can be seen as an attempt to settle the matter to the satisfaction of both parties and to do justice to both empiricist and rationalist accounts of self-knowledge, whilst also passing beyond them. For Kant, the self is to be regarded as both "subject" and "object". Unlike his predecessors he elaborates a dualistic yet monist theory of the nature of cognition. There is a distinction between the "phenomenal" self and the self of transcendental apperception. The phenomenal, empirical or experiential self is an object in the world and is a product of the process of "objectification", as are other objects in the world. Objects in the world present themselves to us as outer appearances and are always changing and, likewise, the phenomenal self. As Kant puts it "No fixed and abiding self can present itself in this flux of inner experiences" (A107). Thus, Kant agrees with the empiricist claim that we can have a kind of knowledge of ourselves through "inner sense". However, he also agrees with the rationalists that this does not end the matter. For as well as the phenomenal self or self as "object" we must also recognise the self as subject, the "transcendental subject" or "transcendental unity of apperception", which he describes as "a pure unchanging consciousness". This is not a sense of self occasioned by experience but is the ultimate condition of all experience or knowledge of the objective world. In fact, it is a necessary prerequisite of our having any sort of experience at all that the self for which the

experience exists should be thought of as a unitary and unchanging subject. Otherwise put; it is a presupposition of experience itself that every object should relate to a subject and that this subject should be aware of its own identity.

Kant makes clear that it is illogical to claim that this "transcendental subject" is the experiential self, since if it is the ultimate prerequisite of experience, it cannot then itself be defined by appeal to experience. Unlike Descartes, Kant does not attempt to establish his philosophical system on an Archimedean point, on an allegedly absolutely indubitable foundation. Instead, he begins with the objective reality of experience and explores the conditions for its possibility. The objective reality of experience does not necessitate any external proofs, since "the *possibility of experience* is … that what gives objective reality to all our *a priori* modes of knowledge" (B195-6/A156-7). Kant's point, in a nutshell, is that it logically follows from the premise that the apperceiving self is the condition of all experience that one cannot become aware of it through our ordinary cognitive capacities since this would be to objectify it which is logically incoherent. For to conceive the apperceiving self as such is to make it an item within consciousness, and thereby to remove it from its role as subject. As he says in the Preface to the second edition of the *Critique*:

> My soul (...) cannot indeed be known by means of speculative reason (and still less through empirical observation) (…) For I should then have to know such a being as determined in its existence, and yet not determined in time - which is impossible, since I cannot support my concept by any intuition (B xxviii).

Nevertheless there is, for Kant, a way in which we can become aware of the transcendental self, through a direct unprocessed recognition. For he also says:

> [I]n the synthetic original unity of apperception I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am (B157).

Kant does not mean that we can infer from the unity of apperception, the "I think" the certain existence of a Cartesian conscious subject. The immediate awareness we have of the subject self in apperception does not give us "knowledge" of the nature of the self in the way that Descartes claims. As he says: "We do not have, and cannot have, any knowledge whatsoever of any such subject" (A350). The "unity of apperception" is the mark not of the consciousness of a Cartesian self but of the consciousness of anything at all.

In Descartes' view the mental is absolutely distinct from physical processes; the two take place in different and distinct "substances". To account for their interaction Descartes asserted that the pineal gland was where the thoughts enter the body. Today a new form of dualism (property dualism) also substantialises or objectivises mind as well as matter, and eventually leads to an explanatory gap (Levine, 1983) and a modern-day version of the mind-body problem. It also leads to Cartesian-like attempts to prove the existence of a non-physical fact about consciousness; that there must be something real beyond the known physical world that must account for it. On the other hand, in the manner of Hume, modern functionalist explanations deny there is a problem; that any explanation in terms of processes would leave exactly nothing further to be explained, no "hard problem" exists above and beyond the explanation in terms of mechanical processes. In other words, the Cartesian and Humean approaches to cognition continue to hold sway today. Functionalist approaches to cognition are, more or less, updated Hume. Although it was Thomas Hobbes who was first to theorise about the mind as a computational device, it was Hume who first devised a "mental mechanics" inspired by Newton's mechanical philosophy of (non-human) nature. It was also Hume who shed the theological baggage which had encumbered his rationalist predecessors and founded a true science of human nature. Daniel Dennett is an avowed Humean and is highly vocal in his contention that cognitive science retains the often unacknowledged remnants of Cartesian dualism, which he calls Cartesian Materialism. Tellingly, he regards Hume as pivotal in the history of functionalism and credits him with formulating the central problem of cognitive science, which he calls Hume's problem, which is how to discharge the homunculus or little man. According to Dennett, Hume's trouble was that it did not occur to him that simple homunculi might work, small subsystems, the sum total of which gives us the illusion of consciousness. The mind, for Dennett contains several sub-persons or agents and each of those in turn contains other agents and so on, right down until we reach individual neurons which Dennett, following McCulloch and Pitts, conceived of as being simple logical switches. This is known, famously, as Homuncular Functionalism, the postulation of a hierarchical series of mind levels (homunculi) which become simpler and simpler in terms of complexity and organisation until at the level of neurons the homunculi are discharged.

On the other side of the debate, those who think there is a "hard problem of consciousness" simply take Descartes' notion of immediate thought as given as a brute fact. Chalmers writes, for example, that "if it were not for the fact that first-person experience was a brute fact presented to us, there would seem to be no reason to predict its existence" (Chalmers 1990). Whatever the theory of consciousness proposed, the core idea that there is a hard problem of consciousness that needs explaining eventually leads to the modern view that consciousness is merely epiphenomenological, an accompanying event to cognition that lies outside the chain of physical causation. Like the Cogito, consciousness is something that evades all causal explanation. However, this way of thinking gives rise to a host of philosophical perplexities that are as intractable today as they were in the past. Much of the contemporary discussion of consciousness remains mired in conceptual problems from centuries ago that should by now have been left behind.

Kant was sharply aware of this tendency to be misled by this kind of reasoning and called Descartes' position "sceptical idealism". The sceptical idealist holds that whilst one may be certain about the existence of one's own states of consciousness, the external world will remain an unproven assumption. For Descartes it is the direct knowledge of one's own existence, the Cogito, that will supposedly furnish the starting point for any attempts to establish the existence of other objects, but never manages to re-establish the connection. Hume, on the other hand, although sceptical of the Cartesian self, was also sceptical of the external world since all there are are sense impressions. Our natural instinctive belief that there is a real world external to our senses, however, cannot be rationally justified. Perception for the empiricists pertains to "ideas" or sense experience, not directly to external objects. From this point onwards, there appears to have been an undisturbed tradition of these central methodological commitments right up to present-day cognitive science. It is for this reason that Kant's reversal of approach to those thinkers before him can provide genuine insight. Although his psychology was secondary to his main concern and a mere by product of his epistemology, his refutations and solutions to what he saw as the problematic philosophy of his predecessors were not only important in his own time but continue to have relevance today. They have immense value in helping unravel the philosophical perplexities that have arisen in contemporary science of the mind. In particular, they are able

shed light on the modern day problem of consciousness, because they address the *sense* in which one can say that a person is a subject of his own awareness. They focused on the question of whether one is one ever directly aware of such a subject. These questions have to do with actual phenomenal consciousness *per se*, what we find when we introspect, when as Hume put is "we enter most intimately into ourselves". Hume's denial that there is introspective awareness of a subject is based on the model of sense-perception. Thus, he argues that although we are aware of perceptions of objects, we are not aware of a perception of a self among them.

This method of characterising self-awareness has been repeated ever since by philosophers and cognitive scientists of various persuasions. Many inadvertently and unknowingly support the view that introspection is to be conceived on the model of perception and accept the Humean model that the self is not among the objects perceived. This has also been regarded as having important implications concerning such topics as the nature of self-reference and self-knowledge. Thus, Hume's denial that we can have perceptual knowledge of a self, that we perceive a self by what Kant termed "inner sense" has been influential and continues to beguile and perplex. Empiricists during the earlier part of the last century followed his interpretative model and claimed that when we introspect we are not aware of a self but only of "sense data". This view is to be found, for example, in the works of Bertrand Russell, (1912, 1927); Broad, (1925); G. E. Moore, (1953), and Ayer, (1956). Indeed, this is the natural conclusion of the idea that access to our own minds is to be conceived on the model of sense perception. Kant saw the error of this way of thinking about the mind; for, as discussed, if perceptions are mental particulars *other* than the self which serve as objects of introspective awareness then the self cannot itself be such an object. He recognises that we are all prone to an all-pervasive "transcendental illusion", which is very hard to recognise, let alone to compensate for, which is a pervasive intellectual illusion, modelled on the perceptual, which predisposes even the "wisest" of people to accept as sound certain invalid arguments for substantive theses about the nature of the self. Kant claims that human reason is *inevitably* caught up in this highly problematic way of thinking: one that it cannot dispense with, even though it is entirely groundless as a description of reality. He says in the Preface of the first edition of the *Critique* that his purpose will be to help the reader at least *be aware* of transcendental illusions, even though their effects

on their thinking will remain unchanged. The core idea is that because metaphysics by its nature exceeds the bounds of experience, it is easy, in fact all too easy, to take the mere logical coherence of a position as a sign of its truth, which leads us into "transcendental illusions": the substitution of logical for real possibility.

3. 2. The Paralogisms - apperceptionis substantiatae

The Paralogisms of Pure Reason are a series of chapters which concern Kant's views on the self, where he seeks to contrast them with his predecessors, Descartes, Leibniz and Hume. To return to Hume, his argument, simplified, is: I can only be aware of objects via impressions, I am not an impression, and therefore I cannot be aware of "a self"". Nevertheless, Hume appears to have some conception of what a self might be despite his denial that he is aware of it because having a mental state (an impression) consists in perceiving it and the act of perceiving is relational. There cannot be a perception without a perceiver. Awareness of objects via impression implies a relationship between perceiver and perceived. Hume seems to have taken it for granted that all mental states are relational; impressions are had, but is puzzled that he finds no impression of a "haver". In an appendix to the *Treatise*, he writes:

> When I turn my reflection on *myself*, I never can perceive this *self* without some or more perceptions; nor can I ever perceive anything but the perceptions. 'Tis the composition of these, therefore, which forms the self (*A Treatise of Human Nature, Appendix*).

If he is aware of any self at all it is "a bundle of perceptions". However, he is puzzled because he cannot find the "principle of connection" that unites this bundle. Hume is looking for "something". In other words, he supposes that awareness of self is akin to perceptual awareness of objects. This can be seen by the fact that the conclusion of the above argument would only follow if for "self" we could substitute "object". He also writes the following in the later *Inquiry Concerning Human Understanding*:

> As our idea of any body, a peach, for instance is only that of a particular taste, colour, figure, size, consistency, etc. so our idea of any mind is only that of particular perceptions without the notion of anything we call substance, either simple or compound.

The Humean conception of introspection or self-awareness is analogous to seeing. Seeing something is relational. For Hume when we see a peach there is a relation set up between the properties of the peach and whatever it is that perceives them, although he is perplexed because we can find no impression of that which perceives, no impression of "substance", just as there is no impression of a substance in which the properties of the peach inhere. Descartes also models self-awareness on perception. Implicit in the Cartesian picture is the assumption that in the "I think" we are aware of some entity. Hence Descartes' move from the Cogito Ergo Sum "I think therefore I am" to "What am I?" and his answer, that he is "a thinking thing", a "*res cogitans"*. On the Cartesian picture awareness of self is analogous to awareness of any other particular. This is a result of the natural tendency to reify the "I" into a kind of intuitable non-bodily entity. The reasoning would be as follows: Bodily entities are individuated by their properties and relations to other things. Since the self or *soul* is a non-bodily entity it must be individuated by non-bodily properties. This is evident from the way that Descartes describes the self in negatives of the descriptions given to bodies, as non-spatial, non-observable, and non-material. Kant talks of our natural tendency to fall prey to this illusion which he calls *apperceptionis substantiatae* (A402), the reification of the "I think" into a kind of object. Because we can speak about the self both empirically and "transcendentally", we are prone to confusion and often speak about the self in an empirical way. In so doing we make the "I" into a thing, *reify* it and speak of it as if it were an object. Although it is legitimate to speak of the "I" of the "I think" we must do so "transcendentally". We must not make the mistake, a mistake to which we are so prone, of speaking about the self as if it were an entity or thing. Kant says in the first Paralogism that although the "I" of "I think" is formal, the mere "form" of consciousness and therefore insubstantial, it must nevertheless be *conceived* as substantial. The formal condition of experience involves an inbuilt illusion, viz. that my experiences seem to be had by a genuine unitary subject.

A paralogism is a syllogism, and a syllogism is a mode of deductive argumentation consisting of three propositions, the third being a deduction from the first two, one of which is a major premise under which everything is subsumed "under the condition of the rule". [43] However, despite giving the initial appearance of a valid argument because both the premisses are true, it is fallacious. The four

psychological paralogisms that Kant examines are all said to have the form of a *sophisma figurae dictionis* in which the middle term is used ambiguously. According to Kant we are liable to mistakenly reason the following:

That which is a subject is substance.

The "I" of "I think" must be *represented* as a subject.

Therefore the "I" of "I think" *is* substance.

However, Kant shows that it does not follow from the premise that the "I" is necessarily conceived or represented as a subject, that it *is* a subject, and therefore it does not follow that the "I" is substance. The fallacy rests on the ambiguity of the middle term, "being a subject" and "being represented as a subject". What Kant terms the "analytic unity of apperception" is the subjective unity which underlies all experience of objects, and where reason goes astray is in postulating a *real* self as corresponding to this merely "formal" unity of the self. As he writes:

> [T]here is nothing more natural and more misleading than the illusion which leads us to regard the unity in the synthesis of thoughts as a perceived unity in the subject of these thoughts. We might call it the subreption of the hypostatised consciousness [*apperceptionis substantiatae*] (A402).

Nevertheless this is a natural illusion that "even the wisest of men cannot free himself from" (A339/B397). Wittgenstein says something similar in the Blue Book:

> We feel … that in cases where "I" is used as subject we don't use it because we recognise a person by his bodily characteristics, and this creates the illusion that we use the word to refer to something bodiless, which however, has its seat in our body. In fact *this* seems to be the real ego, of which it was said *Cogito, Ergo, Sum* (Wittgenstein, 1972, pp. 69-70).

The difficulty is that both Descartes and Hume make the mistake of assuming the "I" functions like the objects of "outer sense". Hume's difficulty arises from the fact that he constantly seeks for evidence of a subject, an impression from which the idea of a self can be derived. But although the relations between our ideas may be traced through time by memory there is no evidence of any self that connects them. No impression of a core self that ties them together. Descartes' error lies in supposing that in introspection one is aware of "a thinking thing". Thus the mistake of both Descartes and Hume derives from the supposition that consciousness of self must be an experience of "something". Descartes inflates it into a *res cogitans*, a thinking thing or "substance" whereas Hume, finding no impression of it concludes

that it does not, in some sense, "exist" or if it does there is no more than the bundle of impressions. Both Humean scepticism and Cartesian dualism are based on the perceptual model. The implicit assumption underlying both is that in introspection we must be provided with identifying facts about ourselves and because we are not we either assume we are non-bodily entities (souls) or that we do not in some sense "exist" save as an idea of a bundle of impressions. Kant diagnoses the error in both cases. We are not presented with identifying features at all. Kant tells us not only that we have no evidence for a self (agreeing with Hume) but that there can be no such evidence.

Descartes, along with other rationalist philosophers mistakes the absence of any intuitions for the intuition of something with remarkable properties, the empiricists (Hume) that there is no introspective awareness of a self at all. In fact, Kant agrees with Hume that there is no perceptual awareness of a persisting self by "inner sense". He writes: "For in what we entitle soul everything is in continual flux and there is nothing abiding" (A381). He also agrees with Descartes that there is a kind of awareness of a self, for he continues "except (if we may so express ourselves) the "I", which is simple solely because its representation has no content, and therefore no manifold, and for this reason seems to represent, or (to use a more correct term) denote, a simple object (A381-2). However, for Kant, we must be careful of our tendency to confuse speaking about the "I" in a transcendental way with speaking about it as if it were a kind of object, for this would be to characterise the "I" as it exists beyond our possible knowledge of it and is a mistaken inference from a transcendental to an empirical claim. Against this tendency Kant frequently reminds us:

> The consciousness of myself in the representation "I" is not an intuition, but a merely *intellectual* representation of the spontaneity of a thinking subject. This "I" has not, therefore, the least predicate of intuition (B278).

> [T]he synthesis of the conditions of a thought in general…..is not objective at all, but merely a synthesis of the thought with the subject which is mistaken for a synthetic representation of an object (A397).

> The "I" is indeed in all thoughts but there is not in this representation the least trace of intuition, distinguishing the "I" from other objects of intuition. Thus we can indeed perceive that this representation is invariably present in all thought,

> but not that it is an abiding and continuous intuition, wherein the thoughts, as being transitory, give place to one another (A350).

Moreover, crucially, from a Kantian perspective, there is an intimate reciprocal relationship between phenomenal consciousness and the intelligent thought and activity of human beings. The human capacity for conceptual understanding or thought is so inextricably bound up with the very capacity for phenomenal consciousness, that each necessarily presupposes the other. For Kant the original consciousness of the identity of the self *is at the same time* a consciousness of the world. Conversely, the same consciousness which reveals the world also reveals the subject. The self becomes conscious of itself by seeking out and bringing into consciousness its own contents. That is to say, by synthesising the manifold of spatial conceptions we gain knowledge of things outside us and it is through this knowledge of outer things that the self can know itself. Kant writes:

> I am just as certainly conscious that there are things outside me which are in relation to my senses, as I am conscious that I myself exist as determined in time (Bxii, note).

Thus, for Kant, the world both requires and guarantees the subject. Moreover, his methodological idealism states that human cognitive subjectivity and empirical objectivity are mutually interdependent for their very intelligibility. This is what Jay F. Rosenberg calls Kant's mutuality thesis:

> [T]he conditions according to which an experienced world are constituted as an intelligible synthetic unity were [on Kant's view] at the same time the conditions by which an experiencing consciousness was itself constituted as a unitary self. That an experience represents the encountered world as catagorially structured in space and time, Kant claimed to show, was a condition of the very possibility of his representing *himself* as a unitary subject of his experiences of that world, or indeed any world at all! At the centre of Kant's critical philosophy, then lies a thesis of self and world...subject and world are two inseparable poles of a single dynamic process of representation (Rosenberg, 1986, 2008, p. 6).

Rosenberg argues that three levels or grades of apperception: situatedness in time and space, multi-positionality, and "x- objectivity", i.e. the impersonal representation of the world that selves gain when they adjust their personal representations to harmonise with reported representations of other selves. Furthermore, these three grades of apperception guarantee that one can distinguish oneself from others. For

Rosenberg, a being who is capable of apperceptive consciousness just *is* one that possesses a conceptual system rich enough to afford a global representation of the world, through three levels or grades i) situatedness in time and space, ii) pure positional awareness (he uses the example of a cat stalking a quail which, although object-directed is not an apperceptive, self-conscious kind of consciousness, but non-conceptual) and iii) since, conceptual capacities are, for Rosenberg, connected to linguistic development, also part of a linguistic community. For Rosenberg, non-conceptual representations of objects are an integral part of both human and animal perceptual experience. Our reflective consciousness is grounded in the non-reflective consciousness we share with other animals. According to Rosenberg "our own form of self consciousness is an *elaboration*" of the cat's "pure positional awareness" (ibid., p. 103). Conceptual representation rests on a "fundamentum of non-conceptual representation" - the cat's pure positional awareness "has the structure of a perceptual field partitioned into figure and ground". This is not simply a fact about the cat's "inner states" but can also be construed as involving embodiment and situatedness-in-a-world. In fact, Rosenberg offers a plausible account of the sense in which non-conceptual animal cognition in humans and other animals succeeds in representing and tracking a world of objects in space and time. (This notion of embodiment and situatedness-in-the-world will be discussed further in Chapter 6).

According to Rosenberg's mutuality thesis, then, an objective "synthetic unity of experience" is correlative to the subjective "transcendental unity of apperception" (ibid., pp. 6-7). Synthesis or "the combination of the manifold of sensory experience" produces not only the objective world, but at the same time analyses out the experiencing subject. Thus, consciousness of self and consciousness of the world are born together through the spontaneous activity of synthesis. The output of our synthesising activities implies both a concrete empirical subject in time and space and the concrete world of objects of experience. The empirical subject and empirical object are thus reciprocally generated through the transcendental *actus* or act of synthesis, in the sense that even though we do not introspect such a continuing subject of thought, the existence of a thinker of thoughts is inferred from the very mental states with which we are presented. This leads us nicely into the topic of the next section.

3.3. Synthesis, Relational Unity, and Consciousness of Self.

As discussed, Descartes had claimed that although he could doubt everything in the world, the one thing that he could be sure of was that the "I" that thinks should be something. Although everything that he is aware of in the external world is a dream, or a deception resulting from the work of an evil genius, he cannot doubt that he, himself, exists as a "thinking thing", a *res cogitans.* As he declares "whilst I thus wished to think all things false it was absolutely essential that the "I" who thought should be somewhat" (*Discourse on Method, Part IV*). Of note here is that a key feature of his position is that the self, insofar as it encounters the world, is itself unencounterable through experience. Being thus unencounterable then it cannot be encountered falsely. Hence the immunity of the *Cogito* from the hyperbolic doubt that afflicts all experience. From this indubitable truth Descartes then goes on to derive a number of *a priori* truths about the self, viz: that it is simple, substantial, unitary and persistent. This notion of the encounterability of the self in experience is reaffirmed by Leibniz. He agrees with Descartes that the "I" is able to be apprehended by the mind directly and not through sensation. It can be known through what he termed "acts of reflection" which are akin to Descartes' *a priori* reasoning and through which we are able to know that the self is simple and substantial (*Monadology* 30). However, for Leibniz there is also knowledge of a self gathered from one's experiences, for he adds to this that when we consider the constitution of the mind there is nothing besides "perceptions and their changes" (*Monadology* 17). In other words, Leibniz introduces a relation between the "I" that encounters and that which it encounters. He writes "It is well to make a distinction between perception, which is the internal condition of the monad representing external things, and apperception, which is consciousness or the reflective knowledge of this internal state (Leibniz, *The Principles of Nature and Grace*, sec. 4). He makes the same distinction almost word for word in the Monadology (14). It is interesting that he makes this distinction between the "I" that encounters and that which it encounters, and that he terms such awareness, respectively, "apperception" and "perception". This terminology is later taken up by Kant. Consideration of this distinction leads Leibniz to conclude that if over time minds have no distinct perceptions then they are "monads which are wholly bare" (*Monadology* 24). For

Leibniz, then, the "I" without distinct perceptions is a "bare" mind. In a certain sense, it is "knowable" only insofar as it encounters the world; there is an emptiness in the idea of the self considered apart from its role as experiencing subject. Hume takes this notion further and propels it to its natural empirical conclusions. He argues that the self is not only unencounterable in the world, as the rationalists claim, it is also unencounterable introspectively, not even as a "bare self". He writes:

> For my part when I enter most immediately into what I call *myself*, I always stumble upon some perception or other…. I never catch *myself* at any time without a perception, and never can observe anything but the perception (*A Treatise on Human Nature p. 252).*

Hume sees it the self is a nothing but a mere "bundle or collection of different perceptions which succeed each other with an inconceivable rapidity". The unity of the *Cogito* is an illusion. There is no diachronic unity, (unity over time) for the self is not a continuant subject but only a sequence of representations. Neither is there synchronic unity (unity at a time) but a complex of ontologically more basic entities with no necessary connection between them. As discussed, Kant can be viewed as adjudicating the claims of his predecessors, synthesising their views and passing beyond them; he also takes some inspiration from Liebniz. He agrees with Hume about the unobservable character of the "I". Empirical self-consciousness or "inner sense" presents us with no "abiding" self but a complex of representations only. In fact, he points this out in terms reminiscent of Hume's own discussions on the matter, that:

> Consciousness of self according to the determinations of our state in inner perception is merely empirical, and always changing. No fixed and abiding self can present itself in this flux of inner appearances (A107).

However, that the unity of consciousness cannot be determined by anything in the contents of consciousness is not the whole story. Kant thought that it must be considered as something more than this; the self does exist and that we are, in some sense, "conscious" of it. As he writes:

> I am *conscious* of the self as identical in respect of the manifold of representations given to me in an intuition, because I call them one and all *my* representations (B135) [my italics].

As Rosenberg points out, Kant regards the unity of consciousness at a time as *necessarily* involving self-consciousness, or at least as involving the potential ascription of contents to itself by the subject:

> The "I think" must be capable of accompanying all my representations for otherwise something would be represented in me which could not be thought, in other words, the representation would either be impossible or else would be nothing to me……The unity of this apperception I call the transcendental unity of self-consciousness, in order to indicate the possibility of *a priori* knowledge arising from it. For the manifold representations which are given in an intuition would not all of them be my representations if they did not all belong to one self-consciousness (B132-3).

The "I" of "I think" is a problematic concept, given the philosophical presuppositions and constraints of Kant's predecessors. Since the self is never encountered in experience yet is always present in every experience, what are we to make of it? I am aware of myself as "something" which has my thoughts but in what way am I so aware and what account are we to give of this "something"? Kant, in fact, recognises that the views of his predecessors present us with an antimony which is both instructive and illuminating. Descartes' conception of the self is that of a unitary simple substance, Hume's that the self that experiences is nothing over and above a set of those experiences. Kant accepts the Cartesian concept of the self, but only as a "form of representation" that is necessary for us given the character of our experiences. However, he also agrees with Hume that the knowledge of the actual character of the "I" is inaccessible to us. However, this is not because it does not in some sense "exist" as Hume claims, but because the "I" is not an intuition, and since our knowledge is restricted to intuitions "thought through the understanding" (A19/ B33) nothing can be *known* about it in the strict sense, although there is cognitive access of a different kind. From this it follows that we cannot know the self to be substantial in the way that Descartes claims. On the Cartesian picture awareness of oneself is analogous to awareness of any other particular. This is evident from the way that Descartes describes the mind, in negatives of the descriptions given to objects, viz., as non-spatial, non-observable and non-material. For him mind and body are both "substances", both "things" with attributes and properties. However,

Kant claims that to assert that the soul is substance is to make "an empirical… but, in this case inadmissible, employment of the category" (A403).

Gilbert Ryle in his famous 1949 book *The Concept of Mind* is regarded by many thinkers as having eliminated the immaterial mind and dissolved the mind-body problem. In it he also accused Descartes of error, and in particular, of making a category mistake - the conceptual fallacy of putting mind in the same category as body. He called Cartesian Dualism "the official doctrine" and claimed that it represented the facts of mental life "as if they belonged to one logical type or category (or range of types or categories), when they actually belong to another". Ryle alleged that it was a mistake to treat the mind as an object made of an immaterial substance because predications of substance are not meaningful for a collection of dispositions and capacities. He also termed it the Dogma of the Ghost in the Machine. There is therefore *prima facie* similarity between what Ryle later called a category mistake and what Kant calls the fallacy of "subreption" (A402). The difference is that Kant has deep epistemological reasons for this, whereas for Ryle it is simply a mistaken language application. Kant's great insight was his recognition that one gains awareness of oneself as subject, in a different manner from the way in which one gains awareness of one's own psychological states. Rather than appealing to "clear and distinct ideas" and the introspective awareness of a simple unitary "substance", as had Descartes ("substance" for Descartes is a thing that does not depend on anything else for its existence, i.e. it is a self- subsisting thing or entity), Kant regards spontaneous synthesis as the source of this unity. His point is that it is a kind of category mistake to claim that the subject is "knowable", by means of a kind of rational intuition, in a way analogous to the way that we "know" objects as Descartes insists. For Kant the subject should not be considered as substantial but as a "formal" unity preceding experience as its necessary condition and capable of knowing its identity through its spontaneous acts.

The transcendental unity of apperception is distinguished from empirical apperception in order to account for the possibility of subjective experience which must display a relational unity between the "I" that experiences and what is experienced. The claim that subjective experience must display a *relational* unity (rather than simply Hume's "bundle of perceptions" held together by the principles of association) is justified because the use of the *a priori* categories of the

understanding is shown to be necessary. Hume's "principles of association" are too impoverished and too subjective as an account of the object of experience to be capable of explaining its unity. The difficulty with it is that whilst it purports to explain why it is we might take various qualities to constitute one object, it is insufficient to show how these qualities are related together *in the object*, but only how perceptions are connected together in our minds; in addition, it fails to explain why certain qualities are *invariably* found to be united together in our experience of the object. Kant's solution is that if consciousness is to be unified in such a way as to mean anything to me, i.e. to be a single complex thought, then it must be possible for a subject to be aware of the "I" that thinks each of the elements of an object (in general) together with the I that thinks them all together. I must be able to think of each element, and each of them as mine. In the manifold of both synchronic and diachronic representations, consciousness of each element is inseparable from consciousness of each as belonging to me. Thus, the subject must be able to be aware of each of the elements in the contents of consciousness, both diachronic and synchronic as belonging to him. This, however, requires an awareness of the act of combination or synthesis by means of which the elements are thought together:

> I am conscious of the self as identical in respect of all the manifold of representations that are given to me in an intuition, because I call them one and all *my* representations, and so apprehend them as constituting *one* intuition. This amounts to saying that I am conscious to myself *a priori* of a necessary synthesis of representations - to be entitled the original synthetic unity of apperception - under all representations that are given to me must stand, but under which they have also first to be brought by means of synthesis (B135-136).

Thus, the transcendental unity of apperception involves a kind of awareness, i.e. consciousness of oneself as active in synthesis, which is different from consciousness of oneself as empirical in inner sense. Self-consciousness as inner sense or our awareness of what we passively "undergo" as we are affected by the play of our senses is distinguished from consciousness of "what we are doing" i.e. spontaneously synthesising. For Kant the self of apperception involves the notion of the existence of a unitary subject. He writes in the Paralogisms "The proposition "I think" insofar as it amounts to the assertion "I exist thinking" determines the subject (B429), and that the "I think" is "something real that has been given (…) something

that actually exists" (B423n). In the Transcendental Deduction he states "my existence is not mere "appearance" (much less mere illusion) (B157). However, he then qualifies this in a footnote:

> The "I think" expresses the act of determining my existence. Existence is already given thereby, but the *mode* in which I am to determine this existence (...) is not thereby given (B157n) [my italics].

Apperception does involve the "real" existence of a unitary self but does not present that existence in "determinate form". For this, representations of objects *other* than oneself are required. We are conscious of our existence as active agency. This existence is not existence as a category. The self whose existence is indicated by the "I think" is not known in the way that Descartes claims because it does not involve an intuition. If one were to know oneself in this way then one would need a special intuition of oneself "which gives the determining in me (...) prior to the act of determination, as time does in the case of the determinable" (B157n). However, no such intuition is forthcoming and one must conclude that one does not have knowledge of oneself as substantial in the way that Descartes insists.
Kant makes a similar point in the first Paralogism:

> The 'I' is indeed in all thoughts, but there is not in this representation the least trace of intuition, distinguishing the 'I' from other objects of intuition. Thus we can indeed perceive that this representation is invariably present in all thought, but not that it is an abiding and continuing intuition, wherein the thoughts, as being transitory, give place to one another (A350).

And he concludes: "We have no knowledge of the subject in itself, which as substratum underlies this 'I' as it does all thoughts". We cannot perceive the "I" because it is not an intuition (intuitions are representations of empirical objects as indeterminate appearances). Since this is the case we cannot ascribe to it the permanence and simplicity which is constitutive of substance. If we do so we are making an empirical but illegitimate use of the *a priori* category of substance. The proposition that the self is substance implies that the mind is representable as empirical intuition. If this were the case this would mean that one must be able to pick it out by means of "predicates of its intuition" (A399-400). However, this is

exactly what we cannot do with the "I" of "I think". The consciousness of myself in the representation "I" is not an intuition, but the merely intellectual representation of the spontaneity of a thinking subject. This "I" has not therefore the least predicate of intuition (B278). Thus, there is no representation of the mind as it is by inner sense. In this Kant agrees with Hume, as discussed earlier. At most introspection gives us the phenomenal self which is appearance only. If we take away from this content we are left with a "bare representation". Nevertheless, Kant writes "I am conscious of my own existence as determined in time" (B 275) and that "I exist as an intelligence which is conscious solely of its power of combination" (B158-9). I take it that for Kant the *fact* of self-consciousness is in itself sufficient to establish the existence of a self. As he says "My existence is (...) already given by the act of consciousness" (B157n). For, without the active agent of synthesis, neither knowledge nor experience would be possible at all. We have no right to claim that we *know* the subject as it is in itself, however. As he puts it "It is (...) very evident that I cannot know as an object that which I must presuppose to know any object"(A402). Being aware of oneself as subject through acts of synthesis is very different from awareness of oneself as object. Kant asserts that "consciousness" of oneself is very different from "knowledge" of oneself (B157). All we *know* of the self is what is formally identical in our various acts of synthesis. Abstracting from these acts we are left with bare consciousness, no content at all. For Kant unity of consciousness depends on synthesis, this consciousness is not akin to perceptual experience. We are, Kant says, "conscious" of the existence of the apperceptive self in the spontaneous activity of synthesis. He writes "Synthesis (...) as an act is (...) conscious to itself, even without sensibility" (B153). One is aware of oneself as "subject" by doing acts of synthesis not via intuitions. In fact, it is on this self-consciousness that the unity of the contents of consciousness necessarily depends. The clearest statement of this comes at B133-4:

> [The] thoroughgoing identity of the apperception of a manifold which is given in intuition contains a synthesis of representations, and is possible only through the consciousness of this synthesis. For the empirical consciousness, which accompanies different representations, is in itself diverse and without relation to the identity of the subject. That relation comes about, not simply through my accompanying each representation with consciousness, but only in so far as I conjoin one representation with another, and am conscious of the synthesis of them.

3.4. The Transcendental Subject.

Kant proposes that there are certain principles given *a priori* on the basis of which we acquire knowledge of all possible objects of perception. Knowledge must, however, be confined to what is given in experience. Any attempt to introduce into thinking what lies "beyond sensation", that is, metaphysical entities such as "value" and "freedom" necessarily lead to antinomies and error. Kant reasoned, against his predecessor, Hume, that something (an "x") exists that unifies our discrete sense impressions, an enduring and coherent subject that is a necessary precondition of experience. Kant called this unifying self the "transcendental subject". But at the same time he stressed that although we can have reasonable faith in the capacity of the transcendental subject to order our experience, we still can never apprehend it directly. It is always beyond the reach of our empirical capacity to "know". In the course of addressing Hume's scepticism in his *Critiques*, Kant distinguishes the phenomenal aspect of empirically knowable things (the world measurable by Newtonian physics) from the noumenal aspect of things-in-themselves. He claimed that although we can "know" a lesser faculty of our minds, the "empirical self", through introspection or inner sense, the transcendental subject is beyond this, thus:

> Through this I or he or it (the thing) which thinks, nothing further is represented than a transcendental subject of the thoughts = *X*. It is known only through the thoughts which are its predicates, and of it, apart from them, we cannot have any concept whatsoever (A346/B404).

It is Kant's aim in the Transcendental Deduction to explain what is meant by this "transcendental subject" of thought. What he says is that although it cannot have empirical predicates applied to it, it can have "transcendental predicates" (B114), that is to say, all we can say about it concern the limits and conditions of validity of knowledge of the most universal kind. Nothing whatsoever of an empirical nature can be predicated of it, and this is against both the Cartesian and Empiricist notions of the subject, which model such knowledge on sense perception. For Kant, the *grund* or ground of all experience is the experiencing subject considered transcendentally, and the constitution of the subject is such that all thought is rule-governed in accordance with the categories of the understanding.

As has been emphasised, it is significant that Kant also argues for a deep interconnectedness between the ability to have self-consciousness and the ability to experience a world of objects. It is through the *a priori* process of synthesis, that the mind spontaneously generates both the structure of objects and its own unity. The "I think" is an act of spontaneity, or pure apperception, not a state of passive empirical observation or empirical apperception. In the *Anthropology* Kant discusses the role of apperception in relation to the *Gemut* (inner sense) and makes it clear that we are to differentiate inner sense from pure apperception, and judges "psychology" on the basis of its failure to apply this distinction. This is articulated in a long footnote in the first book in the Anthropology:

> If we consciously imagine for ourselves the inner action (spontaneity), whereby a concept (a thought) becomes possible, we engage in reflection; if we consciously imagine for ourselves the susceptibility (receptivity), whereby a perception (*perceptio*), i.e. empirical observation, becomes possible, we engage in apprehension; however, if we consciously imagine both acts, then the consciousness of one's self (*apperceptio*) can be divided into that of reflection and that of apprehension. Reflection is a consciousness of the understanding, whilst apprehension is a consciousness of the inner sense; reflection is pure apperception, but apprehension is empirical apperception; consequently, the former is falsely referred to as the inner sense. In psychology we investigate ourselves according to our perceptions of the inner sense; but in logic we make the investigation on the grounds of what the intellectual consciousness supplies us with. Here the self appears to us as twofold (which would be contradictory): (1) the self, as the subject of thinking (in logic), which means pure apperception (the merely reflecting self) of which nothing more can be said, except that it is entirely simple perception. (2) The self, as the object of the perception, consequently also part of the inner sense, contains a multiplicity of definitions which make inner experience possible. To ask whether or not a man conscious of different inner mental changes (either of his thoughts or of fundamental principles assumed by him) can say that he is the selfsame man, is an absurd question. For he can be conscious of these changes in the first place only on condition that he represents himself, in the different situations, as one and the same subject. The human ego is indeed twofold as regards its form (manner of representation), but not with respect to its matter (content). (An Ak. I. V. pp. 17-18, footnote).

Thus, Kant illustrates that awareness of a particular conscious experience includes a tacit awareness of oneself as subject of that experience, a "bare

representation" x . This means that a subject, in perceiving Q is also tacitly aware of itself perceiving Q. In other words all conscious states involve the possibility of *self*-awareness, though this is not awareness of a "self" in the substantial sense, as a thinking thing or Cartesian *res cogitans*: the "self" is not an entity or *quasi* entity that exists apart from, or above, the phenomenal experience. Descartes had declared, famously, "I think, therefore I am" *cogito ergo sum* but for Kant, such knowledge of a self is impossible. The "I" is simply a "logical" requirement of the "unity of apperception" and lacks the experience of direct intuition that would make such self-knowledge available. Although "I" seems to refer to the subject of experience, it is not really a permanent feature but simply the formal characteristic of a unified consciousness. For Kant we have a *sense* of self, which forms an integral and ubiquitous part of our experiential life. This is the "transcendental unity of apperception" which is at the same time the "synthetic unity of apperception" that unites sensory experience. Indeed, there is no experiential dimension whatever without this sense of self. It is useful here to consider the following remark from Wittgenstein in the *Tractatus Logico-Philosophics,* where he discusses this sense of self on analogy with the eye and the visual field: "[T]his case is altogether like that of the eye and the field of sight. But you do not really see the eye. And from nothing in the field of sight can it be concluded that it is seen from an eye (*Tracatus* 5.633).

A notable advantage of Kant's view, then, is that it can deflect the various difficulties of theories of consciousness that are entangled in the conceptual confusions of the "object" consciousness paradigm, requiring a complex theory in order account for the so-called "explanatory gap" between mind and matter. If conscious states are not construed as kinds of secondary "objects" relative to a conscious mind; and conversely, if self-awareness is an intrinsic feature of those states that possess it, there is no need for complicated theoretical tactics to account for the intuitive immediacy of those conscious states. It also prevents infinite regress. If consciousness of a thought is supposedly different from the thought, then an infinite regress is generated, for if I have the thought of *x* then I must be conscious of *x*, but this consciousness of *x* is also a thought, requiring a further thought, that I am aware of *x*, and so on *ad infinitum*. However, Kant's view avoids this problem of infinite regress, since there is no Cartesian homunculus, the "little man" in the head

observing mental states in the theatre of the mind. Rather, Kant claims that consciousness is best understood as involving a non-objectifying self-givenness which is the ground of all experience of the world. That is, an object of phenomenal experience is not simply a "this," but rather, a "this-such-for-me," by which is meant an intuition which has already been conceptualised. And understanding something (it becomes an object for me) is only possible because I classify it in some way, through the pure categories of the understanding by which I make a judgment that brings the intuition under the concept. According to the Transcendental Deduction, objects can only become objects for us insofar as they conform to the conceptual structures with which we cognise objects. As Kant remarks:

> Thoughts without content are empty, intuitions without concepts are blind. It is, therefore, just as necessary to make our concepts sensible, that is, to add the object to them in intuition, as to make our intuitions intelligible, that is, to bring them under concepts (A51/B75).

Thus, Kant thought that human self-consciousness can be regarded from two perspectives, as the logical subject of thought and as the object of inner sense, the empirical self. How can this dual unity of consciousness be brought out in full? Transcendental apperception, the transcendental unity of self-consciousness is the fundamental condition for the cognition of objects in the phenomenal world. The fact that we can make judgments at all presupposes this unity of consciousness in a subject who is synthesising or combining representations according to the categories. Empirical apperception, on the other hand, refers to consciousness of the particular contents of the subject's own mental states. When we refer to the self "transcendentally", such consciousness is not of a kind of intelligible object and Kant admits that it is difficult to describe exactly what it is. The kind of understanding to which it could belong is itself a problem. But it is "a kind of understanding", i.e. consciousness of oneself not as an object but of oneself through acts of synthesis. It is difficult to characterise this special kind of self-awareness. One way of characterising it is to say that when one is aware of oneself this way, one is aware of one's being, or one's mind "as it is", by being a spontaneous synthesising agent, and not by being intuitively aware of contents of a representation for such, that may or may not present me as I am. Another way of putting it is to say that it is

a “mode of being” in a contentful state which is known *instrinsically* through reflective judgement. Kant’s point is that we must be able to have some kind of cognitive access to ourselves as a condition of our ability to perceive objects, i.e. as a condition of our capacity to “apply the categories to the manifold of experience”, which is independent of and logically prior to such ability.

Put in a nutshell, Kant’s claim is that a unity of consciousness is possible only on condition that we have a consciousness of unity, and *vice versa*. His argument for our being able to conceive of ourselves as unities of consciousness, that is, as temporally determined conscious agents, is based on our being able to discern and distinguish a real world of moving objects through the application of *a priori* concepts which order and unify our perceptual input. According to Kant, we are logical subjects of thoughts, “transcendental unities of apperception” that are “logically” necessary for the very possibility of coherent cognition. We look for the self, we reflect, and we try to find something there as the locus of thought. We may conjure it up in the concept of a mental “thing” (Descartes), or a “bundle of perceptions” (Hume). But we would be mistaken, because we are looking in the wrong direction. Self-consciousness requires the existence of a perceiving and conceiving being that acts and interacts with other objects and the environment in an objective world. The self is not the mind: the self is active agency within the world. Descartes had declared, famously, “I think, therefore I am” *cogito ergo sum*, but for Kant, such reasoning is faulty. For him the “I” is simply a “logical” requirement of the “unity of apperception” and lacks the experience of direct intuition that would make such self-knowledge available. Although “I” seems to refer to the subject of experience, it is not really a permanent feature but simply the logical characteristic of a unified consciousness, the “transcendental unity of apperception”.

For Kant the transcendental unity of apperception, which involves synthesis, is an *actus*. It is something we do. It is by means of the act of synthesis, the combining of the raw data of experience, that the unity of apperception is simultaneously made manifest. The concept of an “object in general” signifies the generic product of this synthesis (which takes place according to the categories). The manifold given in intuition has categories applied to it, rendering it possible to think of what is given in that manifold as an object (something given to a subject). The transcendental unity of

apperception is not a thing, however rarified; rather, it is the "formal" condition of synthesising a manifold into a cognition of an object. Such a synthesis could not take place unless done within a single consciousness, which requires a special kind of unity. The self, in other words, is not an object of which one can become conscious like other objects, and we can only become conscious of it as that which unites all the representations. Kant adds that we can only think of a number of representations as belonging to the same self in this way through being conscious of, or seeing them as, an act in which we synthesise (or combine) these representations with each other.

Kant puts it another way, to *apperceive* a representation is roughly to see it as mine, or to say to myself of it "I am having this representation" or "I am thinking this" (thus Kant refers to the act of apperception as the "I think"). The *analytic unity of apperception* is the unity or united character that a number of my representations have through my seeing them as all belonging to one and the same "I". The synthetic unity of apperception (or *original-synthetic unity of apperception*) is the unity that they have through my being conscious of an act of synthesising or combining them with each other, which is a necessary condition of the analytic unity of apperception. So, for Kant the pure concepts of the understanding (of which all our everyday concepts are specifications or combinations) are grounded in self-consciousness, which in turn is grounded in an original act of synthesis. Here is how he summarises his view:

> [A]n *object* is that in the concept of which the manifold of a given intuition is *united*. Now all unification of representations demands unity of consciousness in the synthesis of them. Consequently it is the unity of consciousness that alone constitutes the relation of representations to an object, and therefore their objective validity and the fact that they are modes of knowledge; and upon it therefore rests the very possibility of the understanding (B137).

On the Cartesian picture subjectivity is somehow inner, independent, private and unsupported; conscious states, thoughts and experiences, can be the whole of what is real, without requiring the reality of anything else, and in particular without requiring the reality of the *objects* of our thoughts and experiences. Descartes had to postulate a benign God in order to guarantee the objectivity of the external world. For Kant, however, the first-person point of view does not alone determine a domain of objects of awareness. This is because subjectivity is impossible without objectivity; the one presupposes the other. For it is a necessary condition of our

having thoughts and experiences at all, that the objects of those thoughts and experiences have a certain character. Understanding something (it becomes an object for me) is only possible because it is classified in some way, by an act of judgment which brings the intuition under the concept. In the Transcendental Deduction, Kant declared that objects can only become objects for a subject insofar as they conform to the conceptual structures with which we cognise them. At the risk of belabouring the point, he writes, we must "make our intuitions intelligible, that is, to bring them under concepts" (A51/B75). Moreover, to *apperceive* is to recognise the experience as mine, it is something "for me"; in other words, there is a "for me" factor that necessarily accompanies all our awareness of the world. What makes this for-me factor possible is a kind of *action* that is performed, *viz*., the act of *judgement*, and this is a matter of synthesising, or bringing together, the manifold of intuition (i.e. the totality of the separate elements of sensation) through certain logical operations, i.e. the categories. In judgement, a *unity* (a sense, or a meaning) is created, i.e. the (synthetic) unity of consciousness itself. As Kemp Smith noted in his *Commentary to Kant's Critique of Pure Reason*, published in 1918:

> Kant maintains that human consciousness is always an awareness of meaning, and that consequently it can find expression only in judgments which involve together with their other factors the element of recognition or self-consciousness (Kemp Smith, 1918, pp. 47-8.)

To judge is also to invoke both the productive and reproductive functions of the imagination. [44] Without this, nothing can mean anything to a subject, there can be no representation of anything, no intentional object for consciousness. Spontaneity is an act of the imagination described in both editions as "a blind but indispensible function of the soul, without which we should have no knowledge whatsoever, but of which we are scarcely ever conscious" (A77-78/B104); this is not because it is *un*conscious, but rather, as the engine of synthesis, it is the very ground or *grund* of such consciousness. Transcendental imagination is what Kant refers to as the unknown common root uniting sense and understanding. It seems clear that imagination is viewed by Kant as pre-conceptual and non-intellectual, especially in the A edition of the *Critique*, where it has the important function of "mediating" between sensibility and understanding (A124). The proper domain of the

imagination is a contentious issue in Kantian scholarship. But it is being increasingly recognised that it is related to the body - specifically that Kant's categories of experience are derived from the architectonic structure of embodiment. Although imagination's use in the Transcendental Deduction is not to do with physiological bodily activity *per se*, what it achieves is a remarkable bodily understanding of the world. In fact, recent scholarship has seen the development of several accounts of Kant's theory of the imagination as embodied imagination.[45] This will be the focus of Chapter 6.

The following section is a consideration of the problem of indexical self-reference, the idea that Kant may have been the first to describe the peculiar logical semantics of self-reference, and how this philosophical analysis of the nature of reference to self, has profound contemporary relevance. It is suggested that taking advantage of it might have helped avoid some of the problems that have beset contemporary theories of consciousness that have emerged in the wake of and in reaction to functionalism.

3.5. The Peculiar Logical Semantics of Self-reference

As discussed above, mainstream cognitive science has retained a large part of its Cartesian and empiricist legacy in the sense that they take as the root notion the distinction between the mental and the physical, the mind and the world, the subject and the object. This is precisely what leads to the notion that cognition must be representational and must involve the notion of intentionality or aboutness. It is the view that the mind cannot reach to the objects themselves, and that it is therefore necessary to introduce some kind of representational medium or interface between mind and world if we are to explain intentionality. (Later, it will be argued that Kant's ideas can be seen to challenge the thesis of this functionalist mechanistic-symbolic intentionality). Moreover, a neuroscientific theory of consciousness is always a theory of the *empirical* subject of consciousness, one that analyses a kind of "central executive" (to borrow a phrase from Daniel Dennett) implied in the unity of experience into its constituent parts, none of which can itself be a proper subject. As has been discussed thus far, this omits the Kantian insight of "transcendental

apperception" and the spontaneous agent of synthesis. Kant's insight is that a unity of consciousness is only possible on condition of consciousness of a unity and vice versa. That is, the possibility of our being able to understand ourselves as temporally determined conscious agents is founded on the ability to discern a real world through the application of the concepts or categories which order and unify perceptual experience. Furthermore, for Kant, the empirical must be differentiated from the transcendental in order to guarantee the objectivity of our world (and also of the moral law, but that is another topic) as opposed to the merely empirical and contingent, i.e. that which remains bound to sense experience. Kant argues for a deep interconnection between the ability to be self-consciousness and the ability to experience a world of objects. One of the central theses of Kant's critical philosophy is that the conditions that make the unity of objective experience possible also make the unity of apperception (self-consciousness) possible. As was discussed earlier, Rosenberg terms this Kant's "mutuality thesis": through a process of spontaneous synthesis, the mind generates both the structure of objects and its own unity. This is an important insight; all consciousness of an objective world involves a certain type of self-consciousness, the awareness of the self as neither a *quasi*-object (Descartes) nor as one's mental states (Hume), but rather, as the intellectual thought of the identity of the self through time.

On the Kantian account of the infallible self-ascription of judgements, judging is a matter of consciously combining or synthesising some mental states in others and of recognising that combination as such, and it is also to be aware of the unity of these different states in a single subject. He demonstrates how we are apt to easily confuse this transcendental unity of our perceptions and thoughts with the perception of a unity. Although he was writing over two hundred years ago, Kant anticipated some of the most important ideas about the mind of the last forty or so years. For he addressed the sense in which a person be said to be a subject of his own awareness, when as Hume put it, "we enter most intimately into ourselves". As noted, Hume's denial that there is introspective awareness of a subject was based on the model of sense-perception. Thus, he argued that although we are aware of perceptions of objects, we are not aware of a perception of a self amongst them. This method of characterising self-awareness has been carried on and repeated ever since by philosophers and contemporary cognitive scientists up to this very day, who

implicitly support the view that introspection is to be conceived on the model of perception and unwittingly accept the Humean model that the self is not among the objects perceived. For instance, Elizabeth Anscombe had taken it to support the view that the "I" does not refer (Anscombe, 1975, p 58). Bertrand Russell had claimed that the self is not "empirically discoverable" (1921, p. 5) and Carnap (1928) that "the given is subjectless" (2003 [1928], pp. 103-6). Thus, Hume's denial that we can have perceptual knowledge of a self, that we perceive a self by what Kant termed "inner sense" has been influential.

Kant agrees with Hume that whatever we are aware of when we introspect it is always in the form of some content or another. But he makes the point that the notion of content assumes or presupposes the existence of an "x", that is conscious of it. Hume looks for a self among the contents of inner sense and finds no evidence of one. For Kant, however, self-awareness is a spontaneous act which is independent of all evidence. There is a distinction between the self as it is apart from our modes of apprehension and the self as it appears to those modes, the empirical self. The empirical self like any other object is apprehended successively in a "manifold of representations". The self that apprehends the manifold, however, must be apprehended some other way. If this were not the case we would be unable to identify any "manifold" as belonging to our experience. But I must be able to be aware of the fact that what I combine in a succession of representations belongs to one and the same self:

> Only in so far, therefore, as I can unite a manifold of given representations in *one consciousness*, is it possible for me to represent to myself *the identity of the consciousness in [i.e. throughout]* these representations (B133)

Kant also describes self-awareness in a different way, what we might call the self-ascriptive sense of self-awareness, the sense of self-awareness in which I *ascribe* a set or series of perceptions to myself. "Only in so far as I can grasp the manifold of the representations in one consciousness do I call them one and all *mine"* (B134). I must have the ability to identify the mental contents I do have as belonging to my own mental history. But the self that is temporally continuous and which lies at the foundation of our ability to synthesise any manifold cannot itself be an object of intuition. This requires a self that is numerically one throughout time yet is not perceivable. All we perceive is that certain contents belong to our mental histories

but we are not acquainted with the self to which these contents belong. Objects are known by applying the categories to a manifold of intuition. In order to know myself as object I must be able to apply the categories to a manifold. However, since the self "in itself" is the ground of the possibility of the categories it cannot know itself through them, but "knows the categories, and through them all objects, in the absolute unity of apperception, and so through itself" (A401). Kant writes "I cannot know as an object that which I must presuppose in order to know any object" (A402). By this he means that whatever I must use to know something as an object cannot, without circularity, be an object for me. The claim that the self is an object of consciousness is circular since it appeals to the application of the very characteristics of what we want to know as conditions of our knowing it. However, this does not mean that we do not have cognitive access to ourselves at all. The categories are presupposed in us knowing perceptual objects but they are not presuppositions of every act of awareness. Rather, Kant's point is that we must be able to have some kind of cognitive access to ourselves as a condition of our ability to perceive objects, (i.e. as a condition of our capacity to apply the categories to the manifold of experience) which is independent of and logically prior to such ability. Traditional cognitive science has little say about this kind of consciousness, focusing instead on consciousness of psychological states.

Let us recapitulate to get our bearings: Kant's work can be seen as an attempt to reconcile the claims of reason and experience in order to do justice to both rationalist and empiricist accounts of self knowledge. He therefore elaborates a dualistic theory of the nature of consciousness. He agrees with Hume that we have a kind of knowledge of ourselves through inner sense. But he also agrees with Descartes that this does not end the matter. For as well as the empirical self we must recognise the self as subject. However, this is not a Cartesian subject that exists in the world as a kind of conscious object, but the "transcendental subject". This is not a sense of self occasioned by experience but the ultimate condition of *all* experience. As he puts it in the *Anthropology*:

> Inner sense is not pure apperception, consciousness of what we are doing, for this belongs to the power of thinking. It is, rather, consciousness of what we undergo as we are affected by the play of our own thoughts (An Ak. VII p. 161).

In the first *Critique* he writes:

> [A]s regards inner sense, that by means of it we intuit ourselves only as we are inwardly affected *by ourselves*; in other words, that, so far as inner intuition is concerned, we know our own subject only as appearance, not as it is in itself. On the other hand, in the transcendental synthesis of the manifold of representations in general, and therefore in the synthetic original unity of apperception, I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am (B156-7).

The difficulties of Cartesian and Humean pictures are the result of confusion between the self as encountered empirically or *quasi*-empirically and the self of transcendental apperception. The belief that the numerical identity of the "I" is determined by something analogous to perception is an illusion caused by the confusion between the "I" of inner sense (in which I am aware of mental states (Hume's bundle) or an immaterial substance (Descartes) and the "I" of transcendental apperception. We do not determine our own identity by some kind of inner perception. To think that we do is to succumb to an illusion caused by a confusion between empirical and transcendental claims. Implicit in Kant's reasoning is the recognition that there is a disanalogy between perception and self-awareness. Perceived objects can be identified or misidentified as being a one of a particular thing or of a certain kind. They can also be, (to use Strawsonian terminology) "re-identified", identified at one time with something perceived earlier. All these facts are based on relationships between the observed properties of individual objects or groups of objects by means of which they are individuated. None of this applies in the case of introspective awareness of a self. We do not pick out and distinguish ourselves by means of properties. We only apply the categories where there are individuated temporally located objects, where by means of their properties we are able to identify them as being particular things or of certain kinds. To be of a particular thing or of a certain kind is also to be different from others. However, in apperceptive self-awareness we do not distinguish a self as a subject from any other. This means that neither is there the possibility of misidentification. There is no sense to my saying when I have stubbed my toe. "Someone's toe is hurting all right, but is it mine?' Or consider the statement "I see a house in front of me now". This statement is true just in case there is a house there in front of me and I am looking at it, paying attention, etc. I cannot be mistaken about the fact that it is me so doing. No

possibility of misidentification exists. Kant calls the unity we find in awareness of self more than numerical unity, being one and the same as oneself. Rather it is "synthetic" unity which is difficult to characterise. He writes "This unity is not the category of unity" (B131) i.e. it is not a matter of being one of something. Rather it is a representation which has no properties. The unity of the subject is not like the unity of a chair as a chair which can be individuated by means of its properties. Unlike representations of particular objects it "does not contain in itself the least manifold" (A355). My self-awareness is not the result of any individuating judgement, any application of the category of unity to myself. Indeed my self-awareness is the very precondition of my making any judgement at all.

This is akin to the notion of "self-reference without identification" found in the writings of Chisholm, Castañeda, Perry, and Sydney Shoemaker. As Shoemaker has pointed out, it is pointless to suppose that there is introspective perception of a self unless it plays some part in explaining our introspective self knowledge. However, such reasoning leads to absurdities. He writes in *The First Person Perspective and Other Essays*:

> The introspective observation of a self being hungry is not going to yield the knowledge that *I* am hungry unless I know that that self is myself. How am I supposed to know this? If the answer is that I identify it as myself by its perceived properties, we have to point out that this requires that I already know that I have these properties. Indeed it requires that I know that I am the unique possessor of that set of properties because otherwise the observation that the perceived self has them would not suffice to identify it as me. So I would already have to have some self-knowledge, namely the knowledge that I have certain identifying properties in order to acquire any self knowledge by self observation. If it is supposed that this self knowledge is in turn acquired by self observation, then still other knowledge is required, namely the knowledge that one has whatever identifying properties one used to identify as oneself that one observed to have the first set of identifying properties And so on. On pain of infinite regress it must be allowed that somewhere along the line I have some self knowledge that is not gotten by observing something to be true of myself (Shoemaker, 1996. p. 13).

Shoemaker is making the point that in order to know that anything is true *of* me I must first know that it is *me* of whom it is true. This knowledge is not gained by identifying myself via a further set of properties. Put otherwise, in order to be aware of any property of myself I must first be aware of myself independently of this. I do

not know that "I" am hungry or in pain because I observe my mental states. I do not become aware that I am hungry or in pain because I find that "someone" is hungry or in pain and then identify that individual as myself. In the case of standing before a tree, my experience cannot represent someone standing before a tree without also representing myself as so standing.

As he writes in an earlier work:

> If I say "I feel pain" or "I see a canary", I may be identifying for someone else the person of whom I am saying that he feels pain or sees a canary. But there is also a sense in which my reference does not involve an identification. My use of the word "I" as the subject of my statement is not due to my having identified as myself something of which I know, or believe, or wish to say, that the predicate of my statement applies to it (Shoemaker, 1968, p. 558).

Shoemaker calls the usage of "I" in such cases as "self-reference without identification" and as involving "immunity to error though misidentification relative to the first person". Kant appears to be an early proponent of this view. His writings strongly indicate that he had insights into the nature of reference to self, that he was aware of the peculiar semantics of self-reference. This is because his arguments strongly suggest if not entail that introspective self-awareness could not be inferred from any property of myself unless I already "know" that it is me who has the properties. Kant's genius was to have seen that there are two ways in which we can refer to ourselves. In one way self-reference depends on identification of oneself as "object" and on identification and re-identification. The other does not. It is in the First Edition version of the Paralogisms that he writes: "I cannot know as an object that which I must presuppose in order to know any object" (A402) whereas in the Second Edition version he elaborates the point thus:

> The subject of the categories cannot by thinking the categories acquire a concept of itself as an object of the categories. For in order to think them, its pure self-consciousness, which is what was to be explained, must itself be presupposed (B 422).

Shoemaker argues that it is incoherent to base introspective self-awareness on the model of sense perception. Kant says the same - the condition of the possibility of all judgements relies on the act "I think" which at this level designates *only transcendentally*, no conceptual mediation is involved. As has been discussed, Kant

recognised two types of self-awareness. The first is direct awareness of properties to oneself. This is empirical awareness of particular psychological states such as sensations, memories, beliefs, which is akin to the Humean and Cartesian model of introspection and is a picture of self-awareness similar to that underlying most contemporary accounts. Kant called this "empirical apperception", "inner perception" or "inner sense" (A107). In addition to this is a sense of awareness of self which is quite different from normal perception. It is the knowledge and belief that one is the very entity to whom those properties belong. In fact, the former depends on the latter, for in order for anything to "be something" to someone there must be at least potential for recognition of the properties as one's own. For, in order to make reference to oneself via ascribing properties one must first be able to make reference to a self that does not ascribe properties. This is the difference between "empirical" and transcendental or "apperceptive" self-awareness. In the case of empirical self-awareness one is aware of oneself as having certain experiences or properties. When one is aware of the subject of transcendental apperception one realises that one is the single thing or "x" that has them. One is aware, for example, not only of believing f but of oneself having the belief. One is also aware of oneself as the subject of other psychological properties, that one is the single unified being that not only believes f, but sees g, imagines h and so on. One does not have to rely on "intuitions" in order to come to see this. One has only to consider it to see that it is true. Kant writes "In the synthetic original unity of apperception I am conscious of myself, not as I appear to myself, nor as I am in myself, but only *that* I am" (B157). No recognition of properties is required for reference to oneself as oneself. In ascribing properties to oneself one distinguishes one set from another. But in referring to myself "transcendentally" I am simply aware of myself... "without noting in it any quality whatsoever" (A355).

Some commentators take Kant's rejection of intuitional object awareness of ourselves as amounting to a claim that the I of "I think" is not a referring expression, that it does not refer to one and only one particular individual. Allison (1983), Kitcher (1993), and Powell (1990) all subscribe to a version of this view. For example, Allison writes " 'I'" designates only "something in general", which is to say that it does not refer to anything at all" (Allison, 1983, p. 282). Patricia Kitcher writes that "our knowledge of the self does not rise to the level of knowing that it is

an ineffable something" (Kitcher, ibid. p. 194). All I can be aware of is a collection of representations in inner sense. The rest is an illusion engendered by the formal prerequisites of experience. Is it reasonable to attribute to Kant this view? Whether or not "I" is a referring expression has been debated by a number of philosophers, both past and present. Descartes certainly held that it is so. It refers to a "substance", a "thinking thing" that necessarily exists and which can be known absolutely through introspection. The truth of the Cogito is one that cannot be shaken by even the most extravagant of sceptics. Hume on the other hand claims that it he is not aware of any such entity. If "I" refers at all, it does not refer to a persisting thing, but a bundle of perceptions. Perhaps one of the most famous papers on the topic is Elisabeth Anscombe's *The First Person* in which she denies that "I" is a referring expression. As part of her argument she asks the question "how even could one justify the assumption, if it is an assumption, that there is just one thinking which is this thinking of this thought that I am thinking, just one thinker?" How do I know that "I" is not ten thinkers thinking in unison? (Anscombe, 1975, p. 31). Kant has a response to this which lies in his particular characterisation of the unity of consciousness. On the Kantian account judging is a matter of consciously combining some states in others and of recognising that combination as such and at the same time also being aware of the unity of different states in a single subject. However, I am mistaken if I think that when I introspect I can be aware of myself as object. As discussed, the unity of consciousness Kant talks about is not a matter of being one of something, as opposed to ten, one thinker as opposed to ten thinkers thinking in unison. The unity he is talking about is "not the category of unity" (B131). Rather, his point is that from my own point of view, I cannot think of my experience as other than one. I may, in fact, *be* "ten thinkers thinking in unison" but this is besides the point. Kant writes in his attack on the Paralogisms that when I am aware of myself as subject, I cannot picture myself as a plurality of any sort. In fact, Kant himself argues that "the unity of a thought (…) may relate just as well to the collective unity of different substances acting together" (A353). In other words, he makes a similar point to Anscombe, i.e. there is no reason against a composite subject having the unity required to synthesise representations into a single intentional object. However, from the point of view of myself as subject I cannot picture myself as other than a unity. Whether or not a self is composite in fact is irrelevant.

As he writes:

> We are talking about a "merely logical qualitative unity of self consciousness (…) which has to be present whether the subject is composite or not (B413).
>
> Although the whole of [a] thought could be divided and distributed among many subjects, the subjective 'I' [the 'I' pictured from its own point of view] can never be thus divided and distributed (A354).

In fact, to even think of myself as a plurality of beings requires that I am aware of this plurality and this requires an undivided "me". I cannot imagine going into a fissioning machine and coming out with two streams of consciousness. This is because appearing to myself as a single subject is a prerequisite of my thinking about myself at all. Kant says of the "I think":

> It must be possible for the 'I think' to accompany all my representations; for otherwise something would be represented in me which could not be thought at all, and that is equivalent to saying the representation would be impossible, or at least would be nothing to me (B131-2).

Here Kant is speaking metaphorically of what happens in conscious awareness of an object. Firstly the subject is aware of a representation. The representation is such that if one has it and if one considers that one has it then one will believe oneself to have it, and the subject is aware of that representation. One is aware of the state but one is must also at least potentially be able to be aware that one has the state. This is the "transcendental unity of apperception", the act of judgement by which I take representations to be mine. When a person is aware that he is seeing something he may also be simultaneously aware of other things. He may also be aware that he is touching something and perhaps hearing something. He is also able to be aware that it is he, himself doing the seeing, hearing, and touching. He must also have a notion of himself as experiencing different things at different times. These facts lie at the basis of Kant's transcendental unity of apperception and describe the sense in which a subject can be said to be the subject of their own awareness. This is that one must not only be aware that one is experiencing certain properties, but one must also be capable of knowing and believing that one is the one to whom the properties belong. There must be potential "recognition" of the fact that these attributes belong to them. That is, thinking something necessarily involves self-ascription. If I have a thought it

is necessarily my thought. However, the "I think" is the "analytic unity of apperception". This presupposes the "synthetic unity of apperception". The "I think", the unity of apperception, displayed by the fact that it must accompany all my representations is a spontaneous act of combination or synthesis. This act, that synthesises the representations in one consciousness, is the very same act that analyses out a common subject. For the self and the object of experience are co-determinative; each is a necessary condition for the other. As Kant writes

> Only in so far, therefore, as I can unite a manifold of given representations in *one consciousness*, is it possible for me to represent to myself *the identity of the consciousness in [i.e. throughout]* these representations (B133).

Kant states that if it is the case that a subject has both A experience and B experience then there is one thing that has both A and B experiences, namely that subject. The subject of experience A has been identified with the subject of B experience. Kant tells us that if this were not the case he should have "as many coloured and diverse a self as I have representations of which I am conscious to myself" (B134). However, the "I" of "I think" is not one with which we have direct acquaintance. The self is not part of the content of experience but the "transcendental subject". Kant referred to the unity of this subject as "the unity of consciousness" (A107), "the unity of apperception" (A105, A108) and "the absolute unity of the thinking being" (A353). In the B edition he claimed that it is "not the category of unity" i.e. not a matter of being one of something. My unity as one subject is not the result of any individuating judgement or any application of the category to myself. Of course, there are cases of perceptual knowledge in which awareness of oneself does provide a way of self-identifying such as when we try to pick ourselves out in a group photograph or spot ourselves among an audience on television. Kant's point is that the unity of the apperceiving subject is different. One's awareness of oneself in this sense, unlike awareness of objects, is "non-ascriptive". The self is not a strange kind of entity to which we ascribe properties. Neither is it the unity of a set of relations. It is not like the unity of the categories which consists of the way their semantic relations to one another make it possible for me to conceptualise the world and which Kant says is a perfect unified system. This is a unique kind of unity that is difficult to characterise. He writes:

> Man, [...] who knows all the rest of nature solely through the senses, knows himself also through pure apperception; and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses (A546/B574).

It is argued that Kant saw that "I" is a referring expression and takes as a premise for his transcendental arguments the undeniable fact that he is aware of his own existence "as determined in time" (B275), something no sceptic could take issue with. However, to suppose that "I" is a special sort of demonstrative pronoun used to refer to an introspectively perceived self which I then identify as myself is incoherent. Awareness of ourselves as subject or "intelligible object" rests on facts about oneself that does not conform to the perceptual model of Descartes or Hume. Rather, such self-reference is based on the special unity of the "I think" which for Kant transcends description. As he puts it:

> In attaching "I" to our thoughts we designate the subject (…) only transcendentally, without noting in it any quality whatsoever - in fact without knowing anything of it either by direct acquaintance or otherwise (A355).

In the A edition he wrote "of the absolute unity of this subject (…) I possess no concept whatsoever (A340-B398). There is something about this kind of unity which cannot be expressed in concepts. Other philosophers have expressed similar views about other things. Saint Augustine, in the *Confessions* for instance, who said of time that he knows what it is until he tries to explain it, then words fail him. A similar point is made by Wittgenstein in the *Tractatus Logico-Philosophicus* when he remarks that "What the solipsist means is quite correct …only it cannot be explained but makes itself manifest" (5.26). Like Kant, they are both saying that there is something that can be grasped by the mind but which we cannot sufficiently describe. It is suggested that Kant was the first to describe the peculiar logical semantics of self-reference, and that this philosophical analysis has profound contemporary relevance. Taking advantage of it would help avoid some of the problems that have beset leading contemporary theories of consciousness, such as Higher Order Theories of consciousness (HOTs), examined below.

3. 6. Higher Order Theories of Consciousness

Higher Order Theories (HOTs) are functionalist cognitive/representational theories of mind which assume there is a *level* at which an explanation of phenomenal consciousness can be derived. For example, David Rosenthal's (1991, 1997, 2005) HOT states that to be conscious of some representational state A, one must have another higher representation (a meta-thought) of which the former state is the object. Rosenthal has made significant contributions to philosophy of mind, particularly in the area of consciousness. However, his Higher Order theory of consciousness resembles in many ways the traditional empirical inner sense theory, a direct descent from Descartes and Hume, in which we are aware of conscious states by some kind of inner perception. Higher Order theorists argue that conscious awareness crucially depends on higher order representations that represent oneself as being in particular mental states, in the sense that to be in a conscious state is to be the *object* of another mental state. Yet for Kant, as discussed, there is only one state: the first-person givenness of experiential phenomena is not merely incidental to their being, but is what makes experience possible at all. That is, phenomenal experience entails a built in self-reference, a primitive self-referentiality, that Kant calls the "I think". Put differently, a purported "representation" would itself have the power to make us aware of it and of ourselves. Kant's synthetic apperception and the analytic unity of apperception are isomorphically correspondent elements of, and also reducible to the transcendental unity of apperception, the *grund* or ground that connects together awareness of all distinct particulars. Moreover, the unity of apperception is not only a necessary but also sufficient condition for our representing of objects, as many Kantian scholars have noted: Jay Rosenberg, 2005, p. 135; Henry Allison, 1983, pp. 142–4; Richard Aquila, 1989, p. 159, and Dennis Schulting, 2013. [46]

Proponents of HOTs often mention Kant as an inspiration for their theories, since Kant claims that for a mental representation to be something for me it must be possible for the "I think" to accompany that representation (see Gennaro, 2004); but a proper analysis of Kant's claim shows that this is wrong. HOTs construe a mental state as self-aware, and hence conscious, in virtue of being an object of a numerically distinct second-order state. The idea is that conscious mental states, by which is

meant sensory states possessed of phenomenal feel, (or as Chalmers puts it "what it's like" to be in those states) are conscious in virtue of being the *objects* of other higher-order mental states. This is *what it is* for a mental state to be conscious: to be accompanied by a higher-order thought about that mental state. This, however, neglects the way Kant divides two senses in which we may be said to be self-conscious. The first is through empirical apperception, the attribution of properties to oneself, the self-ascription of items of inner sense which correspond to Hume's bundle of impressions. Here we are aware of representations which are our own. However, there is a second sense of self in which we recognise that we are the subject to whom the properties belong.

> [I]n the synthetic original unity of apperception, I am conscious of myself, not as I appear to myself, nor as I am in myself, but only that I am (B157).

This is a sense of self which does not conform to the perceptual model of Hume. In fact, here we have epistemological access to ourselves which is fundamentally unlike perception of an object, yet is also inexorably linked to the objective. Although Kant wrote that we possess no "knowledge" of such a self; it is not the kind of cognitive "illusion" that many philosophers of mind claim. For example, Daniel Dennett holds the view that once we have explained functions of the mind such as accessibility and reportability there is nothing further to explicate. As he views it, the self is really an "abstraction" brought about by what he terms "the Center of Narrative Gravity", which gives rise in us a spurious sense of a unitary self (Dennett, 1991).[47] Marvin Minsky, also argues self-awareness is a complex, but carefully constructed illusion, (Minsky, 1980). In *The Society of Mind* he argues that minds are simply "what brains do" (Minsky, 1985, P. 287), and proposes that it is the interaction between many functional "autonomous agents" that gives rise to intelligent behaviour. Similarly Thomas Metzinger's thesis is that the self is nothing beyond a special kind of "dynamic representational content" (Metzinger, 2003). No such things as selves exist in the world: nobody ever had or was a self. All that exists are phenomenal selves or "appearances" as they appear in conscious experience.

However, for Kant, there are two quite different kinds of consciousness of self, and for a correct model of the mind it is important to distinguish between them, and not to dismiss one aspect. Few theorists have followed Kant in doing so, however,

and it is suggested that this is due to materialist presuppositions that ignore everything that does not fit into empirical constraints, and a manner of doing science that can be seen as following directly in the footsteps of the empiricist Hume. Yet, Kant's theory of the dual yet unitary nature of self-awareness is a valid alternative to materialist empirical science and compares and contrasts in interesting ways with contemporary theories. These contrasts bring out some of the originality and power of Kant's theory. The main problem with HOTs is how it could be possible that the possession of a higher order state confers subjectivity or consciousness on a lower-order state that did not otherwise possess it? How or why does a mental state come to have a first-person qualitative "what it is like" aspect by virtue of the presence of a HO directed at it? In fact, HOTs seem to turn everything upside down; thoughts about something seem to depend on a subject already having conscious experience of that thing in the first instance. For Kant, just recognising the *object* of a representation is sufficient for the subject to be aware of it and also that it is they themselves who are aware of it. When we look at an object, we are not conscious of our experience of looking at the object, we are conscious of the object *itself*. What makes a mental state conscious is not the subject's awareness of the state, rather, it is the way in which the state makes the subject conscious of *something in the world*. The basis of awareness of an object (and also of oneself) does not involve some separate "higher order" mental state.

Rosenthal thinks it is self evident that a mental state's consciousness consists in our being aware of that state in the same sense that we are aware of external objects, i.e. sensing them or having thoughts about them. However, for Kant, all we need is one state. Moreover, the first-person given-ness of experiential phenomena is not merely incidental to their being, but is what makes experience possible *at all*. Put differently, a purported representational state "A" would itself have the power to make us aware of the object and of ourselves as the subject of the experience. Rosenthal's view is vulnerable to a number of objections such as that of explaining non-human animal consciousness, which conflict with our common-sense intuition that such animals enjoy phenomenally conscious experience, yet are unlikely to have the conceptual sophistication required by HOTs (Jamieson and Bekoff, 1992; Dretske, 1995; Tye, 1995; Seager, 2004). The model also readily leads to a regress of thoughts about thoughts about thoughts, and since there is little independent

reason to postulate such a hierarchy of thoughts, this is a serious weakness. Kant's theory neatly avoids this: Apperception is the taking up and thinking through of content in which the self is given to itself neither as an object or subject but as a mode of being in a state directed to the world, which is known *instrinsically* through reflective judgement. Only an intrinsic or "one-state" theory, according to which the self-awareness involved in conscious states is intrinsic to those states, can explain this distinction. In this way, one's conscious experience and one's tacit awareness of that experience form a single mental act, the spontaneous act of synthesis.

A number of contemporary philosophers have suggested that Franz Brentano's theory is a viable one-state alternative to higher-order theories.[48] It was Brentano who first introduced the notion of intentionality or "aboutness" to contemporary philosophy and psychology. But Brentano's theory, just like higher-order theories, also characterises self-awareness in terms of conscious states being objects to which subjects stand in a certain kind of relation. As Dan Zahavi has noted, "Brentano's claim that every conscious intentional state takes two objects, a primary (external) object and secondary (internal) object, remains committed to a higher-order account of consciousness; it simply postulates it as being implicitly contained in every conscious state" (Zahavi, 2006, p. 5). According to Kant, however, even though our subjective states themselves are "manifested" in experience, they do not become *objects* of consciousness except in the phenomenal sense: intentionality is coessential with the pre-reflective awareness of transcendental apperception, since any consciousness at all exists *qua* conscious of existing. For Husserl too, who was a Kantian, our experiences are conscious not in virtue of being taken as secondary *objects*, but rather in virtue of being "lived through".

Thus, Kant (and Husserl) did not deny that consciousness involves self-awareness, but they deny that self-awareness can be accounted for on analogy with our consciousness of extra-mental objects, i.e. in terms of a subject-object relationship. To be "something it is like" requires "for me" to grasp it in consciousness. In simple terms, conscious awareness of an object has an implicitly dual nature, such that to be conscious of an object is also to be aware that one is in that very state. But this awareness is not, itself, a separable feature of the first order state. One has explained consciousness precisely when one has explained this dual feature. Such consciousness is perhaps more aptly described as a mode of being in a

contentful state directed to the world, rather than it being so in virtue of a relationship to a numerically distinct second order state. By "mode of being" is meant that which expresses the tacit phenomenological veracity of the inherent pre-reflectiveness of conscious experience.

Kant was the first to recognise that phenomenal conscious awareness is essentially bound up with spatial orientation and temporal asymmetry. Spatio-temporal designation is essential to the singular presentation of objects in the world that are experienced, and also to our singular cognitive reference to them. This spatiotemporal orientation is what gives the peculiar subjectivity of experience its own "unique point of view". This is what determines the "what it is likeness" of being in a mental state, e.g. just "what it is like" to smell the aroma of coffee or to have toothache, or to see the redness of a rose that is in front of me now. He also saw that this is in two parts, a duality, yet a unity. There is the qualitative character, the redness, the pain and the aroma in the object that is presented to the senses and the subjective character of the experience - "what it is like" to experience them. How best to capture that ineffability of subjective phenomenal experience is what is at stake. Kant recognises the "for-me-ness" aspect of subjective experience but not only this, he recognises that this is that which makes the conscious state phenomenal at all. An experience of smelling the coffee is different from that of toothache, but what they have in common is the first person perspective in which qualities of a certain kind are presented to the subject. This contrasts with HOTs whereby the subjective character of experience is a higher order access to a lower order one, or a relationship between a higher and lower representation. As discussed, according to HOTs, a phenomenally conscious experience of red would be based on the content of a mental state red with a direct relationship to the meta-thought red (Rosenthal,1997, 2005). That is to say, the higher level meta-thoughts are distinct representations, the latter being phenomenally conscious in virtue of the former which represents it. This does absolutely nothing to account for the distinctive feature of phenomenal awareness, the "for-me-ness" of the experience. To be "something it is" like requires "for me" to grasp it in consciousness. Thus, to account for the first person, subjective grasp of the experience, there needs to be a subject of that experience situated in time and space. A subject must be something that holds the mental states together. Simply being represented by a higher order state is insufficient to explain the something it is

like, or "for me-ness" of phenomenal consciousness. Kant terms this "transcendental apperception" and it is linked to his idea of "transcendental designation" (A355), [see p. 117]. That is to say, all conscious states involve *self*-awareness, although this is not awareness of a "self" in any substantial sense of the word such as Descartes had claimed: the self is not something that "exists" apart from, or above, the experience and, for that reason, something that might be encountered in separation from the experience. Rather, we have an awareness or sense of self that is described by Kant in terms of the kind of unity we have, and which is an integral and ubiquitous part of our experiential life. Indeed, there is no experiential dimension whatever without this sense of self.

Kant is thus making the revolutionary move that this kind of awareness is best understood as involving a non-objectifying self-givenness which is the ground of all experience of the world. This is linked to his particular notion of unity and of the special sense in which he uses the term "existence". In traditional accounts, experience, including sense perception, is assumed to consist of a succession of states of the subject, conscious states being objects to which subjects stand in a certain kind of relation. And that once that is accepted, it will be natural to expect descriptions of perceptual experience to focus on something *within the subject.* For Kant, however, this neglects "time" as the form of inner consciousness that structures appearances in a series (A37/ B54) and, in the schematism of imagination, is the ultimate form of synthesis that enables experience in general. Olli Lagerspetz, noting that "the mainstream of cognitive science has retained Descartes' core methodological commitments, but in a setting that renders them largely meaningless", puts the matter well:

> The category of "existence" implies determination in time. Whatever exists subsists over a time span. I establish the passage of time by using some enduring object as a point of reference. Crucially, this object must exist independently of my consciousness in order for it to constitute a real and not just an imaginary object of comparison. Hence, I can have any positive knowledge of myself as subsisting in time "only through the existence of actual things that I perceive outside me" (Lagerspetz, 2002, p. 12).

This is due to the fact that the *a priori* principle of causal interaction immediately links our inner representations with objects outside of us. Without this objective basis for their associability, we would not be able to conceptualise experience.

Neither would we be able to represent ourselves as a numerically identical subject through different experiences.

To conclude this chapter: Descartes, in speaking of the *existence* of a thinking *thing* or *entity*, had made the logical fallacy of deducing, from the logical analysis of the concept of thinking, the metaphysical hypothesisation of an immaterial object. Hume had derived from this the supposition that any consciousness of self must be an experience of "something", a bundle. As Kant diagnoses it, a Copernican Revolution or reversal of approach is needed. Self-awareness involved in conscious states cannot be construed along subject-object lines. The unity of intuition is subject to the synthetic unity of apperception because the unity of the act of combining *is* the unity of consciousness, the "I think". This is the identity "at a time" that goes with the unity of a subject, and it is very different in kind to the unity of the rationalists, the unity of a persisting "substance". This is related to his doctrine that existence is not a predicate (A598/B626) and also to what Kant says about reference to self. Being aware of oneself as oneself is something that "transcends" being aware of "qualities" of oneself. Thinking, consciousness or, to put it in the idiom of cognitive science, the having of conscious states, obviously involves someone who does the thinking. What does not follow from this is that there is a substance, or any of its contemporary equivalents: a representational medium, level of organisation, or neural correlates of consciousness operating behind the scenes as the locus of thoughts. The unity of apperception transcends such empirical considerations. Physicist Erwin Shroedinger pondered on this problem in the epilogue of his classic *What Is Life?* and his answer is remarkably Kantian:

> Consciousness is never experienced in the plural, only in the singular. Even in the pathological cases of split consciousness or double personality the two persons alternate, they are never manifest simultaneously. In a dream we do perform several characters at the same time, but not indiscriminately: we are one of them; in him we act and speak directly, while we often eagerly await the answer or response of another person, unaware of the fact that it is we who control his movements and his speech just as much as our own.

He concludes:

> (i) My body functions as a pure mechanism according to the laws of nature.
>
> (ii) Yet I know, by incontrovertible direct experience, that I am directing its motions, of which I foresee the effects, that may be fateful and all-important, in which case I feel and take full responsibility for them (Shroedinger, 1944, pp. 86-88).

It has thus far been argued that a certain picture of the mental has given rise to a host of contemporary philosophical problems. This is largely the result of the lasting influence of a generally forgotten or neglected philosophical heritage which has led to the situation where the many significant and far reaching discoveries concerning cognition and the workings of the brain have been obscured by their presentation within an incoherent conceptual framework. It is proposed that although Kant's views about consciousness and "the self" may differ drastically from the ways in which these topics are currently discussed, his theory of the transcendental subject is an invaluable tool in unravelling these philosophical complexities. Kant claimed that the rationalists, in particular Descartes (and Leibniz) were seduced by what he calls a *Transcendental Illusion*, a pervasive intellectual illusion, modelled on the perceptual, which predisposes even the "wisest" of people to accept as sound certain invalid arguments for substantive theses about the nature of the self. Kant, in the Paralogisms, provides an ingenious account of how the rational psychologist might arrive at such a view. He identifies the source of the fallacy as concerning the very nature of conceivability itself and in this identification makes a philosophical contribution of lasting significance. Kant's intellectual insights concerning the mind have stood the test of time, and are able to shed light on the modern day hard problem of consciousness, deemed the most puzzling scientific problem of this age. They provide a conceptual tool for making sense of consciousness, the "what it's like-ness" of phenomenal experience, the me-ness that accompanies all thought. He shows that the nature of subjective experience and its phenomenal *qualities* is impossible to understand in terms of dualism of substances or properties and addresses the problem of trying to explain introspective first-person aspects of mental states (the mind) and consciousness in general in terms of third-person quantitative neuroscience (the brain/body). For Kant, there is a confusion with the idea of "mind" and "body" as conceptually separable in the first place. The nature of the question is misleading and leads us to believe that there is something tractable that can be substantiated in the realm of empirical knowledge. He held that certain mistaken beliefs about the mind and consciousness arise from reification of first person phenomenal experience, because certain philosophers (Descartes) have projected the particularity of private, first person experience onto a third person entity called "the mind" or "thinking substance". This is not to deny that

consciousness involves self-awareness, it is to deny that self-awareness can be accounted for on analogy with our consciousness of extra-mental objects, i.e. in terms of a subject-object relationship. The modern day mind-body problem is therefore the result of a kind of category error, where the *a priori*, transcendental character of subjectivity is confused with psychological "facts" that are amenable to scientific explanation. The insights of Kant applied to the contemporary mind/ body problem reveal that the problem *is the very disease of which it takes itself to be the cure.* The very debate, the actual effort to find the "solution" to the various fragments of the hard problem, indeed the conceptualisation of the problem itself, is based on faulty foundational premisses, premisses which not only ensure that "the hard problem" remains insoluble, but perpetuate its existence, seemingly *ad infinitum.*

Thus, Cartesian and Humean approaches to cognition continue to hold sway today; functionalist approaches to cognition are, more or less, updated Hume, and Chalmers' "hard problem of consciousness" is Cartesian. This has resulted in an ongoing and apparently tireless game of what Dennett has called "burden tennis" (Dennett, 1993) where the field of play is conceptual space and each side claims that the ball is in the other's court. Kant's intellectual genius and depth of insight into the sources of our cognition can be of great significance in addressing this impasse; between those who claim there is a "hard problem" to be addressed by a science of consciousness and those who deny it, (Dennett) since this picture of mentality is a resurrection of the conceptually flawed understanding that Kant addresses in the *Critique.* Although Kant may not have used the concept of consciousness in the now dominant sense of phenomenal *qualia*, his theory of the "transcendental subject" is a valuable tool in unravelling the philosophical complexities and confusions that are the bane of current theories. In order to begin this enterprise, the following chapter examines the extent to which Kant's transcendental psychology can be considered compatible with modern functionalist theories and, most importantly, where it diverges.

4. Transcendental Psychology and Functionalism

It has become more and more respectable to claim that Kant's ideas about the mind are consistent with functionalism, the prevailing approach in cognitive science at present, and consequently it has become more commonplace for commentators to regard Kant a functionalist of some sort and to identify accounts of the transcendental psychology with an accommodation of it to some of the central tenets of traditional cognitive science. Patricia Kitcher, Andrew Brook, Ralf Meerbote, Wilfred Sellars, Daniel Dennett and C. Thomas Powell are among the members of this group; although differing in details, their common general view is that Kant can be viewed as early proponent of functionalism, what might be termed a proto-functionalist. For Kant, the way to understand mind was through *a priori* reasoning, the analysis of what the mind *must* be like and what capacities it *must* have to represent things as it does. This is his "transcendental method" which is seen by contemporary philosophers to have had a major influence on the research programme of contemporary cognitive science. Kant's question was how is it possible for something to be a thought? What are the *a priori* conditions of the possibility of experience? Contemporary cognitive scientists, in a similar vein, also try to justify their assertions by showing that they provide the necessary presuppositions for the possibility of meaningful discourse about certain aspects of the mind. In other words, in order to give epistemic justification to their claims, they also try to show what must be the case in any system in order for a particular phenomenon to occur, and to do so without appealing to hidden psychological "facts", but rather by appealing to certain "conceptual prerequisites".

4.1. What is Functionalism?

Functionalism is a philosophical theory or family of theories according to which cognitive or psychological states are essentially states of whole systems. As discussed in Chapter 1. 2., the first formulation of a functionalist theory of mind was put forward by Hilary Putnam [49] in 1960 and was inspired by the analogies which

Putnam and others noted between cognition and the theoretical or hypothetical "machines" or computers capable of computing any given, and which had been developed in 1936 by Alan Turing, called Turing Machines. Turing machines compute various tasks deterministically, given certain instructions and a machine table that specifies how various states of the machine relate to one another and to inputs. Turing himself was of the opinion that a machine operating this way would quite literally be performing the same computations a human performing computations would. He claimed that the brain must be organised for intelligence, that all mental operations are computable and hence realisable on such a machine, and that *all* mental functions of the brain could be accommodated within this model and not merely those of a mind following definite rules (Turing, 1950). Two years after Turing's hypothetical computing machines, Claude Shannon (1938) demonstrated that simple on-off electrical circuits could carry out basic mathematical procedures, an idea that ultimately led to the development of "information theory" (Shannon, 1938, 1948; Shannon and Weaver, 1949, 1998). Five years later Kenneth Craik wrote a book entitled "The Nature of Explanation" (Craik, 1943) in which he discusses possible ways to link mental and mechanical operations, and settled the notion of internal models, as he wrote:

> Thought is a term for the conscious working of a highly complex machine, built of parts having dimensions where the classical laws of mechanics are still very nearly true, and having dimensions where space is, to all intents and purposes, Euclidean. This mechanism, I have argued, has the power to represent, or parallel, certain phenomena in the external world as a calculating machine can parallel the development of strains in a bridge (p. 85).

Craik's vision was that minds create models and use them to predict the future, the idea being that the organism is able to choose, from the array of possible futures, the one that would be the most adaptive. Putnam's extension of this idea was that mental states or *representations* can be thought of as having functional roles within a system, where a system, in the case of human beings, is a Turing Machine with a probabilistic transfer between different states (rather than a deterministic one) called a Probabilistic Automaton, probabilistic because human beings are not predictable in the way a Turing Machine is. Since Putnam's original formulation, its apparent sturdy empirical grounding is viewed as providing a solid foundation on which to build a more concrete conception of mind than the earlier "brain state" theories of

U.T. Place, Herbert Feigl, J.J.C. Smart and David Armstrong (see Chapter 5). As a result of this, cognitive science proceeded within a symbolic framework that required little or no contact with the brain. Functionalism offers convincing arguments as to why any physico-biological implementation of cognitive functions is derivable from an abstract picture of the logical structure of the mental representations and transformations involved in those functions (Block, 1980).

Functionalism was first systematically developed as a response to questions about the relationship between mind and body, in the context of a debate between two opposed views in the philosophy of mind, *dualism* and *materialism* or *physicalism.* It was derived in order to be compatible with materialism, whilst avoiding the difficulties of behaviourism and the identity theory of J. J. C. Smart, among others, whose theories entail that for every type of mental state there corresponds a type of physiological brain state with which it is identical. Putnam and others were arguing against the view that every mental kind is identical to some as yet undiscovered neural kind. There were several serious problems with this "brain-state" view and this is what motivated the idea that the relation between mind and body is analogous to the relation between software and hardware in a computer. For example, Putnam argues, against the brain state theorists:

1. The brain state theorist maintains that *every* psychological state is a brain state in the sense that there will always be one and the same physical "correlate" of the same psychological state.
2. But we can find at least one psychological state (such as "feeling hungry") which can clearly be applied to both a mammal and an octopus, but whose physical-chemical "correlate" is different in the two cases.
3. Therefore, the brain state theory must be false. [50]

The advantage of functionalism was that mental states are describable in terms of what the functions bring about, *what they do* in a whole organism or system rather than *what they are*. Computational states are defined, not in terms of specific hardware configurations, but in terms of their relations to inputs, outputs, and other computational states. According to functionalism, the phenomenon to be explained is usually defined as a logical function, e.g., the logical inference reported by a subject who is having at that moment a particular experience, say discriminating a colour or noting a flashing light. Any machine able to make the same inference (through the manipulation of symbols in a language translatable to the subject's) can be said to simulate or predict it, without having the phenomenal experience of the colour or the

flashing light. The functionalist concept of explanation is one that takes such logical simulations/predictions as sufficient to explain cognitive phenomena, and consciousness or phenomenological experience is secondary and of little or no consequence.

As a theory of mind, functionalism is very appealing to scientists. This is because firstly, it has scientific rigour since it is grounded in mathematical proof and secondly, it provides a means of constructing analogies to aid guidance through a systematic, abstract understanding of cognition. Computation-representation functionalism is thus a clear example of a scientifically "rigorous" model; the core presumption is that digital computers and people are both Turing machine describable, and because of this, the computer/brain analogy can be supported by direct reference to computational theory. This is how Turing machine equivalence has played such a central role in supporting functionalist intuitions (Fodor, 1981).

This idea has been transported into cognitive neuroscience, and has its roots in the proposal advanced by D. Marr (1982) who, along with Tomaso Poggio, treated cognition as an information processing system that can allow cognition to be studied on the basis of three complementary "levels" of analysis whereby it can be explained in terms of: computation, algorithm, and implementation (Marr's Tri-Level Hypothesis). The methodology itself is founded upon four main pillars:

i) The measurement of behaviour.

ii) The attempts to find empirical correlation (using EEG, invasive electrodes, or, more recently PET, MEG or fMRI) of such behaviour with brain regions or networks allegedly responsible for them.

iii) The formulation of mathematical/computational functions able to account for them.

iv) If the simulations of such functions satisfy the measured behaviour, it is taken as a proof that the activity of the correlated brain regions/networks "implements" the mathematical computational function.

This methodology is said to make available a complex map of the "localisation of function" that are within the brain, along with a set of mathematical/computational functions that formally describe the inner "mechanics" of the correlated brain regions. This functional, computational or "information processing" view aims to understand the mind in terms of processes that operate on "representations". The

underlying assumption is that any cognitive process, animal, machine, human, could be thought as a computable function. It entails that cognition is explicable in terms of discrete, mental representations (symbols) and that cognitive processes are transformations of such representations or symbols described in terms of rules or algorithms. As professor of cognitive science at Rutgers University, Zenon W. Pylyshyn writes:

> [T]hat certain behavioural regularities can be attributed to different representations (some of which are called "beliefs" because they enter into rational inferences) and to symbol-manipulating processes operating over these representations, is a fundamental assumption of Cognitive Science. This idea is an instance of what is a fundamental claim about intelligent systems: Intelligent systems (including animals and computers) are governed by *representations* (Pylyshyn, 1999, pp. 7-8).

Also:

> What makes it possible for systems - computers or intelligent organisms - to behave in a way that is correctly characterized in terms of what they represent (say, beliefs and goals), is that the representations are *encoded* in a system of physically instantiated symbolic codes. And it is because of the physical form that these codes take on each occasion that the system behaves the way it does, through the unfolding of natural laws over the physical codes (ibid., p. 4).

Moreover, Pylyshyn is adamant that computation should not be regarded as merely a handy *metaphor* for cognitive activity, but as a definite *scientific* empirical hypothesis:

> The question of what's in the mind should be answered in psychology the same way that the parallel question is answered in physics. There, a question such as what's in this table or what's in the sun is answered by looking for properties, entities and causal laws which explain the important regularities that define that particular science (ibid., p. 1).

This understanding of cognition was shared by Alan Turing, Allen Newell and Herbert Simon, Marvin Minsky, Hilary Putnam, and Jerry Fodor. Newell, for example, couched cognition in terms of the physical symbol hypothesis, according to which being a physical symbol system (a physical computer) is a necessary and sufficient condition of thinking. Marvin Minsky claimed, famously, that "the brain is just a computer made of meat." (Minsky, M. Quoted in: Michalowski S. Science, Man, and the International Year of Science). In fact, ever since the development of

powerful modern computing machines around the 1940's and the beginning of the computer age, many scientists (and some philosophers) have simply taken it for granted that it was only a question of time before satisfactory machine models of human intelligence were produced.

Cognitive science mushroomed rapidly on the basis of this paradigm, resulting in detailed theories of cognitive processes including perception, attention, memory, language and decision-making. Rather than using mental representations, another kind of model, proposed by cognitive linguist Ray Jackendoff, focused on mental "structures" which differ from mental representations in terms of their intentionality or "aboutness". Intentionality is a philosophical concept regarding the power of minds to be about, to represent, or to stand for, things, properties and states of affairs. According to Jackendoff our use of language already presumes such rich representational structures that can be linked to what we see, so that "there is no reason to be paralysed by the absence of a solution to intentionality, as Fodor seems to be" (Jackendoff, 2002, p. 280). In other words, meaning is a separate combinatorial system not entirely dependent upon syntax. However, whatever the particular flavour of the symbolic modelling approach, the foundational claim is that an algorithm can be used to model or simulate cognitive processes and produce output that corresponds with human behaviour. In fact, Jackendoff staunchly defends functionalism and the symbol manipulation paradigm and claims that although "some neuroscientists say we are beyond this stage of enquiry, that we don't need to talk about "symbols in the head" anymore, I firmly disagree" (see Lakoff, 2008).

A further type of functionalist model, that has become more commonplace, is the *connectionist* kind which focuses on what are termed "neural networks", i.e. networks comprising neurons or neuron-like elements and the description of connections between these elements (Rumelhart & McClelland, 1986). These later connectionist models were inspired by the neural architecture of the brain, where the neurons they employ are a much simplified version of the real thing and relegated to the role of "mere implementation". Connectionism also entails the commitment to mental representations as the distributed patterns of neural activity. These kinds of "neural network" models proceed according to "feed forward" or recurrent connections, all based on algorithmic computation. In fact, David Chalmers is a

connectionist functionalist, and writes the following of the implementation of computation in a system:

> The relation between an implemented computation and an implementing system is one of isomorphism between the formal structure of the former and the causal structure of the latter. In this way, we can see that as far as the theory of implementation is concerned, a computation is (...) an *abstract specification of causal organization* (Chalmers, 2011, p. 6).

Steven Pinker, Harvard Professor in the Department of Psychology at Harvard University also adheres to a connectionist, computational theory of mind. Relying on the work of Newell and Simon, Minsky, Putnam, Fodor and Marr, he defends the view that the human mind is a naturally selected system of "organs of computation". In *How the Mind Works* he frequently refers to "the mechanisms of cognition", and considers computation or "information processing" as simply what the brain does to allow us to see, feel, think, choose, and act. Pinker claims both i) that thinking is a kind of computation used to work with configurations of symbols, and ii) that the mind is organised into specialised modules or mental organs. Pinker also explains the mind according to the supposed principle of "reverse-engineering"; that is scientists are able to come up with the evolutionary narrative of what nature intended the mind to be able to do as it evolved, simply through tracing the process of natural selection. The mind, he writes, is a system of "organs of computation" that allowed our ancestors to understand and outsmart objects, animals, plants, and each other. He also repeats this computational language in a later article defending his theory, *So How Does the Mind Work?*

> Mental life consists of information processing or computation. Beliefs are a kind of information, thinking a kind of computation, and emotions, motives, and desires are a kind of feedback mechanism in which an agent senses the difference between a current state and goal state and executes operations designed to reduce the difference (Pinker, 2005, p. 24).

In short, Pinker retains a computational theory of mind, one premised on the idea that information processing is the fundamental activity of the mind and it is "function" that makes human beings intelligent. The mind is akin to an information processing machine, a kind of biological Turing machine or "virtual mental computer" (Pinker, 1997, p. 69). The central point of his neural computation model

is “that symbols both stand for some concept and mechanically cause things to happen” (ibid., p. 71). Cognition works, in part, via combinatorial symbol-manipulation, not just associations among sensory features, as in many connectionist models. Moreover, he avers, “what makes a system smart is what the parts of the machine stand for and how the patterns of change inside are designed to mirror truth - preserving relationships” (ibid., p. 77).

Whatever the functionalist theory of mind, all share the common assumption that our cognitive activities and mental life can be reduced to information processing, a deterministic, scientific, materialistic, mechanistic explanation. This is the core premise - an almost universally held general belief that there are no *a priori* reasons why thought processes are not reducible to materialist, information processing, mental representations, and that the question of whether or not they are is purely an empirical matter. Daniel Dennett, a well known adherent to the functionalist endeavour, who, with Jackendoff, is co-director of the Centre for Cognitive Studies at Tufts, puts the matter clearly when he claims that “the materialistic, empirical study of mind aims to find reductive explanations that make mental representations “demystified, unified, [and] placed on more secure foundations” (Dennett, 1995, p. 89) and do so in order to demonstrate *why* a particular physical state counts as a mental state. Dennett asserts that this is the correct view of human cognition and human nature, and it is one where consciousness itself should be left out of the picture, because it does not *do* anything at all. In a paper illustrating two opposing teams tackling the problem of consciousness, he writes: “Turing’s great contribution was to show us that Kant’s question could be recast as an *engineering* question. Turing showed us how we could trade in the first-person perspective (...) for the third-person perspective of the natural sciences and *answer all the questions*–without philosophically significant residue (Dennett, 2001, p.1). According to Dennett, Turing’s Kantian question is “How could we make a robot that had thoughts, that learned from experience and used what it learned the way we do?” The next section evaluates this claim.

4.2. Kant and Functionalism

A major doctrine of functionalism is encapsulated in the fundamental dictum "function does not determine form". Contemporary functionalists hold four main principles:

i) We know very little about the mind as it is structurally.

ii) Therefore the way to model the mind is on what it does or can do, on how it works or functions, rather on how it is constituted or on its introspective contents.

iii) Given this functionalist explanation nothing follows concerning its actual nature.

iv) Therefore, mental functioning could, in principle, be realised in systems of many different forms.

Putnam defined the "functional organisation" of minds or "mental systems" in terms of their role within a representational system and characterised them in terms of their function within that system. Moreover, he points out, descriptions of the functional role of a system are of a logically different kind from descriptions of its actual nature, i.e. nothing follows from them concerning any underlying composition. Therefore, to use his own words, "any physico-chemical system" (that can be represented functionally, by a machine table) "is functionally isomorphic to a denumerable infinity (at least) of systems with quite different physical-chemical constitutions" (Putnam 1967, p. 242.) This is broadly akin to Kant's own conception of what we can and cannot capture in a model of the mind. As mentioned, Kant claimed that we know little of the mind's structure. Thus he writes:

> Through this I or he or it (the thing) which thinks, nothing further is represented than a transcendental subject of the thoughts = *X*. It is known only through the thoughts which are its predicates, and of it, apart from them, we cannot have any concept whatsoever (A346/B404).

As discussed earlier, Andrew Brook likens this to Kant's "doctrine of the noumenal" - that for Kant we know absolutely nothing of the noumenal mind, the "substrate" that underlies functioning, of which the senses give no knowledge but whose bare existence can be inferred from the nature of experience, which he says gives credence to a functionalist interpretation of the *Critique*. Wilfred Sellars also points

out that according to Kant we do not know mental processes "*save as processes which embody these functions*" and that his "revolutionary move was to see the categories as concepts of functional roles in mental activity" (Sellars, 1974, pp. 62-90). In the Second Paralogism Kant writes that we do not know whether "the thinking "I" is different from or the same as body. It may be the same as body but to assert this is to move fallaciously between different logical categories, descriptions of composition and descriptions of function. That is, a synthetic description of the composition of the "I" of "I think" involves an illicit slide between statements of different logical types given the purely analytic, functional, *a priori* characterisation of the mind with which Kant is concerned. Brook claims that Kant's attempt to characterise the necessary characteristics or prerequisites of a thinking subject led him to the recognition that through such analysis nothing about the actual composition of the subject need be inferred; in other words, that Kant's attempts to characterise such prerequisites within the framework of his transcendental idealism led him to a recognition of the dichotomy between function and material or physical composition, which is akin to the functionalist claim, that mental functioning could be realised in a denumerable infinity of different forms, the argument from multiple-realisability.

So, to recapitulate in order to get our bearings: the functionalist model of the mind is the prevailing model in cognitive psychology. Proponents hold that the way to model the mind is through its functions, what it does and can do. The claim is that this general picture is akin to Kant's own model which was also centred on how the mind works at an abstract level rather than on how it is constituted or on its introspective contents; that like modern cognitive scientists, Kant was concerned with the "functions" or conditions needed for functions to work. Concepts are *functions* based on acts of unity of judgement. Kant writes about concepts as functions as follows:

> Whereas all intuitions, as sensible, rest on affections, concepts rest on *function*. By "*function*" I mean the unity of the act of bringing various representations under one common representation. (...) Now the only use which the understanding can make of these concepts is to judge by means of them (B 93) [my italics].

As discussed in Chapter 3, Kant claimed that knowledge is the result of "acts of synthesis", that judgements have the particular content they do have in virtue not of their immediate causal relationships to objects but only through their dependence on intuitions (A68/B93); that in order that representations of objects be anything *to anyone* they must belong with others "to one consciousness"(A116), i.e. they must be synthesised or combined with others into one unified representation, and that this in turn requires the application of concepts. His emphasis throughout is on the workings of the mind.

Kant's functionalism is of a very general kind, however. Unlike contemporary cognitive scientists he was uninterested in specifics. Nevertheless, it is reasonable to state that this aspect of his theory can be viewed as an early form of functionalism, what we might term a *proto*-functionalism. One of the most intriguing aspects of Kant's alleged proto-functionalism, is that it can be viewed as akin to the negative doctrine of contemporary cognitive science encapsulated in the dictum "function does not determine form". Contemporary functionalists hold that we know very little about the mind as it is structurally, and Kant's own conception of what we can and cannot capture in a model of the mind appears to attest to the same. Andrew Brook claims this is an aspect of his doctrine of the "noumenal" and maintains that Kant's insistence on the unknowability of the noumenal mind implies a broad agreement with functionalism, viz., that (i) mental functioning could be realised in principle in objects of many different forms; and, (ii) we know too little about the form or structure of the mind at present to say anything useful at this level in any case, except that mental functions will never be straight-forwardly mapped onto any forms that may be associated with them, whatever these forms might be like (Brook, 1997, p.13). He writes that Kant accepted a variant of both these positions; that concerning (ii), Kant maintained not just that we know little about the "substratum" (A350) that underlies mental functioning but that we know nothing (or we can never know that we know anything) about it. If the noumenal mind is unknowable, however, (i) immediately follows; the mind "as it is" could take different forms. Indeed, "function imposes so few constraints on form that, so far as we can infer from function, we cannot determine even something as basic as whether the mind is simple or complex (A353)" (ibid., p.12). In short, Brook claims, Kant not only accepted the notion that function does not dictate form, but "accepted a very strong version of it" (ibid., p.

14). For Kant not only do we now know little of the mind's structure - we will never know it. As far as the nature of the mind is concerned Kant advocated strict ontological neutrality.[51]

Discussing the subject of the role played in thought by the "I" of transcendental apperception Kant writes:

> Since, in thinking in general, we abstract from all relation of the thought to any object (…) the synthesis of the conditions of thought in general (...) is not objective at all, but merely a synthesis of thought with the subject, which is mistaken for a synthetic representation of an object (A397).

The "I" of transcendental apperception is not objective, i.e. something to which we could ascribe properties. For Kant, thought of what is objective must be conceptualised by the categories, which are the very conditions of our knowing anything at all. Moreover, the categories themselves are not to be considered entities of any kind but the most general logical constraints on anything that can function as a mind. They are concepts of functional roles in mental activity formed by abstraction, not by reflecting on the self as "object" but from consideration of the mind's conceptual capacities. That is to say, the categories are not "things" of any kind however rarefied, but are characterisable purely in terms of their function within a system of representation. Their material composition cannot be inferred, just as the nature or "substrate" of the "I" cannot. Kant's description of the mind in the *Critique* is quite readily translatable into functionalist terminology, since all the conceptual terms he uses can be construed in terms of their role within a representational system and characterised in terms of their function within that system. For Kant, as for modern functionalist theories of mind, abstract descriptions of mental content are logically different from descriptions of actual material composition. Kant claimed that nothing can be inferred from them of the actual composition of that which realises these states.

Putnam writes the following on the idea of multi-realisability:

> The functional organisation (...) of the human being or machine can be described in terms of the sequences of mental or logical states respectively (...) without

reference to the nature of the "physical realisation" of these states (Putnam, 1975, p. 373.).

[D]escriptions of the functional organisation of a system are logically different in kind from descriptions of psycho-chemical composition (ibid., p. 424).

Echoes of Putnam's thoughts can be discerned in the 18th century language of Kant, when he writes:

If anyone propounds to me the question: What is the constitution of a thing which thinks, I have no *a priori* knowledge wherewith to reply. For the answer has to be synthetic - an analytic answer will perhaps explain what is meant by thought but beyond this cannot yield any knowledge of that upon which all thought depends for its possibility (A398).

The notion that Kant propounded a proto-functional doctrine of the mind is compelling. For from the Kantian characterisation of a mind that is advanced in the *Critique*, nothing follows concerning its actual nature. It does not tell us whether the "I" is the same as or different from the body. It may be the same but to assert as such is to move illicitly between talk of composition and talk of functions. Putnam wrote that "any physico-chemical system which possesses a "functional organisation" which can be represented by a machine table is functionally isomorphic to a denumerable infinity (at least) of systems with quite different physical chemical constitutions"(Putnam, 1974, p. 242). Similarly, for Kant, the cognitive acts of the self *qua* representational system are independent as to the type of physical mind of a subject that might engage in them. We do not even know something as basic as whether or not it is simple or composite. If we know nothing at all about the mind "as it is" it seems to immediately follow that there are an infinite number of ways in which it could be instantiated. Kant does give a detailed account of the functions of the "I think", but from this nothing follows regarding the compositional basis of the mind.

Kant attacks the assumptions of the proponents of rational psychology (Descartes) who claim as the foundation of their beliefs that in the "I think" one must be aware of an actual substantive thing. He wites: "I do not know an object merely in that I think (...) I do not know myself through my being conscious of myself as thinking

(B406). Kant insists that the "I" of "I think" or "the unity of consciousness" is not substantial in the way rational psychology claims. Neither is awareness of such unity analogous to our perceptual experience of objects; rather it is the very condition of all such perceptual experience. Kant's claim is that the use of "I" does not involve any intuition of a subject of experience because the "I think" is not an inner perception of an immaterial object at all. What we are aware of in the "I think" is a not an entity or "substance" with special properties of unity, simplicity and absolute persistence, as Descartes had hypothesised. As discussed, earlier, Kant argues in the Second Paralogism that the unity of thought "may relate just as well to the collective unity of different substances acting together" (A 353). Knowing that the elements in cognition need to be synthesised into a single object by a single subject tells us little about what kinds of structure might instantiate a subject with such abilities. The rationalists had taken their claims to be a demonstrative science the propositions of which were analytic. However, Kant makes the point that no substantive claim about thinkers can be derived from analytic propositions. Rational psychology is mistaken. Kant diagnoses the error in both editions, where in the B edition he writes:

> From all this it is evident that rational psychology owes its origin simply to misunderstanding. The unity of consciousness, which underlies the categories, is here mistaken for an intuition of the subject as object and the category of substance is then applied to it (B421-2).

...whilst in the A edition he writes: of the "thinking I":

> [T]he thinking 'I' (...) does not know itself through the categories, but knows the categories, and through them all objects, in the absolute unity of apperception, and so through itself. Now it is indeed very evident that that I cannot know as an object that which I must presuppose to know any object. (...) Nevertheless there is nothing more natural and more misleading than the illusion which leads us to regard the unity in the synthesis of thoughts as a perceived unity in the subject of these thoughts (A402).

And again,

> In attaching "I" to our thoughts, we designate the subject of inherence only transcendentally, (…) without noting in it any quality whatsoever—in fact, without knowing anything of it either by direct acquaintance or otherwise (A355).

There can be seen to be several threads of functionalist thought in Kant; the parallels are clearly there. Firstly, Kant frequently treats concepts, both the "categories of the understanding" and ordinary empirical concepts, as functions that make it possible to transform the raw content of experience into judgments. Secondly, and perhaps more relevant to cognitive science, Kant organises the mind into "faculties", responsible for different phases of the constructive process that takes raw sensory experience as *input* and produces thoughts or judgments as *outputs*. A functionalist interpretation of Kant would focus on the theory of synthesis that contains Kant's division of mental labour into the more elementary tasks that are necessary for consciousness to occur. Kant's views are also compatible with contemporary functionalist theories in the sense that they allow consciousness to be characterised at a high level of abstraction that would allow instantiation into any number of physically realisable systems. They are also compatible with the functionalist endeavour to describe the causal dependencies and relationships between various mental states, as well as system inputs and outputs. This is Kant's theory of "synthesis" which is his division of cognitive labour into more elementary tasks necessary for cognition; where he systematises the mind into functional modes responsible for different stages of the cognitive process that receive the disorganised flux of sensory experience as input which correspondingly produce thoughts or judgments as output. It is therefore reasonable to credit Kant's insights into the nature of cognition as a kind of proto-functionalism and compatible with contemporary functionalist theories. His *a priori* characterisation of our human cognitive faculties certainly gives *prima facie* evidence for this. The unknowability of things in themselves which entails neutrality concerning the underlying composition of the mind means that he would have had to allow that multiple-realisability is at least *open* to intellectual possibility. The Kantian notion that sense experience has to be conceptually ordered if there is to be the possibility of coherent thought is also evident. However, *crucially*, and this is an important caveat, functionalist interpretations of Kant tend to uncouple his transcendental psychology from his epistemology and metaphysics and do not do full justice to Kant's overall strategy. Therefore, it is important to emphasise that this is only a part of Kant's legacy. In fact, insofar as we are talking about the mind of a human being and not a mere machine, some of the crucial ways in which his views depart from

functionalism, illuminated through a proper examination of transcendental method he employed, can continue to provide significant guidance to cognitive science that is obscured by such an interpretation. For functionalism persists as a backdrop to a range of puzzles and obsessions in the philosophy of consciousness that are ongoing, in particular, "the hard problem of consciousness", puzzles which a more fruitful interpretation of Kant can address.

4.3. Beyond Functionalism

So Kant's "transcendental psychology" is regarded by many to be, in several important respects, an early form of functionalism or a *proto*-functionalism, the merit of which is that it allows a better understanding of Kant's theory of mind, whilst simultaneously isolating it from the scorn of analytic positivist philosophers such as Peter Strawson and others who dismiss it as incomprehensible, speculative and unverifiable. Reinterpreted as proto-functionalist, Kant's "transcendental psychology" or his theory of *a priori* synthesis is enabled to enjoy a legitimised status that was impossible during the era of behaviourism and positivism which allowed no discussion about the mind. As Ralf Meerbote writes in the introductory section of his influential book *Kant's Functionalism:*

> Kant's transcendental psychology, often maligned, is a cognitive psychology. More specifically, it is a faculty psychology which speaks of capacities and abilities of various sorts which are needed for empirical cognition. The exercise of such capacities and abilities typically consists in mental actions of several types. An activity-characterization of cognitive mental life is the indispensable core element of transcendental psychology (Meerbote, 1998, p. 161).

Kant characterises human cognitive capacities, apperception, synthesis and *a prior* structure this way, provides *prima facie* evidence and justification for a functionalist interpretation. However, one limitation of this construal is that this does not appear to sit easily with Kant's overall purpose, i.e. his transcendental idealism. As such, interpreting Kant's theory of mind as purely functionalist in nature does him several injustices. One reason for this is that, as a consequence of his transcendental idealism, his transcendental method for analysis of the mind's activities is importantly different from the methods of contemporary functionalism in

several regards. This is in relation to the transcendental unity of apperception, as discussed in Chapter 3. There it was noted that Kant's view was that although we cannot "know" anything about the self in apperception, as it has an unknown nature, it does have transcendental "properties" that we cognise not by being passively aware of the contents of consciousness but by being "active". We are "aware" of our spontaneous synthesising activity, and to be so is also to be aware of an *actor*. There can be no activity without an agent and to acknowledge the existence of an activity is to acknowledge the existence of "something that acts". This suggests that human cognition involves a necessary relation between the elements of cognition and its overall grasp by a human subject or active *agent*. This is an integral relationship, and one which is often described by Kant as "purposive". Another important reason is that this is linked to the notion of human freedom. Kant identifies *spontaneity* as the theoretical analogue of freedom in the discussion of the third antinomy where he discusses the problem of free and natural causality. Here he tries to show that there is no logical contradiction between taking something as causally determined and as free at the same time, insofar as a distinction between the dual aspects of *phenomena* and *noumena* is accepted. The idea from Kant is that causality and freedom refer to two different ways of viewing things. Although under the "phenomenal perspective" mental activities are causally determined, to engage in cognitive activities means to consider ourselves not as "objects", but as "subjects" of experience. This means that we must therefore consider ourselves *apart* from the conditions to which all objects of experience are subject, as rational free agents. Spontaneity is "the mind's power of producing representations from itself" (A51/B75); more generally, it is the capacity for creative mental activity that is either strictly underdetermined or else wholly unconditioned by natural or physical causation (A448/B476). Here the definition of freedom is purposive, the power to *spontaneously* originate behaviour and generate a new causal series. Therefore, Kant's views differ profoundly from functionalism in its direct appeal to spontaneity, a naturally occurring indispensable feature of the mind without which subjective thought would be impossible. Kant terms this power of self consciousness variously as "the original synthetic unity of apperception", "the transcendental unity of self consciousness", the "I think", and "pure apperception".

Functionalism is more or less updated Hume. Hume's bundle of perceptions is held together through the principle of association, with no necessary causal connection between them. However, for Kant, causal interaction immediately links our inner representations with objects outside of us. Without this objective basis for associability, there would be no means of conceptualising experience. You would not be able to represent yourself as a numerically identical subject through different experiences without this. Kant argues in the Second Analogy of Experience, that we can know *a priori* that "[all] alterations take place in conformity with the law of the connection of cause and effect"(A189/B232). For example, that we can determine that one of our mental states follows another in time shows that we have a fundamental concept of cause. For Hume, the order of time is empirically given as a sequence of ideas and impressions and the associations between them, which Kant says is "something merely *subjective*, determining no object, and may not, therefore, be regarded as knowledge of any object, not even of an object in the (field of) appearance" (A195/B240). Causality requires a necessary rule of connection between events which transforms a merely subjective temporal sequence into an objective one. To illustrate this Kant compared the perception of a house with that of a ship moving downstream. There is a series of perceptions changing in time, but the house itself remains unchanged, just as the position of the ship changes, but that does not reflect a change in the subject which is the same throughout (A191-2/B236-7). He says in his proof that the concept of causality is necessary for distinguishing an objective sequence of events in the world from a subjective sequence of perceptions. In the case of the ship, what we apprehend is an objective process, and its order cannot be arranged otherwise than in this very succession; we first see the ship up river and then we see it down river. Also in the case of a seeing a house: "the apprehension of the manifold in the appearance of a house which stands before me is successive" but "no one will grant" that "the manifold of the house is also in itself successive" (A190/B235). The progression of the perceived parts of the house is not due to those parts coming to be and then passing away, but rather to the movement of the eyes in a particular *body* in space and time. (However, this does not contravene the strictures of his transcendental idealism since the causal link is not something we perceive, but is an *a priori* principle, and the transcendental self is never itself directly present as an object of empirical inquiry, since this self is

essentially, or transcendentally, the point of view from which any inquiry can take place).

Thus Kant showed that any unity such that might be said to be a product of sense alone presupposes a "transcendental affinity", that is, the power to distinguish a mere succession of perceptions from a representation of objective succession. Knowing that the successive perceptions of a house do not count as the perception of a succession in the world requires this kind apperceptive unity:

> This objective ground of all association of appearances I entitle their *affinity*. It is nowhere to be found save in the principle of the unity of apperception, in respect of all knowledge which is to belong to me (A122).

From this it is clear that for Kant, the activity of a mind turning its own powers of thought onto itself creates a set of issues that are not found in functionalist accounts. And although modern cognitive science might benefit from the depth of Kant's insights gleaned through the functionalist reading, at least some of the differences evident in his transcendental method can provide guidance to cognitive science that is *obscured* by such a reading. Of particular significance in this regard is that Kant's ideas can shed light on the philosophical perplexities that arise from functionalism itself, especially in relation to the "hard problem of consciousness" (Chalmers, 1995) and the related problem of an "explanatory gap" (Levine, 1983). This is the problem of how physical processes in the brain are supposed to give rise to subjective experience. According to Chalmers, one of the main difficulties with functionalism in describing the mind is that it does not give sufficient justice to qualitative phenomena. This puzzle involves the inner aspect of thought and perception: the way things feel for the subject. In his seminal paper "Facing up to the Problem of Consciousness", he resurrects this ancient puzzle of philosophical perplexity that was brought into sharp focus by Descartes. Chalmers argues that there really *is* an "explanatory gap" from the objective to the subjective, and criticises all scientific explanations of mental experience that eliminate or ignore the subjective aspect of cognition. This has led to the numerous attempts to try to solve what is seen as a real problem, currently considered by many as one of the most important unsolved problems in contemporary science. Twenty or so years ago

neuro-functionalist Thomas Metzinger, commonly regarded as one of the foremost contemporary thinkers and researchers in the philosophy of mind, wrote:

> Today, the problem of consciousness — perhaps together with the question of the origin of the universe — marks the very limit of human striving for understanding. It appears to many to be the last great puzzle and the greatest theoretical challenge of our time (Metzinger, 1995, p. 3).

Twenty years later, despite the tremendous scientific advances since then, the hard problem of consciousness seems as intractable as ever. For decades, the study of phenomenal consciousness had been shunned within functionalist oriented cognitive science; the then prevailing view being that science, which depends on objectivity and third person verifiability could not accommodate something as subjective as that. The principle of objectivity, on which the sciences are based, required the avoidance of all first person perspectives and distancing oneself as far as possible from all individual points of view. All this changed during the latter part of the last century; as John Searle wrote, "raising the subject of consciousness in cognitive science discussions is no longer considered to be "bad taste", causing graduate students to "roll their eyes at the ceiling and assume expressions of mild disgust" (Searle, 1990 p. 585). In fact, not only is consciousness no longer considered a taboo subject, but quite the reverse; it has become a vibrant topic in both philosophy of mind and the sciences: in psychology, neuroscience and even quantum physics. This is mostly down to David Chalmers and his framing of the "hard problem" of consciousness. Because of his arguments many now see fundamental problems with the functionalist theory of the mind and all reductive explanations and ask: How can it be possible that as a means of investigating cognition, its leading characteristic is the elimination of all subjective perspectives? How can that help us to understand the mind? Something crucial is left out, and that is the mind itself. For functionalists the mind just *is* the functioning of the brain, which appears to process information mechanically. But "the hard problem of consciousness" is the question of why, in addition to the information processing that the brain engages in, there is a feeling of "what it's like" associated with it. It is also the related problem of how conscious experience can "emerge" from the grey matter of the brain. Thus, David Chalmers, along with predecessors, Joseph Levine and Frank Jackson, is most responsible for

the outpouring of work on this issue, and of opening up a seemingly unbridgeable gap between the physical world and the realm of phenomenal consciousness. Each made use of different arguments and thought experiments that purport to demonstrate that functionalist/physicalist stories about the mind cannot capture the qualitative features of experience. This caught the imagination of philosophers and scientists alike and thousands of articles and papers have been written which attempt to solve the problem. The association for the Scientific Study of Consciousness (commonly referred to as the ASSC), was set up in 1997 with the aim of encouraging such research on consciousness, through organising annual meetings, and promoting the academic study of consciousness in a number of ways, including the publishing of articles and papers in their open-access journal *Psyche*, now *Neuroscience of Consciousness*, and also in a freely accessible e-print archive of papers. There is also the *Journal of Consciousness Studies*, currently edited by Professor Valerie Gray Hardcastle of the University of Cincinnati, (who views consciousness as a "lower level dynamical structure underpinning information processing in the brain" (Hardcastle, 1995). One of the most famous books published, *The Astonishing Hypothesis*, focused on how the activity of the brain's neurons might give rise to conscious experience - the view that consciousness is correlated with a biological state of the brain, the so-called neural correlates of consciousness or NCCs (Crick and Koch, 1994). There is also the "global workspace" perspective of Bernard Baars (1988), further explicated his book *In the Theatre of Consciousness* (Baars, 1997) and various accounts in terms of higher order states, called HOT theories, discussed in the previous chapter. There are other quite radical solutions, including assigning to consciousness one of the most fundamental and mysterious causal roles in quantum physics, that of "collapsing the wave packet" in physical measurement, where a causal gap within quantum theory makes it an open system into which free choice can enter (Henry Stapp, 2009, 2014) or by assigning quantum mechanics a causal role in brain function through orchestrated collapse in the brain's microtubules, as in Hameroff and Penrose Orch Orr model (1996, 2011, 2014). As pioneering neuroscientist Benjamin Libet, declares in a review of the journal: "there is clearly a need - or a demand - for an interdisciplinary journal devoted to the subject", and that despite the fact it focuses on one topic, "it has the widest range of contributors of any academic journal he has

ever read". This illustrates the extraordinary interest in the quest to find a theory of consciousness, deemed by many as the most puzzling scientific problem of this age.

However, it is proposed that paying attention to the insights of Kant would show that accepting that there is a hard problem of consciousness, at least in the way that is currently stated, is a philosophical mistake. Proponents of a hard problem have more or less substituted the Cartesian conception of mind as immaterial substance in favour of a materialist ontology, and this does not dissipate the inherent problems with the model. On the contrary, these very foundational commitments constitute the very source of the current difficulties. The "hard problem" is not just difficult to answer, it is impossible to answer as it is currently formulated. The way the problem is set up gives rise to dualism, as it involves a reification of the subjective awareness of *qualia* and creates the explanatory gap between subjectivity and information processes. All efforts to "solve" the supposed hard problem and close the gap intrinsically risk both perpetuating its pre-eminence and further guaranteeing its insolubility. In fact, it promotes, in a particularly powerful way, the continuation of the problem, and of keeping it alive indefinitely. The following chapter fleshes out the modern day "hard" problem of consciousness along with the historical and philosophical background from which it emerged, in order to begin untangling the web of conceptual confusion in which it is enmeshed.

5. The Problem of Consciousness and the Explanatory Gap

> It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physicochemical mechanisms: it seems that even the behavior of man himself will one day be explicable in mechanistic terms. There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness. That is, for a full description of what is going on in a man you would have to mention not only the physical processes in his tissue, glands, nervous system, and so forth, but also his states of consciousness: his visual, auditory, and tactual sensations, his aches and pains (J. J. C. Smart, 1959, p. 122).

The above passage, written almost sixty years ago by J. J. C. Smart in his pioneering paper *Sensations and Brain Processes*, marks the beginnings of the re-emergence of the old philosophical problem that was initiated three hundred years ago by Descartes. There had been an unbridgeable abyss between an immaterial *res cogitans* (thinking thing) and material *res extensa* (bodily thing) on Descartes' account, and this chasm has been resurrected today, where is known as the problem of the "explanatory gap" (Levine, 1983) and which is linked to the closely related "hard problem of consciousness" (Chalmers, 1995). Smart was one of the earliest proponents of identity theory. This is the view that all phenomenal experiences correspond to actual neurological states in the brain (such as the interaction of certain neurons and axons). Mental states and events are identical with neurological states and events (literally) inside our brains. For example, pain is literally the firing of c-fibres. The model here is empirical scientific identification: water with H2O, lightning with electrical discharge, and genes with segments of DNA molecules. The major proponents of identity theory were deeply influenced by the logical positivists of the analytic tradition such as Moritz Schlick, Rudolf Carnap, Carl Hempel, and Bertrand Russell whose aim was to overhaul all of philosophy and convert it to a new *scientific philosophy,* and whose mission was to dispense with the mind-body problem and relegate it to the realm of speculative metaphysics, as discussed earlier. Proponents of identity theory thus regarded themselves as champions of the

scientific materialism that strictly limits knowledge to scientific findings and to a methodology free of *a priori* preconceptions and speculation.

From this point of view, the mind literally *is* the brain - they are identical in terms of kinds. David Armstrong was a later proponent of this view. In his influential book, *A Materialist of the Mind* (1968), he made the radical claim that *all* mental states (including intentional ones, i.e. beliefs, desires, etc.) are identical with physical states. The identity theorist, when asked why a certain conscious experience correlates with some physical state, would claim that the experience just *is* a physical state, and that there is nothing further to be explained. However, from this premise there began a period of on-going philosophical dispute. Adhering vehemently to this scientific, materialist stance, that human beings are nothing more than physio-chemical mechanisms, Smart, and to a lesser degree his contemporaries Place (1958) and Feigl (1958), posed a challenge to philosophers of mind with the declaration that the task for philosophy was to work out an account of the mind which is compatible with this view. Thus, Smart helped lay the groundwork for future physicalist theories of the nature of mental states, even those that reject his specific proposal that sensations and brain processes are strictly identical. As Jaegwon Kim notes, "the brain state theory helped set the basic parameters and constraints for the debates that were to come…" (Kim, 1998, p. 2). At the root of identity theory is the following premise: "there is no conceivable experiment which could decide between materialism and epiphenomenalism"; therefore, the statement "sensations are brain processes," although not a straight-out scientific hypothesis, should be adopted on other grounds (Smart, 1959, p.156). Occam's razor cited in support of this claim, that is, the invocation of the scientific rule of thumb which states that when you have two competing theories which make exactly the same predictions, the one that is simpler is the better. Since dualism is the view that there are two kinds of substances or properties in the universe, mind and matter; identity theory that there is one, identity theory wins. So the basis for the materialist position is empirical parsimony: that of the two competing theories, dualism and mind-brain identity, identity theory is the simpler since it commits to fewer entities.

This led more or less directly to the rise of functionalism; since the perceived virtue of functionalist theories of mind was that they avoid some pitfalls suffered by this naïve, simplistic physicalism, as they are based on the premise that cognitive

functions can be characterised on a high level of abstraction that would allow instantiation into any number of physically realisable systems, the thesis of multi-realisability (Putnam, 1967). Hilary Putman convinced philosophers to reject strict identity theory, through his argument that mental states are multiply realisable and that multi-realisability is incompatible with it. This level of abstraction allowed the functionalist account freedom from the problems of strictly identifying consciousness with certain "type" or "token" physical states that constitute it in humans. Token identity is weaker than type identity: type identity is the claim that mental kinds themselves are physical kinds; that mental states such as pain itself, and not merely instances of pain, are identical with physical states such as c-fibre stimulation. Token identity, on the other hand, is the view that some individual instance, a token, of a mental state, a pain for example, is identical with some individual token physical state, whatever that state may be; hence argues that mental events are unlikely to have "steady" or definite biological correlates, e.g. the anomalous monism of Donald Davidson (1970). It is the claim that mental events are contingently associated with some physical property or another but says nothing about that relationship. The brain and mental states are still identical, but not in a way that can be directly typed: the same type of brain state may produce different types of mental states from token to token. As Fodor (1974) declares, token identity is entailed by but does not entail type identity. Functionalism thus avoided such problems as the finding of "the neural correlates of consciousness", i.e. the actual neural representational systems whose contents systematically match the "contents of consciousness" and which cause specific conscious states. Furthermore, it seemed to promise that there are more possibilities for conscious systems than just human ones, i.e. for animal and machine consciousness.

As discussed, the functionalist deals with questions about what the mind does, how its states are related, and what supposedly gives each type of mental state its own identity, whilst ignoring what sort of physical matter there is, or what the system must be made out of. Most functionalists are agreed that one of the biggest advantages of this approach is that it makes it possible to answer questions about the nature of consciousness by explaining that mental states are constituted by their causal relations one to another and to sensory inputs and behavioural outputs. As Jaegwon Kim writes:

> Pain is to be understood in terms of its function as a causal intermediary between sensory input behaviour output, and other mental states……. This view of psychology (…) is, arguably, the received view of the nature of cognitive science (Kim, 1997, p. 580).

However, Kim (1993, 1989) is critical of functionalism precisely because it leads to the re-emergence of the mind-body problem, the problem of finding a place for the mind in a world that is fundamentally physical. For him, *qualia*, or phenomenal mental states such as the visual sensation of seeing red, or the olfactory sensation of the aroma of coffee, cannot be reduced to physical states or processes. [52] He holds a weaker version of functionalism, that something like functionalism or multi-realisability applies to intentional states but not to *qualitative* or mental states. Mental states are also species specific. Instead of pain, there is human pain, dog pain, pig pain, etc. Each such species-specific mental state or event is identical with some first-order physical or material state. Consequently, species-specific mental states are not multiply realisable. For Kim, it is the mental-state *concept*, rather than the mental state or physical event, that is multiply realised. What is significant in terms of this thesis is his famous causal closure thesis: "No physical event has a cause outside the physical domain" (Kim, 1993). No causal chain involving a physical event will ever cross the boundary of the physical into the non-physical: If *x* is a physical event and *y* is a cause or effect of *x*, then *y*, too, must be a physical event. In other words, materialism implies the causal completeness of physics. He writes, "to reject the closure principle is to embrace irreducible non-physical causes of physical phenomena. It would be a retrogression to Cartesian interactionist dualism, something that is definitive of the denial of materialism" (Kim, 1998, p. 47).

The physical causal closure thesis can take many forms, but at its simplest it states that if a physical event has a cause, then it must have a physical cause. A stronger version of the thesis, and one that is more widely held, is that all physical events ultimately have first-order physical causes; the possibility of there being novel, irreducible, second-order physical causal powers is eliminated. In short, the physical causal closure thesis is primarily a metaphysical framework to which materialists and physicalists are committed and which permits no place for what is metaphysically independent of the physical world to have any causal effects. Thus,

the study of the mind based on the functionalist/physicalist theory of mind gives rise to the problem of consciousness, where conscious properties are seen as nothing but epiphenomenal, caused by physical occurrences, but themselves causally redundant having no physical effect on the material world. [53]

The functionalist tries to infer, from a variety of methods, what sort of causal states must be connecting the two externally observable states (input and output). The resulting theory is testable in terms of the predictions it makes about a system's behaviour, and can be modified, depending upon the sort of empirical evidence the investigation reveals about the functioning of the system. Some questions in cognitive science, such as those concerning the ability to discriminate, categorise and react to environmental stimuli, may be explicable to some extent by functionalist methods. They may be able to describe the causal process of what happens when one sees the colour red, smells the aroma of coffee, or remembers a sequence of events. However, this omits first-person phenomenal consciousness itself, which some take as a very significant omission. This kind of thinking led to the formation of the "explanatory gap", the seemingly fathomless abyss between the physical and mental realms. That is, the problem of understanding how something physical like the brain, can generate something non-physical, or how experience arises from this functionalist explanation. David Chalmers, and others, claimed that cognitive science and cognitive neuroscience can explain certain aspects of cognition and define them in functionalist terms as well as neural terms. However, consciousness *per se* cannot be given a purely functionalist definition. Only things defined in terms of functionality can be explained (e.g. the functional mechanism that explains memory function). The problem is that phenomenal consciousness, the "what it is like" to feel a certain way cannot be given a functional definition. In his seminal book *The Conscious Mind*, Chalmers suggests that scientists and philosophers should take consciousness seriously, and that "[t]o take consciousness seriously is to accept just this: that there is something interesting that needs explaining, over and above the performance of various functions" (Chalmers, 1996, p. 215). For Chalmers, even if there is a full discovery of each and every neural mechanism that exists and the functions that are carried out, these mechanisms may explain all kind of psychological functions; the workings of memory, attention and so forth. Among these there will be functions associated with consciousness, such as when one is

experiencing a sharp stabbing pain or feeling hungry, in the sense that the function of pain is to bring one into an awareness that causes one to withdraw from whatever is causing it, or the function of the hungry feeling is to cause one to seek out food. The problem, he claims, is that the neural or functional mechanisms involved only concern transactions or correlation among states and not the phenomenal nature of the states themselves. That is, the neural or functional story fails to capture what is distinctive about pain, or hunger; precisely, the phenomenal sensation of "what it's like" to be in pain or to be hungry, the so-called "hard problem of consciousness". The "hard problem" of figuring out what phenomenal concepts refer to is distinct from the "easy problem" of showing how psychological concepts of causal roles are physiologically or functionally realised. There would be no special challenge either about, say, understanding colour vision, if all that was required was the discovery of which physiological mechanisms realise the psychological colour responding differentially to coloured objects. This is also one of the easy problems; easy problems are easy precisely because they concern the explanation of cognitive *abilities* and *functions*. To explain a cognitive function, scientists need only specify a mechanism that can perform the function. The methods of cognitive science are well-suited for this sort of explanation, and so are well-suited to the easy problems of consciousness. By contrast, the problem of consciousness *per se* (what it's like to experience red) cannot be given a neural or functional definition, and is the "hard problem".

Chalmers' point is that even if there is a full discovery of each and every neural mechanism that exists and the functions that are carried out, these mechanisms may explain all kind of psychological functions, for example, the workings of memory, attention and so on. They may also assist in developing neurophysiologic profiles of the mechanisms of sensory processing, which may serve as valuable biomarkers for diagnosis and monitoring of certain conditions, as well as for the monitoring of therapeutic interventions. However, the neural mechanisms only concern transactions or correlation among states and not the phenomenal nature of the states themselves. That is, the neural story fails to capture what is distinctive about pain, or hunger or thirst, namely the supposed sensation of "what it's like" to be in pain, hungry or thirsty. The neural mechanistic explanation of the supposed pain would work just as well for a "zombie" which could possess the same mechanism to

indicate that it suffers damage (bleeding, bruising, writhing) or exhibit the hunger behaviour, by seeking out food, but be devoid of the subjective feeling of "what it's like" to have the phenomenology. In other words, there could be "no one at home" despite empirical/observable evidence to the contrary. David Chalmers famous "zombie argument" is a thought experiment, the purpose of which is to argue that the neural or functional explanation is blind to the state of consciousness itself: its best explanations can only ever capture input-output causal relationships. In that way, neuroscience may explain certain aspects of experience/consciousness, but it cannot ultimately place consciousness in the material world. This, he claims, is a great unsolved problem in the science of the mind, and also the sciences in general. As he puts it:

> The problem of consciousness, also known as the Mind-Body Problem, is perhaps the largest outstanding obstacle in our quest to scientifically understand reality. The science of physics is not yet complete, but it is well-understood. The science of biology has explained away most of the mysteries surrounding the nature of life. Where there are gaps in our understanding of these fields, the gaps do not seem intractable; we at least have some idea of the direction in which solutions might lie. In the science of mind, things are not quite so rosy. Much progress is being made in the study of *cognition*, but *consciousness* itself is as much of a problem as it ever was (Chalmers, 1996, p. xi).

Chalmers was not original in pointing out that there is something quite peculiar about consciousness. One of the first arguments meant to illustrate the problem of consciousness was given by Thomas Nagel, who claimed that since consciousness is subjective, i.e. directly and privately accessible solely by the person who has it, we are barred from ever understanding it fully, including the question of whether, and if so how, it could be physical. For even if we did know everything there possibly is to know about bat brains, we would never know "what it's like" to be one, simply because a bat's conscious experience would be so very different from our own, as their means of navigating the world are through sonar and echolocation. His famous "What it is like to be a Bat?" paper (1974) provides a now much used and repeated description of phenomenal consciousness; the "what it is like" criterion which aims to capture this subjective notion of being a conscious organism. According to Nagel, a being is conscious just if there is "something that it is like" to be that creature, to be a bat or to be a person. Later arguments put forward by Frank Jackson (1982) and

Joseph Levine (1983), hinged on the claim that this is a serious problem that needs to be taken into consideration if there is to be a cogent account of the mind.

It was Levine who first coined the term "explanatory gap", when he transformed into an epistemological version a modal argument given by Saul Kripke (1980) against the identity theory of Smart, discussed above. This is that the explanatory gap opens up in identity theory since nothing in the physical or functional correlates of a mental state *explains* why this state subjectively *feels* a certain way. A physical theory cannot explain the phenomena of consciousness in the way that it can explain the behaviour of water. Reductive strategies to explain how something feels a certain way seem to leave a gap in the explanation, in that, strictly speaking, such explanations cannot really be *understood.* This notion, that there is somehow an "explanatory gap" between the physical and the mental is seen as leading current challenge to reductionist science.

Chalmers' "zombie argument" is really a reworking of the Cartesian real distinction between mind and body in the *Sixth Meditation* [54] modernised for compatibility with contemporary modal logic, and which has given rise to a new form of dualism (property dualism). It also leads to Cartesian-like attempts to prove the existence of a non-physical fact about consciousness; that there must be something real beyond the known physical world in order to account for it. Chalmers' claim is that the body can exist without consciousness because we can conceive of a possible world containing our physical bodily duplicates but which lack phenomenal consciousness or minds. Levine accepts that Chalmers' zombie argument might show that zombies are conceivable, but denies that this implies their possibility, that they could exist. He argues that "one's ideas can be as clear and distinct as you like, and nevertheless not correspond to what is in fact possible" (Levine, 1993, p. 123). What the zombie argument shows for Levine is the presence of a gap that is not a metaphysical or ontological one, but simply a gap in our understanding. However, Frank Jackson and David Chalmers are "property dualists". Property dualism is the view that non-physical, mental properties (sensations, emotions, beliefs) *supervene* on the material world.

Supervenience is a metaphysical notion introduced into analytic philosophy by R.M. Hare in the 1950s [55] and can be used to formulate relations of dependence between two different domains. According to Chalmers, whereas all high-level facts

about natural phenomena are logically supervenient on the totality of physical facts, those about consciousness are not. That is, physicalism is true in all natural domains except for the mental. Chalmers claims that consciousness supervenes naturally on the physical, but not "logically" and that "psychophysical" laws will explain how these conscious experiences depend on physical processes. Both Chalmers and Jackson view the zombie argument as ontological, rather than an epistemological - that physicalist and functionalist stories about the mind cannot capture the real *qualitative* features of experience. As Jackson puts it:

> Tell me everything physical there is to tell about what is going on in a living brain, the kind of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a rose, hearing a loud noise or seeing the sky (Jackson, 1982, p. 127).

In *Facing up to the Problem of Consciousness*, Chalmers writes:

> Consciousness poses the most baffling problem in the science of the mind. There is nothing that we know more intimately than conscious experience, but there is nothing that is harder to explain. All sorts of mental phenomena have yielded to scientific investigation in recent years, but consciousness has stubbornly resisted. Many have tried to explain it, but the explanations always seem to fall short of the target. Some have been led to suppose that the problem is intractable, and that no good explanation can be given (Chalmers, 1995a).

The "easy" and "hard" problems of consciousness are distinguished thus:

> Researchers use the word "consciousness" in many different ways. To clarify the issues, we first have to separate the problems that are often clustered together under the name. For this purpose, I find it useful to distinguish between the "easy problems" of consciousness and the "hard problem" of consciousness. The easy problems are by no means trivial - they are actually as challenging as most in psychology and biology - but it is with the hard problem that the central mystery lies (Chalmers, 1995b).

According to Chalmers, there are multiple "easy" problems of consciousness, but only one "hard" problem, namely that of explaining *qualia*, or the phenomenal or subjectively felt aspects of experience. The easy problems are easy for Chalmers precisely because they concern the explanation of cognitive *abilities* and *functions*

[and] to explain a cognitive function, "we need only specify a mechanism that can perform the function". He gives the following example, amongst others: "To explain reportability (...) is just to explain how a system could perform the function of producing reports on internal states" (1995a). As he writes:

> The easy problems of consciousness include the following: How can a human subject discriminate sensory stimuli and react to them appropriately? How does the brain integrate information from many different sources and use this information to control behavior? How is it that subjects can verbalize their internal states? Although all these questions are associated with consciousness, they all concern the objective mechanisms of the cognitive system. Consequently, we have every reason to expect that continued work in cognitive psychology and neuroscience will answer them. The hard problem, in contrast, is the question of how physical processes in the brain give rise to subjective experience. This puzzle involves the inner aspect of thought and perception: the way things feel for the subject. When we see, for example, we experience visual sensations, such as that of vivid blue. Or think of the ineffable sound of a distant oboe, the agony of an intense pain, the sparkle of happiness or the meditative quality of a moment lost in thought (Chalmers, 1995b, p. 81).

What makes the hard problem hard, then, is that no mechanical or reductive explanation of a cognitive process, however complex, would be able to account for the further question of why mental processes should be accompanied by felt experiences, the question of why should physical processing give rise to a rich inner life at all? This is a consequence of his commitment to property dualism, discussed below.

5.1. Property Dualism

Genuine property dualism occurs when, even at the level of individuals, the ontology of physics is insufficient to explain what there is. Chalmers claims that since cognitive science and neuroscience do not begin to explain how subjective experience emerges from neural processes in the brain, conscious experience must instead be understood in a new light, as an irreducible entity that exists at a fundamental level and cannot be understood as the sum of simpler physical parts. He proposes that consciousness be understood as a fundamental feature of the universe alongside such ontological categories as mass and space-time. It had once been

supposed that electromagnetism could be explained in terms of more basic mechanical processes. James Clerk Maxwell and his contemporaries realised that this was impossible, and so added electromagnetism to the list of basic elements of reality. Maxwell introduced new fundamental laws to explain the phenomena, as this was the only way it could be explained. The same should be the case for consciousness. Chalmers believes that consciousness is due to "proto-conscious" properties that must be ubiquitous in matter and that "psychophysical" laws, will account for how conscious experience arises out of those properties. He writes:

> On this view, the world still consists in a network of fundamental properties related by basic laws, and everything is to be ultimately explained in these terms. All that has happened is that the inventory of properties and laws has been expanded, just as happened with Maxwell (Chalmers, 1995a. p. 113).

For Chalmers, therefore, property dualism is the view that the world has physical properties which are "casually closed" (Kim, 1993) in addition to "non-physical" mental properties. Chalmers suggests that property dualism would allow the mental to "supervene" on the physical, and "consciousness" to be active in the world.

Some scientists and philosophers think that Chalmers is right to draw a distinction between the "hard" problem and the "easy problems" of consciousness, as if research is to move forward in this area, scientists must be clear that there is a difficulty here. But is this really so? The really hard problems we might think are the ones that the scientists are already dealing with, such as trying to find a cure for Alzheimer's and Parkinson's disease; when these are discovered then these hard problems will have been solved. But what is being talking about here, of course, is a problem of a different kind, a philosophical hard problem, and it is suggested in the spirit of Wittgenstein that certain philosophical problems require dissolving rather than solving. This is because the whole distinction between the hard problem and the easy problems is the result of confusion in the conceptual scheme. As Wittgenstein noted, whenever a question appears difficult and seemingly intractable, this is often due to the fact that its philosophical underpinnings may not be formulated correctly, which requires philosophical, and not scientific analysis. The initial assumptions may need to be reinterpreted in radical ways in light of the analysis.

It is of note that the greatest figures of the first two generations of twentieth-century neuroscientists: Sherrington, Eccles and Penfield, were also Cartesian dualists. There was a distinction between on the one hand, an immaterial mind and the other, corporeal body. The third generation of neuroscientists retained the basic Cartesian structure, but transformed it into brain-body dualism: the Cartesian substance dualism was abandoned, but a kind of dualism still remains. Although few neuroscientists openly endorse Cartesian dualism, a careful reading of most neuroscientific texts, from prominent neuroscientific articles to books intended for a lay audience, reveal implicit dualistic intuition in the sense that they ascribe much the same kind of mental predicates to the brain as Descartes had to mind, and conceive of the relationship between thought and action, experience and its objects, brain and person in much the same way, essentially by merely replacing the mind by the brain. Benjamin Libet's view of intentional action falls into this category, so does Damasio's explanation of vision as the production of mental images in the brain (1994), also some of the theoretical work by Stephen Kosslyn (2005) on mental imagery. Chalmers belongs to this dualist tradition, although he views the mind or conscious experience as neither physical nor material, but a fundamental "property" and also a distinctively non-material entity. However, it is neither obvious, nor even *prima facie* plausible to assume that the two ways in which we grasp physical and phenomenal properties, the former by descriptions of their causal role and the latter by a supposed "rigidly designating" direct awareness, reflect two metaphysically distinct properties. There is an un-argued for assumption of dualism. It is already there in the language used to set up the problem. For example: "How do physical properties *give rise* to phenomenal properties?", or "How could a physical brain made of lumpy grey matter *produce* consciousness?". Since writing *The Conscious Mind* in 1996 Chalmers has established himself as a significant figure in philosophy of mind and metaphysics. His primary weapon against materialism is the aforementioned "conceivability argument", involving a thought experiment about the logical possibility of zombies to illustrate that there is no logical entailment from physical facts to facts about consciousness. Since then this argument has been reinforced against the standard means for debunking conceivability arguments by means of his 2D (two-dimensional) possible world semantics. Two-dimensionalism is an approach to semantics in analytic philosophy, a modal theory of how to

determine the sense and reference of a word and the truth-value of a sentence. The next section examines this in detail.

5.2. Chalmers' 2-Dimensional Semantics

As mentioned above, Joseph Levine had different views of the metaphysical consequences of the explanatory gap than Frank Jackson and David Chalmers. Levine argued that the gap is epistemological, and compatible with the thesis that facts about consciousness supervene on the physical facts, whereas both Jackson and Chalmers argue that the fact that there can be no conceptual analysis of consciousness supports metaphysical dualism: consciousness is neither identical with nor logically supervenient on the physical, but is a further fact. Phenomenal properties naturally supervene on physical properties even though they do not logically supervene. It is significant that both Chalmers and Jackson make use of a two-dimensional appropriation of Kripke's possible world semantics for modal discourse (Chalmers 2002a, 2002b, 2004, 2006, Jackson, 1998). Two-dimensional (henceforth 2D) semantics is a modal framework, introduced by Gareth Evans and developed by Martin Davis used to characterise the meaning of certain linguistic expressions and the entailment relations among sentences containing them. In his 2004 paper "Epistemic Two-Dimensional Semantics" Chalmers seeks to develop a version of 2D semantics which can vindicate the rationalist claim that there are constitutive connections between possibility, a priority and meaning. There he appeals to a framework for connecting modal, epistemic, and semantic issues which he terms the "golden triangle", the basic elements of which are "modal rationalism", Kripkean semantics, and, what allegedly follows from them, the "*a priori* entailment" thesis (AE). Kripke's metaphysics of modality was created from the linguistic, conceptual realm of necessity, the realm in which rationalist philosophers since Descartes have made their home by way of mere stipulation. Rather than standing in for some description, as in Russell's theory, names under Kripke's theory "designate objects rigidly", that is, in every possible world names pick out the same object. He argues that identity statements using alternative names for the same thing have a necessary truth: e.g. "Hesperus is Phosphorus", "Tully is Cicero" and the natural kind identity statement "Water is H2O". For Frege these only have contingent truth, but for Kripke any rigid designators used in an identity statement

are necessarily true, and necessarily true in all possible worlds, even if the statement is not *a priori*. [56] For instance, take the identity statement, "Water is H2O"; this is a necessary truth of physics, but it was an empirical discovery that this is the case, at one time this was not known; hence it is necessary *a posteriori*.

Possible world or intensional semantics is the view that we can model the representational properties of language by assigning intensions to terms and sentences. An intension is any property or quality connoted by a word, sentence or another symbol. Meaning is representational in the following manner: i) the literal meaning of a term or sentence can be equated with how the term or sentence represents things as being in the world, and ii) how a term or sentence represents things as being in the world is encapsulated in its truth-conditions, iii) Truth conditions are truth-value distributions over possible worlds, which are determined by the references or extensions of its terms, across all possible worlds (a possible world is a counterfactual alternative to the way the actual world is). The conclusion drawn by Chalmers is that the property referenced by the term "consciousness" cannot be reduced to any physical property. Consciousness does *not* supervene on the physical, according to Chalmers, because we can imagine another possible world where beings physically and functionally identical to us are not conscious; as mentioned earlier, Chalmers calls these beings *zombies.* According to Kripke's original modal theory, something's being conceivable does not entail that it is possible; however, Chalmers attempts to get around this by noting that "many apparent problems that arise from these Kripkean considerations are a consequence of trying to squeeze the doubly indexed picture of reference into a single notion of meaning or necessity. Such problems can usually be dissolved by explicitly noting the two-dimensional (2D) character of reference, and by taking care to explicitly distinguish the notion of meaning or of necessity in question (Chalmers, 1996, pp. 64-65).

According to Chalmers, mental states, like sentences, are the kinds of things that can have primary and secondary intensions, which seems to imply that mental states are underwritten by bearers of content that are in some sense analogous sentences in a natural language. Chalmers' 2D argument for property dualism assumes that someone in possession of concept *C* has knowledge of how this concept applies in every possible world since they have already grasped the primary intension. The

primary intension represents the narrow, cognitive content of a concept. It captures the motivation for Frege's notion of sense, namely that co-referential terms (such as Hesperus and Phosphorus) may have a different epistemic role, but the same meaning or sense. The concept's application conditions for different scenarios (possible worlds) are *a priori* in that a person knows them in virtue of possessing the concept. The main premise is that although we can be mistaken about the primary intension of the planet Venus, or the natural kind water, we can never be wrong about the primary intension associated with the term "consciousness". We each know the primary intension associated with the term from our own cases. In his words, "[we] can say that a subject grasps an intension when the subject is in a position to *evaluate* that intension: that is, when sufficient reasoning will allow the subject to determine the value of the intension at any world" (Chalmers, 2002b, p. 148). From this he goes on to say that we can conceive a logically possible world in which so-called zombies, beings that are physically like us, but which lack phenomenal consciousness, exist. Otherwise put, there are *a priori* epistemologically possible worlds where the term "consciousness" has an empty extension. For Chalmers, consciousness is a special case where it is not possible that we could be mistaken about the referents of our concepts, about our first-person experiences. A central premise of his argument is that conscious sensations, e.g. "what it's like to feel pain" serve as both the primary intensions and secondary intensions of sensation terms, i.e. they coincide, (in contrast to, say, water, which might not be H2O). The secondary intension (or "property") denoted by a sensation term just is the property of having a certain phenomenal feel. Chalmers' reasoning is as follows:

> What it takes for a state to be a conscious experience in the actual world is for it to have a phenomenal feel, and what it takes for something to be a conscious experience in a counterfactual world is for it to have a phenomenal feel. The difference between the primary and secondary intensions for the concept of water reflects the fact that there could be something that looks and feels like water in some counterfactual world that in fact is not water, but merely watery stuff. But if something feels like a conscious experience, even in some counterfactual world, it *is* a conscious experience (ibid., p. 118).

But it is not so clear how the phenomenal experience of, say, pain is the "primary intension" of pain. Chalmers says he just *feels* that functionalism leaves

something out. He has a gut intuition, something he has sometimes called "direct experience". His method of conceiving his zombie world depends on gut intuitions, what he also calls a "brute intuition" (ibid., p. 96) which he considers is enough to establish that such a world is possible. It is arguable, however, whether a person's "direct experience" or "gut intuitions" can count as evidence or grounding for philosophical conclusions; could a method of conceiving based on such intuitions ever be authoritative enough? Descartes called on the authority of God to guarantee his clear and distinct ideas, which gave rise to the so-called Cartesian Circle. [57] In a similar vein, what might be termed a secular version, Chalmers introduces the notion of "ideal conceivability", by means of which he argues for the logical possibility of Zombie worlds by arguing that they are ideally conceivable, and asserting that something is logically possible if and only if it is ideally conceivable. [58] Or rather, for Chalmers it is ideally negatively conceivable that there is a Zombie world, which means that this Zombie world is not ruled out *a priori*, or that there is nothing contradictory following from this description. The leads on to its logical possibility, for what is negatively ideally conceivable is possible. However, all conceivability arguments rely on the notion of ideal conceivability, and ideal conceivability is a suspect notion at best. One could equally argue that what is conceivable actually depends on what theories one tacitly holds: in this case, finding zombies conceivable might very well count as evidence that one implicitly or explicitly holds a dualistic theory rather than showing what is possible. Moreover, Kripke's idea of *metaphysical, in addition to* logical, necessity is a claim that obviously presupposes that logic *is* connected with metaphysical necessity. However, on Kant's analysis of the demarcation between logical and real (metaphysical) modality, the speculative metaphysician, although they may not commit any logical fallacy, do instead makes an illegitimate transcendental assumption of possible existence, as will be discussed below.

It is suggested that clever, technical rhetoric about rigid designation and necessary truths that apply to all possible worlds, as well as appeals to hunches, intuition and guesses are excuses for not dealing with difficult philosophical questions about the real world. It is difficult enough to figure out what is true in this world, let alone a possible one. There is obviously no reliable way of establishing what *is* true in all possible worlds. An overarching difficulty with modal

epistemology is it is just not clear what *it shows*. The senses can tell you what is going on in the actual world: but how can hunches, intuition and *conceivability* tell you about other possible worlds? Conceivability is, after all, a subjective, psychological and epistemic property, while genuine logical possibility is usually taken to be a mind-independent, modal property, and it is not clear how this gap is to be bridged. Moreover, as Peter Van Inwagen (1998) has cogently stated, it is doubtful that philosophers have capacity to justify modal claims so far removed from everyday life. [59] Chalmers insists that it can be, and has several complex arguments up his sleeve to back up this claim. Chalmers' main argument, in a nutshell, is this: Zombies are conceivable, if zombies are conceivable then zombies are possible, and if zombies are possible, physicalism is *false*. However, in some sophisticated papers on the topic, it has been shown that it is quite simple to neutralise the zombie conceivability argument by its own logic. In fact, a conceivability argument can be constructed that follows precisely the same logic of the zombie conceivability argument, yet reaches the opposite conclusion, i.e. that physicalism is *true*. That is, Chalmers' evidence against physicalism is given in the form of *a priori* conceivability arguments, but there are *a priori* arguments against Chalmers' property dualism of exactly the same variety. Keith Frankish has presented one in an interesting paper on the subject, stating:

> The zombie argument is an elegant and seductive piece of philosophical argumentation. But the idea that we can determine the nature of consciousness by an exercise of the imagination seems too good to be true, and the fact that we can construct an anti-zombie argument suggests that it is not true. When zombies and anti-zombies meet, they annihilate each other, and in so doing reveal that considerations of conceivability have little role to play in debates about the nature of consciousness (Frankish, 2007, p. 15).

Frankish's paper cleverly illustrates that the premisses of both zombie and anti-zombie arguments are equally intuitively plausible. His argument proceeds from the thesis that anti-zombies are conceivable to the thesis that they are metaphysically possible to the opposite of Chalmers' conclusion, viz., that physicalism is true. The argument hangs on the conceivability principle (CP), that whatever is conceivable is metaphysically possible. Property dualism holds that phenomenal properties are extra features of the world over and above the physical, so that consciousness does

not supervene metaphysically on the physical, and physicalism is false. Frankish holds that the converse entailment also holds - that if consciousness *does* supervene metaphysically on the physical, physicalism is true. Frankish goes on to describe anti-zombies as exact microphysical physical duplicates of human beings with all of their phenomenal experiences, but without any non-physical states: "I shall call an object x a bare physical duplicate of an object y if x is a physical duplicate of y and has no further properties of a non-physical kind. Then we can define anti-zombies as beings which are bare physical duplicates of us, inhabiting a universe which is a bare physical duplicate of ours, but none the less having exactly the same conscious experiences as we do". They are exactly like Chalmers' zombies in every way, physically and behaviourally, yet "the lights are on inside" (ibid., p. 4), they have subjective, conscious experience. In the anti-zombie world consciousness is a totally physical phenomenon supervening metaphysically on the physical.

Chalmers' conceivability argument, simplified is: (i) zombies are conceivable; (ii) whatever is conceivable is metaphysically possible; (iii) therefore zombies are metaphysically possible. Frankish's parallel argument is: (i*) anti-zombies are conceivable; (ii*) whatever is conceivable is possible; (iii*) therefore anti-zombies are possible. But (iii) and (iii*) cannot both be true, since if the purely physical facts about anti-zombies make them conscious, then the exactly similar physical facts about zombies make them conscious too, that is, they are not zombies after all (see Kirk, 2015). Frankish writes :"The anti-zombie argument was conceived as a tactical device to neutralize the zombie argument. Its primary function is to show that the CP (conceivability principle) thesis is a two-edged sword and should be rejected" (ibid., p. 14). What this means is that unless we wish to get involved in an endless clash of intuitions about which possibilities should be taken more seriously, we should conclude that this kind of conceivability argument is inconclusive. If we are to understand the nature of consciousness, we need something over and above *a priori* intuitions.

Moreover, there is the question of the analytic philosopher's penchant for thought experiments as a good way of generating intuitive evidence. What evidence would that be? It is not possible to show any logical contradiction in the thought-experiment or hypothesis that zombies exist, because those who postulate it have not made their claim falsifiable. There would be no observable difference between a

world with zombies versus one without such entities. One might even question why a philosopher would go to all those lengths to argue for something that has no impact on the world whatsoever, especially since they are supposed to address the scientific and ontological "problem of consciousness". Intuitions in philosophy are often seen to play the same role that observation does in science, but unlike in science their claim to usefulness is controversial. As Hilary Kornblith, contemporary epistemology's most prominent proponent of naturalised epistemology avers, "philosophy cannot live up to its ambitions" if it continues to emphasise the use of intuitions, since they merely tell us about our concepts (Kornblith, 2006, p. 11).

So, let us recapitulate on what has been discussed thus far: Chalmers' primary intensions are associated with the term "consciousness", and his point is we can imagine so-called zombies, beings that are physically like us, but lack phenomenal consciousness. Expressed alternatively, there are *a priori* epistemically possible worlds where the term "consciousness" has an empty extension, though the world is inhabited by beings physically identical to conscious creatures like us. Put otherwise, our concept of consciousness does not have any *a priori* conceptual connections to any physical concepts; otherwise we could not rationally conceive of zombies. The conclusion that Chalmers (and also Jackson) draw is that the property picked out by the term "consciousness" cannot be reduced to any physical property. There is a question, however, over whether, and/or to what extent, possible world semantics are able to provide a substantive metaphysical grounding for scientific laws concerning consciousness. Kripke's semantics beg the question of what exactly a possible world is, and further, it leaves open where to fit one's epistemology or ontology. Primary intensions, for Chalmers, are supposed to do the work of Fregean modes of presentation that determine reference. Frege had distinguished two aspects of meaning which he called *Sinn* (or "sense") and *Bedeutung* (usually translated as "reference") (Frege, 1892). He had noted that the thing that a word refers to is not necessarily the same as the meaning of the word. To take the paradigmatic example, the planet Venus was once known as Hesperus by those who saw it in the morning and Phosphorus by those who saw it in the evening, yet both groups did not connect the two references to the same object. According to Frege this showed that although these two names have the same reference, Venus, they do not have the same sense. The reference based theory of meaning has been pursued by analytic philosophers

ever since, and the aim throughout has been to find the hook that connects our words to what they mean. For Frege, language serves to represent by means of sentences which say either truly or falsely how things are. The meanings of words are a function not of how things are, but rather, are grounded in the contribution they make to fixing the truth-conditions of those sentences in which they occur. The 2D semantics of Chalmers is a direct descendent of this. Chalmers thinks it is plausible that sentences express entities that are quite closely akin to Fregean thoughts. As was discussed earlier, his 2D semantics entails the view that a sentence has, relative to a context, two sorts of meaning, one of which, its primary intension, is related to its epistemic properties and the secondary intension to its metaphysical or modal profile, which is much like the notions of Fregean "sense". As Chalmers remarks: "An expression's secondary intension (or what Jackson calls its C-intension) is just its familiar post-Kripkean intension, picking out the extension of the expression in counterfactual worlds" (Chalmers, 2006, p. 10).

Thus, Chalmers presents himself as vindicating a Fregean account of meaning and, in fact, uses a framework very similar to the one Carnap uses in *Meaning and Necessity* to define his "modal intensions" in order to provide a "metaphysical foundation" for Fregean senses. His account also embraces the Kripkean notion of rigid designation and necessary truth, and he proposes an analysis of necessary *a posteriori* truths: a sentence is necessary *a posteriori* if it combines a necessary secondary intension with a contingent primary intension. For Kripke, "gold has the atomic number 79" is a necessary truth, rigidly designated in all possible worlds. However, empirical research is required to determine instances of actual gold. Hence "gold is the element with the atomic number 79" is a necessary *a posteriori* truth. In the case of pain, however, it is possible for a person to have a mental property pain without its corresponding physical property of c-fibre stimulation because it is possible to conceive of the situation when the two come apart. This led to Chalmers' anti-physicalist conclusion that there cannot be a relation of *a posterior* and necessary identity between consciousness and the physical world. However, there is the underlying assumption, as there is with all logic, that there is a relationship between names, signs or concepts and referents and the philosophical task is one of explaining how such language connects or hooks onto the world, what is termed the grounding problem (Harnad, 1990). Wittgenstein was very much aware of this

problem in connection with his "picture" theory of meaning in the *Tractatus* stating: "The difficulty of my theory of logical portrayal was that of finding a connection between the signs on paper and a situation outside in the world. I always said that truth is a relation between the proposition and the situation, but could never pick out such a relation" (Wittgenstein, 19e-20e quoted in *Word and World*, 2004.)

As David Papineau cogently claims, the line of thought which Chalmers pursues is a form of fallacy, which he has dubbed the "antipathetic fallacy" (Papineau, 1993). This ability to see the world from two perspectives, does not give rise to the conclusion that there really *are* two metaphysically distinct things. Chalmers talks about subjective experience, the way things feel for the subject, "the ineffable sound of a distant oboe, the agony of an intense pain, the sparkle of happiness or the meditative quality of a moment lost in thought". Although we often feel intuitively in the case of phenomenal experiences that something is left out, we should resist the anti-pathetic fallacy. The "having" of such feelings is just what it is to be in a certain material state when we are in those states rather than something extra over and above them. He writes:

> Chalmers supposes that all terms have a "primary intension", in addition to their referents as normally conceived. This "primary intension" consists of those entities that the term would pick out in other possible worlds "considered as actual" (for example, "water" would pick out XYZ if the actual world's watery stuff were XYZ rather than H2O). Chalmers then assumes that, if the claim that $a \neq b$ is so much as conceivable (for example, water ≠ H2O), this must be because "*a*'s and '*b*'s primary intensions diverge" (there must be worlds in which the terms "water" and "H2O" would pick out different items), from which it follows that there is a genuinely possible world corresponding to the thought $a \neq b$. Applying this to the mind-brain case, we then get the Kripkean thesis that, if it is so much as conceivable that pain ≠ M, where "M" is some material concept, then there must be genuine possibilities where "pain" and "M" pick out different items. Moreover, if "pain" is a priori distinct from *all* material concepts, as the inflationist materialist assumes, then this must mean that "pain" must refer by invoking some distinctively non-material entity (Papineau, 2003, p. 8 n.1).

Papineau denies the first premise: although we can conceive of a possible world where pain does not pick out a material property, this does not imply distinctness. The terms "a" and "b" may simply refer directly, which means that there will not be any "primary intension" that differs from their normal referents. Moreover, the

zombie argument supposes that a physical counterpart of "my" body can be devoid of phenomenal consciousness, and a subjective point of view. But this is conceptually flawed: the zombie argument rests on "an illusion of contingency", the mistaken conceptual intuition that a conscious subject's actual-world body can have a genuine physical counterpart that is *not* subjectively conscious. This is because Chalmers has performed a logical sleight of hand and tacitly shifted an essentially indexical element ("my" body, "his" body, "her" body) to a *non-essentially* indexical rigid designator, David Chalmers' body, viewed as an objective entity in a possible world describable in third person terms through science. According to David Kaplan, the originator of the 2D semantics framework for indexical and demonstrative expressions, *essentially* indexical terms cannot be replaced by descriptive terms, even rigidly designating ones, without loss of meaning. In Kaplan's terminology, the meaning of an indexical term consists in a certain *character,* which takes into account the particular *context* in which it is uttered, in order to deliver an overall *content* to a proposition. "My" or "her" or "his" body in the actual world, therefore, would refer to an actual living and lived body, from which it follows that any physical counterpart in a possible world must also be one, complete with subjective experience and a point of view. For Chalmers, however, a physical counterpart of that body can be devoid of a point of view, as it lacks consciousness entirely. The problem is that he has generalised the 2D semantics framework for indexical and demonstrative expressions that were developed by Kaplan and used it to try to reinstate descriptivism in the philosophy of language. Kaplan (1989a) showed that indexicals such as "me", "you", "he", "she", concern *direct reference,* i.e. that the *content* of an indexical, with respect to a context *c*, is the object to which it refers in *c*; its content is not a *property* or descriptive condition that determines the referent. In other words, that the contents of "I", "he", "she", "you", "that", and similar indexicals, in contexts, are the actual individuals to which those terms refer, in those contexts.

As Robert Hanna and Evan Thompson aver:

> Chalmers secures reference to the relevant actual-world body by means of a term whose semantics includes an element that is irreducibly indexical, and therefore by his own account *not* logically supervenient on the physical facts; and he then tacitly shifts to the use of a referring term lacking this semantic element, a term

> whose semantics *is* logically supervenient on the physical facts. Or in still other words, Chalmers tacitly shifts from treating the relevant actual world body as a *Leib* to treating it as a *Körper"*(Hanna and Thompson, 2003, p.18).

Chalmers has shifted meaning from the actual living and lived body (*Leib)* to a body as an objective material thing that can be picked out in the world and defined third personally by science (*Körper)*. However, it could be argued the body functions as an absolute indexical "here" in relation to which things appear from a point of view or are *inherently* perspectival. The necessary condition of the conceivability argument is that it is logically conceivable that a counterpart of you, having your cognitive capacities could have no feeling of its own body and no pre-reflective awareness of its embodiment, i.e. that experience is causally or explanatory irrelevant to our lived lives in the world. However, this purely analytic argument divorced from the real world is wrong-headed and technically flawed. According to Hanna and Thompson, philosophers should not be allowed to get away with simply *asserting* that the zombie scenario *seems conceivable to them* but need to spell out the scenario that would render such bold assertions intelligible; accordingly, the zombie argument should be rejected outright. Even from Chalmers' own functionalist perspective, the sense perceptual capacities of the body in terms of the realisations of certain representational states, ones that are able to be studied from the third person perspective of science, *necessarily* depend on the subject's lived experience. Every phenomenal experience is not only correlated with a lived kinaesthetic experience of the body but is also functionally tied to that experience and cannot be separated from it. This is because they are co-constituted by a non-analytic necessary equivalence relation between them, one which is also non-reducible. Humans understand the world and successfully live in it because they have individuated objects by means of their bodies, through a wide variety of sensory experiences and from a multiplicity of different points in space and time. Therefore, it is precisely *because* objects achieve their perceptual unity through bodily experience, that a functionally equivalent Zombie world could not exist. In other words, a form of bodily self-experience is a necessary constitutive condition of ordinary perception, therefore there is ultimately no sense in the notion of a completely unconscious being (a "zombie") "whose (functionally defined) perceptual abilities are exactly those of its (physically identical) conscious

counterpart" (Thompson, 2007. p. 233). This is because there could be no supposed "zombie" that could conceivably be *just like us in perceptual abilities*, without also having the same kinaesthetic *experience*. But then, a zombie with bodily self-experience is no zombie at all.

As mentioned earlier, it is often claimed that analytic philosophers can be narrow, trite and superficial in their treatment of philosophical problems. There is often a failure to see beyond the technical methods to which their specialised training best suits them and readily appreciate the bigger picture - in this case to think through the implications of the Zombie argument. What Chalmers' Zombie argument really boils down to is the claim that a physical counterpart of a conscious subject's actual world body could have a bodily life indistinguishable in every respect but lack subjective consciousness. But since phenomenal consciousness entails the lived body, the argument fails. Chalmers' quasi-Cartesian position insists that there is no logical connection between the mental and the physical, between the possession of a particular body and the capacity for consciousness. As Wittgenstein reminds us, the living human body in its everyday environment literally manifests consciousness and cannot be separated from it. He makes the following remark in the *Philosophical Investigations* "Only of a living human being and what resembles a living human being can one say it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (PI §281, see §§282–7, 359–61). This connects consciousness, not with biological life *per se*, but with what manifests or expresses it.

Chalmers maintains that Kant is right that there is a deep link between necessity and *a priority* and uses it to back up his arguments. In fact, however, for Kant any argument attempting to derive properly metaphysical results from logical analysis of modality is doomed to failure. Kant was interested in modality throughout his career, in the *Critique* he talks of possibility, actuality and necessity and in the pre-Critical works he distinguishes real possibility from logical possibility and gives a theory of what grounds real possibility. Kant's theory of real possibility in the *Critique* is largely continuous with the pre-Critical theory. However, whereas his pre-Critical work concerned a metaphysical claim about two different kinds of modality, in the *Critique* it concerns an epistemic claim, and this is that logical possibility is not the same as real possibility. Real possibilities are grounded in

actuality. Contra Chalmers, Kant argues that the mere fact that a concept is logically consistent or logically possible is insufficient for the real possibility of its instantiation. Not only does it not tell us about actual objects, it does not tell us about possible objects either. The mere logical possibility of, say, freedom from contradiction of a concept does not suffice for its objective validity or real possibility. Real possibility would require that the properties or terms that constitute this concept are themselves possible or could be "given" in experience:

> A concept is always possible if it is not self-contradictory. This is the logical criterion of possibility, and by it the object of the concept is distinguishable from the *nihil negativum*. But it may none the less be an empty concept, unless the objective reality of the synthesis through which the concept is generated has been specifically proved; and such proof, as we have shown above, rests on principles of possible experience, and not on the principle of analysis (the law of contradiction) This is a warning against arguing directly from the logical possibility of the concept to the real possibility of things (A596/B624, footnote).

This is also the claim that there are propositions that may be logically possible but not metaphysically possible (See also Kannisto, 2013). [60] Kant would have viewed Chalmers and other modern rationalist philosophers as falling into the same deceptive trap as his predecessors Descartes and Leibniz. They would belong in the category of rationalist speculative metaphysicians who, even though they may not commit a *logical* fallacy, nonetheless would make an excessive, illegitimate transcendental assumption of possible existence. In the Dialectic, Kant pursues a program of transcendental criticism, which champions pure reason's detection and correction of its own excesses, excesses which occur because reason strives to go beyond the bounds of sense in its pursuit of knowledge. In Kant's view, all such attempts to "soar above the world of sense by the mere power of speculation" (A 591/B 619) are doomed to fail. As was discussed in Chapter 3, Kant says in the Paralogisms chapter that this kind of error is due to an unavoidable "transcendental illusion" which is born when "we take the subjective necessity of a connection of our concepts, which is to the advantage of the understanding, for an objective necessity, the determination of things in themselves" (A 297/B 353). This illusion, which the even the wisest of men is not immune from, "does not cease even after it has been detected and its invalidity clearly revealed by transcendental criticism" and is the result of our readiness to mistake the "subjective necessity" of "a certain connection

of our concepts" for an "objective necessity in the determination of things in themselves". For Kant "subjective" necessity, as it concerns the connection of concepts to produce judgements falls under what he would term "logical" modality, whereas objective necessity, which concerns having a reference to objects or things, requires also categorical (transcendental/real) modality. He writes:

> So long as the definition of possibility, existence, and necessity is sought solely in pure understanding, they cannot be explained save through an obvious tautology. For to substitute the logical possibility of the concept (namely, that the concept does not contradict itself) for the transcendental possibility of things (namely, that an object corresponds to the concept) can deceive and leave satisfied only the simpleminded (A 244/B 302).

This division between mere logical or formal truth and transcendental modality is central to Kant's critique of speculative metaphysics, as it marks the distinction between general and transcendental logic (see also Chapter 2.3). For Kant general logic determines the basic structure of the way we think, in that it specifies the forms of judgement on which the categories are based. However, he explains in Part Two of the Transcendental Logic section of the *Critique*, the Transcendental Dialectic, that "general logic, if viewed as an organon [...] teaches us nothing whatsoever regarding the content of knowledge" and is quite indifferent in respect of objects themselves. The goal of the Dialectic is to discover a natural illusion which results from the attempt to apply the principles of the understanding beyond the limits of the understanding to "objects which are not given to us, nay, perhaps cannot in any way be given" (A63/B88). It is worth quoting the passage in full:

> Now it may be noted as a sure and useful warning, that general logic, if viewed as an organon, is always a *logic of illusion*, that is, dialectical. For logic teaches us nothing whatsoever regarding the content of knowledge, but lays down only the formal conditions of agreement with the understanding; and since these conditions can tell us nothing at all as to the objects concerned, any attempt to use this logic as an instrument (organon) that professes to extend and enlarge our knowledge can end in nothing but mere talk - in which, with a certain plausibility, we maintain, or, if such be our choice, attack, any and every possible assertion (A61-2/B86) [My italics].

Moreover, in a tone of admonishment, he concludes:

> Such instruction is quite unbecoming the dignity of philosophy. The title "dialectic" has therefore come to be otherwise employed, and has been assigned to logic, as a *critique of dialectical illusion*. This is the sense in which it is to be understood in this work.

General logic abstracts from all relation to possible objects, therefore cannot offer criteria for metaphysical truth but only for formal truth. That is to say, general logic alone cannot be used to derive metaphysical results. Logic has no content and therefore is not transparent to the ontological structure of reality. Thinking, being exclusively discursive, can provide itself with no content at all, which is given to us only through sensible experience. Kant says, disparagingly, that dialectic is the art of giving ignorance the colouring of a truth, a sophisticated art of empty pretensions or "sophistical illusion" (A64).

Thus, Kant's own version of modality contrasts markedly with Chalmers' whom he would possibly construe as someone well practiced in "the art of giving ignorance the colouring of a truth", or perhaps merely one who confuses the merely "logical" necessity of ideas with the possible objective existence of their objects. Chalmers' version of 2D-semantics seeks to vindicate the rationalist claim that there are constitutive connections between meaning, possibility and a priority. From a Kantian perspective, even if one does accept, for the sake of argument, the complicated 2D framework of primary intentions, and descriptions of zombie world scenarios couched in semantically basic terms for phenomenal properties, logical operators, and so forth, it would tell us nothing about which properties are metaphysically basic, merely something about the conceptual repertoire used to set up the scenario. In other words, being picked out by a semantically basic concept does not make a property metaphysically basic. Therefore, the introduction of metaphysical necessity is gratuitous. There is no reason to presume that anything corresponds to it. This being the case, it cannot be used an avenue for solutions of the mind-body problem. As Joseph Levine, originator of the problem of the explanatory gap writes:

> The fact that one cannot infer from a description couched in terms of all the semantically basic terms but the phenomenal ones to a phenomenal description shows nothing about the metaphysical irreducibility of the phenomenal to the physical, but only that phenomenal concepts must be included in the class of semantic primitives. In other words, being picked out by a semantically basic concept does not make a property metaphysically basic (Levine, 2011).

One could say the "modal rationalism" of Chalmers is akin to that of the "speculative metaphysician" whom, although perhaps not committing any logical fallacy, nonetheless makes an illegitimate transcendental assumption of possible existence. On that basis Chalmers' attempt to derive properly metaphysical results purely from a logical analysis and intuition is found wanting, and his attempts to set up a concordance between epistemic conceivability and metaphysical possibility fails. According to Kant this would amount to no more than a mere "dialectic illusion"; the hard problem of consciousness arises from a kind of category mistake. The illusion that there is one is caused by Chalmers' unbridled faith in the authority of "intuitions" and the pseudo profundity of his many pages of detailed technical rhetoric.

An alternative approach to constructing meaning is found in research in embodied cognition, which will be the topic of Chapter 6, which construes cognition as highly dependent on the physical capacities and actions of a cognitive agent. This abandons the notion that meaning is simply a connection or relation between sense and reference. More specifically, it is a way of viewing meaning as the coordination of action in order to achieve certain goals. There is no need to look for the hooks that ground the reference and referents, the speech acts and what they are about. Instead it concerns the matching of goals and affordances (Gibson, 1979) i.e. what an agent wants to achieve and the opportunities *afforded* by a situation in order that the agent can achieve that goal. On this theory put forward by Gibson, cognitive scientist Arthur Glenberg (1997) proposed that cognition evolved to coordinate effective action; that is, action that enhances survival and reproductive success given the constraints of a particular type of body. In this way cognition is naturalistic, involves no metaphysical concepts such as supervenience, and is not fundamentally different from perceiving and acting, but is dependent on the body, its sensory motor systems, action, and on context in the real world. In fact, the features of cognition are so deeply dependent on characteristics of the physical body of an agent, such that it plays a co-consitutive role in an agents cognitive processing (Barsalou, 2003, 2008; Clark, 2011; Laakso, 2011; Schubert, & Semin, 2009; Stapleton, 2013). That is to say, if we are to understand cognition, it is more fruitful to think beyond the inner and outer distinction of mind/brain and world found on traditional accounts and to consider the natural unity that already exists between an organism and the

environment in which it is actively engaged. This way the hard problem is dissolved; living and lived experience is the point at which a theory begins.

Analytic, functionalist, traditional cognitive science construes meaning as something that arises from the syntactic combination of symbols and expressions that are arbitrarily related to that which they signify or refer. Chalmers views the cognitive system this way, in functionalist terms, in terms of semantics and information processing, and proposes that there is an explanatory gap which needs explaining through his semantic theory and naturalised metaphysics. But the traditional, functionalist picture will automatically generate the question as to how "information processing in the brain" connects to the real world, and then the philosophical task will be one of explaining just how it does so. Part of Chalmers' solution is that psychological words, such as: pain, belief, redness, have two completely different meanings - one where it refers to a non-conscious functionalist process and one where it refers to a state of consciousness. But as Wittgenstein reminds us, an embodied view of linguistic meaning conceives of meaning as what people do with language, about language use. Meaning is not reducible to a purely hypothetical construct (i.e., semantics) able to be studied from a formal objectivist perspective. Rather, embodied experience is a crucial part of linguistic meaning, because it is understanding or grasping the meaning of a linguistic expression, rather than being in a particular mental state as the result of an internal representation, that prepares people for situated action (Barsalou, 1999, 2003, 2008). On this view, meaning includes the perception of physical objects, physical events, the body, and other people in interaction in a real world. The meaningful use of language includes both a depiction of what has happened, potential perceptions and embodied actions that may take place in the future. Put otherwise, linguistic meaning is *essentially* embodied, not only in the sense of what has happened, but in the sense of what is likely to occur next in a discourse situation. The primary problem with views of Chalmers and other analytic functionalists is that they conceive of meaning, as well as human cognition more generally, in terms of abstract and disembodied symbols or representations, and the mind as a syntactic engine that operates on them. It is this very fact that leads almost inevitably to property dualism. But this is to ignore the fundamental problem of how meaning is grounded in ordinary experience (the symbol grounding problem, Harnad 1990).

5.3. Functionalism plus Qualia

Chalmers advocates the idea that the basic furniture of the world (its ontology) should be expanded in order to include experience as a new fundamental "property" besides the material stuff that is amenable to third person science (Chalmers, 1996, pp. 111-12). New laws should accordingly be formulated and enabled in order to be able to describe the relationship between the phenomenological/subjective and physical/objective features of the world. Moreover, Chalmers, as a functionalist, virtually equates the terms "functionalism" and "physicalism" in his ontology of mind. Everything apart from consciousness is physical, yet consciousness is taken to be a natural phenomenon, falling under natural laws. If so, he argues, then there should be *some* correct scientific theory of consciousness, whether or not we can arrive at such a theory. This is what results in the "hard", seemingly intractable, problem; how to close a deep and puzzling explanatory gap. But with a little more care we might discover that it is the basic assumption that all mental processes are neural-computational that is the real problem, and the reason that it seems intractable is that Chalmers (and others) may be begging the question by presupposing that the sort of cognition that is being alluded to when considering thoughtful human activity is the sort that should be characterised in a mechanistic, "functional" or "information processing" manner.

Within cognitive science there have been many significant studies of components or aspects of cognition, to do with, for example: memory, observational learning, colour discrimination, decision making, reward prediction learning, attention control, etc. which have been made by modelling the various experimental results using ever-more sophisticated computer programs. In this manner progressive inroads have been made into gaining a better understanding of these various aspects of cognition. The computer metaphor has also been transferred over the neuroscience where scientists have defined the fundamental features of the brain and its information-processing capabilities in terms of computations' mean firing rates at points in the brain cortex (neurons). In fact, the prevailing view in neuroscience is that neurons can be considered fundamentally computational devices. In operation, such computationally defined neurons effectively sum up their input and compute a complex nonlinear function on this value; output information being encoded in the

mean firing rate of neurons, which in turn exhibit narrow functional specialisation. That is, the idea of a neuron as simply a computational device has crept into neuroscience from the implicit adoption of a computational theory of mind, and with it a concomitant functionalism with respect to the instantiation of cognitive processes. This computational understanding of mentality is what inevitably leads to the non-reductive functionalism with respect to consciousness that Chalmers, and others, adhere to. As Chalmers writes:

> When we think and perceive, there is a *whir of information processing*, but there is also a subjective aspect. As Nagel (1974) has put it, there is something "it is like" to be a conscious organism. This subjective aspect is experience. When we see, for example, we experience visual sensations: the felt quality of redness, the experience of dark and light, the quality of depth in a visual field. Other experiences go along with perception in different modalities: the sound of a clarinet, the smell of mothballs. Then there are bodily sensations, from pains to orgasms; mental images that are conjured up internally; the felt quality of emotion, and the experience of a stream of conscious thought. What unites all of these states is that there is something it is like to be in them. All of them are states of experience (Chalmers, 1995) [my italics].

What is significant here is that Chalmers writes that when we think and perceive there is a "whir" of information processing, and perhaps this is the clue as to the root of the misconception. Proponents of the hard problem take the unconscious whirring of neural cogs as being sufficient to explain everything that humans do. "Consciousness" is then defined as something else - an epiphenomenal veneer or accompaniment on top of the real neurological functions. This gives rise to the hard problem of consciousness and the question of why these whirring informational processes should be "accompanied" by any conscious experience, sensations, or feelings of "what it's likeness". But, to repeat, the real problem is that Chalmers (and others) are begging the question by presupposing that the sort of cognition that is being alluded to when considering thoughtful human activity is the sort of thing that should be characterised in a mechanistic, "functional" or "information processing" kind of way. It is suggested that the "information processing" model of cognition is to blame for the confusion. That is, it is a serious mistake to construe the mind as a machine, whether it is a "meat machine" or "metal machine"; in fact, the brain and the whole organism in which it is embedded, although explicable in

mechanistic terms, may not be like a machine at all. Indeed, as has been discussed, functionalism was not motivated by investigations into the way organisms work, but by the knowledge of how computers do and the hypothesis that this could be generalised. It is suggested that the computer/machine metaphor was uncritically adopted as a model of cognition because it linked in with a certain narrative made popular by the rapid and advances in computer technology.

At the crux of Chalmers' non-reductive functionalism is the Principle of Organizational Invariance, which asserts that "given any system that has conscious experiences, then any system that has the same fine-grained functional organization will have qualitatively identical experiences" (Chalmers, 1995, p. 232). Consciousness does not arise from anything other than functional organisation - the only relevant properties of a system in determining its state of consciousness are functional, information processing ones. The argument for this is presented in the form of a *reductio ad absurdum*. He imagines how he could replace himself with a silicon copy and provides a couple of thought experiments on that basis:

> We can imagine, for instance, replacing a certain number of my neurons by silicon chips. In the first such case, only a single neuron is replaced. Its replacement is a silicon chip that performs precisely the same local function as the neuron. We can imagine that it is equipped with tiny transducers that take in electrical signals and chemical ions and transforms these into a digital signal upon which the chip computes, with the result converted into the appropriate electrical and chemical outputs. As long as the chip has the right input/output function, the replacement will make no difference to the functional organization of the system. In the second case, we replace two neighboring neurons with silicon chips. This is just as in the previous case, but once both neurons are replaced we can eliminate the intermediary, dispensing with the awkward transducers and effectors that mediate the connection between the chips and replacing it with a standard digital connection. Later cases proceed in a similar fashion, with larger and larger groups of neighboring neurons replaced by silicon chips. Within these groups, biochemical mechanisms have been dispensed with entirely, except at the periphery. In the final case, a chip has replaced every neuron in the system, and there are no biochemical mechanisms playing an essential role. We can imagine that throughout, the internal system is connected to a body, is sensitive to bodily inputs, and produces motor movements in an appropriate way, via transducers and effectors. Each system in the sequence will be functionally isomorphic to me at a fine enough grain to share my behavioral dispositions. But while the system at one end of the spectrum is me, the system at the other end is essentially a copy of a silicon robot (ibid., p. 237).

This first thought experiment concerns *fading qualia*. We can imagine a continuum from Chalmers with full conscious experience to the Robot in which consciousness is absent. According to this principle, what matters for the emergence of experience is *not* the specific physical makeup of a system but the abstract pattern of causal interaction between its components. This purports to show that a creature whose causal patterns in its cognitive system are identical to those of a conscious person, but whose neurons are gradually replaced by silicon chips, will continue to make statements about his experiences which are identical to those of the conscious person, though his experiences will become different if he gradually loses consciousness as the replacements takes place. He concludes the fading qualia thought experiment by stating that it supports his theory that consciousness results from *organizational invariance*, a specific set of functions organised in a particular way. For Chalmers, systems which have identical experiences are *functionally isomorphic*; they could be water pipes, aliens or the population of China. [61] This is precisely what leads to epiphenomenalism, a kind of one way dualism in which consciousness is a by-product generated by functional processes in the brain but without itself being able to exert any causal effect on those processes. He writes, "We might put this by suggesting as a basic principle that information (in the actual world) has two aspects, a physical and a phenomenal aspect" (ibid., p. 286). As has been argued, this model of human mentality has led to the so-called "mystery" of the hard problem of consciousness, one supposedly in need of a scientific explanation. On this view, if you set out the premisses from the start, that phenomenal experience emerges from and/or accompanies and/or is correlated with neural functions, then the natural question will be to ask, as Chalmers does "*why is the performance of these functions accompanied by experience?"* (ibid., p. 8). Why should physical processing give rise to a rich inner life? This question seems almost a natural consequence of this model or understanding of the mind.

Chalmers is not alone in thinking that consciousness is something "extra" over and above the physical features of reality. Even cognitive scientists and theorists with explicit antagonism towards dualism, those who disavow it, often betray themselves when they start talking about the physical processes which "generate" or "cause" it, or "give rise to" or "are correlated with" it. These phrases may seem innocent enough, but they implicitly presuppose that conscious properties are some

extra feature of reality, which gives rise to a separate realm in need of explanation. This is to regress to the scientific, mechanistic perceptual model of Descartes and Hume, as discussed in Chapter 3, where the difficulty for both was in assuming that the "I" or the mind functions like the objects of "outer sense" or perception. Hume's difficulty arises from the fact that he seeks for evidence of a subject, an impression from which the idea of a self can be derived. Descartes' error lies in supposing that in introspection one is aware of "a thinking thing". Thus, the mistake of both Descartes and Hume derives from the supposition that consciousness of self must be an experience of something extra. Descartes inflates this idea into a reality, a *res cogitans*, a thinking thing or "substance" whereas Hume, finding no impression of it concludes that it does not, in some sense, "exist" or if it does there is no more than the bundle of impressions. Humean scepticism and Cartesian dualism are based on the perceptual model and the implicit assumption that in introspection we must be provided with extra facts about ourselves and because we are not we either assume we are non-bodily entities (souls/substances) or that we do not in some sense "exist" save as an idea of a bundle of impressions. This understanding of mentality continues today and has resulted in the object-metaphors of analytic philosophy of mind. For Chalmers, consciousness is understood as if it were a kind of entity. He says that "a theory of consciousness requires the addition of *something* fundamental to our ontology (Chalmers, 1995a, p. 210). "Scientists introduced electromagnetic charge as a new fundamental entity and studied the associated fundamental laws. Similar reasoning should apply to consciousness". Consciousness is a "thing" (which we know as a brute fact) with properties in the way a table or a chair is an object with properties. He says that consciousness is "presented" to him, as if consciousness were an object in the visual field which we can come into contact with and manipulate; consciousness "exists", just like an object in the world exists. He "experiences it" as if consciousness were something objective that we experience in everyday life. He asks why cognitive functioning is always "accompanied" by consciousness, as if consciousness were some extant object which can either accompany us or not. "Even if every behavioural and cognitive function related to consciousness were explained, there would still remain a further mystery: "Why is the performance of these functions *accompanied* by conscious experience?" (Chalmers, 1995b. p. 201) [my italics]. It is this additional conundrum that makes the

hard problem hard. Consciousness is understood as a thing which exists and has certain properties such as ineffability, privacy, phenomenal feel.

It is contended, however, that this "hard problem of consciousness" is simply an artefact generated by the twin presuppositions: that i) consciousness is "something extra" and that ii) it is an effect of an underlying cause. There really is no hard question needing an answer as to why functional processes in the brain give *rise* to consciousness, as Chalmers insists. The related problem of closing the "explanatory gap" is based on two further assumptions: (i) that there *is* actually a gap between how we experience the world, our subjective or phenomenal consciousness, and the scientific explanation of the material causal mechanisms and forces that underlie it which constitute nature; and (ii) that the gap can be closed through an overarching scientific explanation. From a Kantian perspective, however, the gap does not exist. This is because there is an intimate reciprocal relationship between phenomenal consciousness and the intelligent thought and activity of human beings. The human capacity for conceptual understanding or thought is so inextricably bound up with the very capacity for phenomenal consciousness, that each necessarily presupposes the other and are mutually entwined. Phenomenal experience, far from being an accompanying factor, is a *constitutive* part of human cognition. For Kant the original consciousness of the identity of the self is at the same time a consciousness of the world. Conversely, the same consciousness which reveals the world to us also reveals the self, the self becoming conscious of itself by seeking out and bringing into consciousness its own contents. It is through synthesising the manifold of spatial conceptions, that we gain knowledge of things outside us and it is through this knowledge of outer things that the self can know itself. He writes:

> I am just as certainly conscious that there are things outside me, which are in relation to my sense, as I am conscious that I myself exist as determined in time (BxIi, note).

Thus, for Kant, the world both requires and guarantees the subject, which means that a phenomenal experience, something "it is like", to be in pain, taste a lemon, or smell a rose, is not simply a "this," but rather, a "this-such-for-me," by which is meant an experience which has already been conceptualised for a subject. Understanding something (it becomes an object for me) is only possible because it is classified in some way, by an act of judgment which brings the intuition (perceptual experience) under the concept. In the Transcendental Deduction, Kant declared that

objects can only become objects for a subject insofar as they conform to the conceptual structures with which we cognise them. As he puts it, we must "make our intuitions intelligible, that is, to bring them under concepts" (A51/B75). The consequence of this epistemology is that it is quite erroneous to suppose that the ascription of genuine thoughts is at all possible to an entity without the capacity to enjoy conscious experiences; therefore ascribing states with phenomenal or qualitative character to what are in effect "information processing machines" results in a conceptually flawed model. The informational or functional "states" of a reductionist, functional mechanistic account of human nature do not possess conceptually expressible content; but the beliefs, thoughts and judgements of human beings certainly do. Drawing on the insights of Kant, there is a strong disanalogy between a functionalist information processing model of the mind and the real experience of a living, human being. The Kantian principle is that our human conceptual capacities, those of our ordinary concepts as well as those relating to such relatively abstract concepts as number and time, are so intimately tied up with the human capacity for perceptual experience, that each entails the other. Viewed this way, the hard problem of consciousness, which is essentially the question of "Why should neural processes be "accompanied" by conscious experience?" is redundant. For the Kantian account *presupposes* consciousness and first person subjectivity, as it begins from the assumption that there is *already* "something it is like" to be a conscious, and cognising agent and gives an account of the necessary conditions for this.

The problem with the reductionist information processing model of human cognition is that it is conceptually confused from the start, and this is what generates the problem of consciousness, with its attending notions of *qualia* and the "hard problem" of "what it's likeness". It also leads to the difficulty that "consciousness" has merely an epiphenomenal role to play, as a peculiar, "tacked on", causally redundant feature of human cognition. According to Chalmers and other functionalists or "information processing" theorists, everything about human cognition apart from for the fact of *qualia* can be or will one day in the future be explained in reductive (computational or neural) terms, the easy problems. But this leaves "the hard problem" of explaining the *qualia* or phenomenal experience. However, the notion of functionalism and information processing is far too

impoverished to be of use in characterising the conceptually articulated structure of human thought and its intimate relation to our capacity for phenomenal consciousness. As Kant shows, although both thought and perceptual experience are necessary for cognition, these two aspects of human cognition are inextricably interrelated. Thought is more than just information-processing and perceptual experience is more than simply the subjective experience of phenomenal *qualia*: *both* are states which are conceptually articulated, necessary and sufficient conditions for consciousness and, each depends for its possibility upon the other.

Let us recapitulate on what has been said thus far: According to Chalmers, the "hard" problem is this: "Why doesn't all this information-processing go on "in the dark", free of any inner feel?" (1995b, p. 203). Since he believes that human thought and cognition in general are just a matter of "information-processing", of a sort which could in principle go on in a mindless computer, he is left with the idea that all that is really distinctive about consciousness is its qualitative or phenomenal aspects (the "what it is like", or "inner feel"), which leads to epiphenomenalism, and to the mysterious puzzle of why should we possess this sort of consciousness *on top* of all our capacities for thought and understanding; these capacities being, for Chalmers, simply capacities for certain sorts of information-processing and storage. Whatever a computer can do by way of information processing, is not by any means to be confused with what a real thinking human being can do. Here again is the spectre of the Cartesian reductionist idea that phenomenal consciousness is an epiphenomenon of cognition: due to the narrowness of reductionist science, and blindness to the consideration that human cognition may not a matter of computation at all but, on the contrary, thoroughly integrated within our own natures as embodied creatures, both acting upon and being acted upon by our physical environment and our social engagement in the world. An object of experience, for Kant, is not a mere "this," not merely a given intuition, but a "this-such-for-me," an intuition which has been conceptualised. We understand or become aware of something (it becomes an object for us) when we can classify it in some way, e.g. as a figure or a triangle, or an equilateral triangle, by making a judgment which brings the intuition under the concept. This requires the unity of apperception, that which unites the thought. As he writes:

> That the "I" of apperception, and therefore the "I" in every act of thought, is *one*, and cannot be resolved into a plurality of subjects, and consequently signifies a logically simple subject, is something already contained in the very concept of thought (B407).

A failure to appreciate the special and *a priori* character of the "I think" misleads us into thinking that there is a deep and puzzling explanatory gap for consciousness. But this is a cognitive illusion. There is no such gap and no such puzzle.

In concluding this section, those involved in the scientific study of consciousness have claimed that conscious properties are material properties, firstly that they are strictly physical properties; that mental states are brain states, the identity theory of Place (1956), Smart (1959) and Feigl (1958). In protest at functionalism's eliminative strategies denying consciousness, later theorising was motivated by a more or less explicit dualism, of conscious properties which supervene on strictly physical properties. Supervenience can be understood on many levels, but at root it describes a relation of dependence between two properties, a set of higher level properties supervenes on a set of lower ones if the higher level properties depend on those lower level ones. Following this, Chalmers argued that consciousness is not logically supervenient on the physical in the sense that "all the microphysical facts in the world do not entail the facts about consciousness" (1996, p. 93). Chalmers argues explicitly that conscious properties are extra to any physical properties, and that the task of a theory of consciousness is to discover which physical process *give rise* to this extra realm of higher conscious properties. But the central question here is whether consciousness and its "properties" is an extra ingredient that we have *in addition* to our abilities to perceive, think or feel or is it an *intrinsic part* of being a human animal that can feel think or perceive? The contemporary world-view on the nature of cognition is a legacy of a scientific and philosophical tradition which has travelled down historically from Descartes, and has resulted in scientists and philosophers treating conscious experience as tractable in reductive terms, (identity theory) or as a process that has a mind-independent qualitative content, like mass in physics, and which should be taken as a fundamental feature of the world (Chalmers, 1995). As discussed, Chalmers assumes the mind is explicable in terms of a materialist, functionalist theory operating within the normal

laws of physics and he sees no particular barrier to the successful creation of consciousness in a computer. He writes, "Epiphenomenalism may be counterintuitive, but it is not obviously false, so if a sound argument forces it on us, we should accept it" (Chalmers, 1996, p. 159).

However, it is not at all clear that we should accept these premisses. Problems often arise when scientists and philosophers try to resolve a paradox on its own terms, rather than question the foundational presuppositions that make it unavoidable. The presuppositions for the existence of a hard problem of consciousness are not theory-dependent and not neutral to the topic, at all. In fact, the truth of the key intuitions, that something is "left out" presupposes the very existence of a hard problem, which cannot then be used to establish its existence. Each of the thought experiments which claim to establish the existence of a hard problem depends on certain intuitions and when we ask under what conditions these could be true, it turns out that they are true if the structure and function of mental states is insufficient to account for our phenomenal states. Chalmers claims that it is a *conceptual* point that functional explanations leave out consciousness, but it is a conceptual point only under the particular conception of consciousness he adopts, i.e. the "something it is likeness" criterion. Thus, he assumes a conclusion about the nature of consciousness, which although intuitive and widely accepted by many, is hardly robust enough to form a sound basis for a theory. He offers a *prima facie* simple critique against functionalist/reductionist explanations of consciousness; since we can conceive of it, there is the logical possibility that a world exists identical to our own in every way but which is populated by experience-free zombies. But the simplicity belies a vast complexity of detail and makes use of the conceivability principle, one of the foundational principles in analytic philosophy. It is of significance that this principle has been used to support a wide array of disparate and controversial philosophical positions. We can indeed imagine a world that has the same physical features but without consciousness, a world where zombies exist. However, in order to do so we have to also imagine a change in the laws of nature, for if consciousness *is* a physical feature of brains then it follows that the absence of it is also a change in the physical features of the world. That is, his argument works to establish property dualism only if it assumes consciousness is not a physical feature. In other words, the argument only works by assuming that

consciousness is not a physical feature, but that was the very conclusion it was initially meant to prove. The zombie argument will work to establish property dualism only if it assumes consciousness is not a physical feature, but this is begging the question at issue, for this was the very proposal the argument was supposed to establish. So, rather than posing the question of whether consciousness and its "properties" is an extra ingredient that supervenes on the physical; a "something extra" that we have in addition to our abilities to perceive, think or feel, why not think, along with Kant, of consciousness as an *intrinsic part* of being a human *animal* that can feel think or perceive. To explain cognition is not just specifying a mechanism that can perform a certain function. The functionalism-plus-*qualia* approach fails to provide an adequate explanation of mind because phenomenal experience is a constitutive part of human cognition and not an accompanying factor. Chalmers calls his position "naturalistic dualism"; "dualism" because, like Descartes, he views the mind as neither physical nor material, but fundamental.

According to Chalmers, the whole story about human cognition *apart from* the fact of *qualia* can be explained in reductive (computational or neural) terms. But this leaves the problem of explaining *quali*a, the "what it's likeness of experience". However, for Kant, this is to misconstrue cognition and its intimate relation to concepts. Kant's insight is that genuine thought, with real conceptual content, is only available to creatures with a capacity for perceptual experiences bearing not only intentional (that which perhaps *can* be captured in functionalist terms) but also phenomenal content, and that both imply the other. The notion of information processing is just too impoverished to be of use in characterising the conceptually articulated structure of human thought and its intimate relation to the capacity for phenomenal consciousness. Our everyday experience of the world as being meaningful is inseparable from our experience of it as looking (sounding, smelling, etc.) particular ways, captured in Kant's well-known and widely-quoted phrase: "thoughts without content are empty, intuitions without concepts are blind" (A51/B76). Chalmers' characterisation of the mind as involving information processing of the sort that could go on in a computer brings about this strange notion of "leaving something out", the omission of immediate subjective experience itself. It then seems like a mystery how we could have this *on top* of our ability to process information. He writes:

> Experience is the most central and manifest aspect of our mental lives, and indeed is perhaps the key explanandum in the science of the mind. Because of this status as an explanandum, experience cannot be discarded like the vital spirit when a new theory comes along (Chalmers 1995, p. 206).

There is a distinct parallel here to the Cartesian *Cogito Ergo Sum*. Immediate subjective experience is the point at which all of our attempts to understand the world begin, because it is directly given to us as a brute fact and is "incorrigible". This understanding of mentality also takes cognitive content to be mediated or inferential and leads to the view that there must be something called *experience* which is supposedly different from the things "out there in the world" being experienced. Because this assumption is so persuasive, and Kant warns us that it is a natural illusion that "[e]ven the wisest of men cannot free themselves from" (A339/B397) it usually becomes a self-fulfilling prophecy for those who accept it who end up seeing the world in these Cartesian/Chalmersian terms. For Kant, however, this way of understanding is to misconstrue cognition and its intimate relation to concepts. As Wilfrid Sellars has persuasively argued in his important and influential book *Science, Perception, and Reality,* against what he terms "the Myth of the Given":

> (...) instead of coming to have a concept of something because we have noticed that sort of thing, the ability to notice a sort of thing is already to have the concept of that sort of thing and cannot account for it (Sellars, 1963, p. 176).

We do not come to have the concept of subjective experience because we have introspected and "noticed" that we have subjective experience. Rather, the ability to note that we have a subjective experience is already to have the concept of it.

Those involved in the scientific study of consciousness have claimed that conscious properties are material properties, as in the identity theory of J. J. C Smart, Place, Feigl and Armstrong (brain states), or functional properties which supervene on strictly physical properties. But as has been discussed, the central question then became whether consciousness and its "properties" is an extra ingredient that we have in addition to our abilities to perceive, think or feel or is it an intrinsic part of being a human animal that can feel think or perceive? The former way of conceptualising the problem gives rise to the explanatory gap. It has been argued that

proponents of a hard problem have substituted the Cartesian conception of mind as immaterial substance in favour of a materialist ontology, and this does not dissipate the inherent problems with the model. On the contrary, these very foundational commitments constitute the source of the current difficulties. The "hard problem" is not just difficult to answer, it is impossible to answer as it is currently formulated. It has been noted that the very language used to set up the problem is dualistic. For example: how do physical properties in the brain *give rise* to phenomenal properties? We all experience an intuition of mind-brain (or body) distinctness, and it is this "feeling" or intuition which lends a spurious plausibility to the arguments against functionalism, materialism and reductionism that there is an explanatory gap. Chalmers argues that the existence of an explanatory gap justifies the ontological conclusion that phenomenal consciousness is a fundamental property *in addition to* the fundamental physical properties - thus begging the question. He has moved from the existence of distinctive non-physical ways of thinking, to the ontological conclusion of the existence of a non-physical property. As mentioned earlier, the very idea of a hard problem to be solved depends on the combination of these twin notions: (i) that phenomenal consciousness is something left out of the scientific explanation that needs explaining over and above the performance of various function and (ii) that there is somehow an "explanatory gap" between the physical and the mental that needs to be closed. Although there may be a *logical* sense to the conceivability argument; we can imagine a scenario or possible world where there could be my Zombie twin, a being physically like me with "no one at home", this does not mean it is metaphysically possible. As was argued in Section 2, if it is at all logically possible that such a doppelganger exists, she will share all my physical properties and therefore my conscious properties, i.e. not be a zombie at all. There is ultimately no sense in the notion of a completely unconscious being whose functionally defined perceptual abilities are exactly those of its physically identical conscious counterpart, simply because there could be no supposed "zombie" that could conceivably *be* just like me in perceptual abilities, without also having the same kinaesthetic *experience*. Moreover, even accepting the logical possibility of zombies, it turns out that the question of whether zombie are "accessible" (as used in possible world semantics) to our world is equivalent to the question of whether physicalism is true at our world. As Gualtiero Piccinini states:

> By assuming that zombie worlds are accessible from our world, proponents of the zombie conceivability argument beg the question of physicalism. In other words, it is a mistake to assume that the metaphysical possibility of zombies entails that physicalism is false at our world. This would seem to indicate the futility of such reasoning in coming to an understanding the nature of consciousness (Piccinini, 2015).

As Kant would diagnose it, such a problem of consciousness rests upon a misrepresention or "subreption" of ideas concerning the nature of human cognition and sensory experience and their intimate interrelationship, and it fails to take account of a vitally important insight. This is that how we conceive of physical objects is inextricably bound up with how they appear to us in perception - how they look, sound, feel and so forth. Thus, to repeat, although conscious thought is obviously not the same thing as perceptual experience, the *conceptual content* of thought is intimately related to the content, both phenomenal and intentional, of perceptual experience. The Kantian duality of receptivity and spontaneity leaves no gap between mind and world as human conceptual capacities become involved in all experience. Kant's famous slogan (A51/B76) encapsulates this necessary cognitive complementarity and semantic interdependence of intuitions, which derive from perceptual experience, and concepts, which come from the understanding. Moreover, Kant's closing words in the first *Critique*, emphasised the point that reason is immanently self-developing (A835/B863). This suggests that human cognition involves a necessary relation between the elements of cognition and its overall grasp by a human subject or active *agent*. This is an integral relationship, and one which is often described by Kant as "purposive". Kant took it be the case that such "purposive" relations were not available to "machines" and this is an important topic within the *Critique of Teleological Judgment*, and in his *On the Use of Teleological Principles* in *Philosophy* where he argued that if experience is real, then those who are in possession of experience must be understood as purposive beings that are also capable of positing their own purposes into nature according to the basic powers of will and understanding with which they are equipped. This leads to the conclusion that experience presupposes purposive beings, i.e., organisms, for only such purposive beings are empirically accessible. In other words, Kant recognised that rational human agents are necessarily also rational human living organisms, i.e. biological animals capable of intentionality whose rational mindedness and rational

directedness towards objects in the world, other real persons, and themselves, is fully continuous with this. As he wrote in the *Critique of the Power of Judgement:*

> It is quite certain that we can never adequately come to know the organized beings and their internal possibility in accordance with merely mechanical principles of nature, let alone explain them; and this is indeed so certain that we can boldly say that it would be absurd for humans ever to make such an attempt or to hope that there might yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws (*CPJ* 5:400).

Chapter 6 examines the implications of this to an alternative view of the mind that has quite recently started to emerge, that of embodied or enactive cognition. From this perspective the brain appears as part of a dynamical process (and not a syntactic one) of real time variables with the capacity for self-organisation (and not as representational machinery). Chalmers has hypostatised consciousness in the manner of Descartes, and taken its direct and ineffable nature to imply that it can logically float free, and this gives rise to dualism. But if we take a different perspective and view conscious experience as an activity, something we do, as Kant suggests, we can begin to understand that conscious experience *is* an activity, especially if we consider and attend to our experiences carefully. Then we can also ascertain what sort of activity it is, and obtain an understanding of how our consciousness relates to the world in which our whole person, and not just our minds, are a part. The rationale behind this is the radical insight that the things that we understand directly are not simply *qualia*, i.e. the phenomenal feel of "what it is like to see the colour red" but, rather, are related to the activities that we ourselves undertake and which we have learned in our normal development as embodied, involved, active human beings. Such understanding derives from our very nature, and from the symbiotic co-constitution of ourselves and the objective world around us through which, and in which, we have developed.

This chapter has presented a philosophical analysis of the modern day problem of consciousness in order to illustrate that Kant's theory of the transcendental subject, though written over two hundred years ago is still relevant today. Proponents of a "hard problem" have reintroduced the Cartesian conception of mind

as immaterial substance in favour of a materialist ontology, but this does not dissipate the inherent problems with the model. Chalmers' "zombies argument" is, in fact, a reworking of Descartes' argument in the *Sixth Meditation* (1641) where he notes that since he can "clearly and distinctly" conceive of himself existing apart from his body (and vice versa), and since the ability to clearly and distinctly conceive of things as existing apart guarantees that they are in fact distinct, he is *in fact* distinct from his body. Similarly, Chalmers uses a conceivability-based rational argument to separate mind and body; his "zombic" claim is that the body can exist without consciousness: this is because we can conceive of a possible world containing our physical bodily duplicates but which lack phenomenal consciousness or minds. Chalmers belongs to the dualist tradition, since he views mind or conscious experience as neither physical nor material, but a fundamental "property" and also a distinctively non-material entity. However, as discussed, it is neither obvious, nor even *prima facie* plausible to assume that the two ways in which we grasp physical and phenomenal properties, the former by descriptions of their causal role and the latter by a supposed "rigidly designating" direct awareness, reflect two metaphysically distinct properties. There is an un-argued for assumption of dualism. Antoine Arnauld in his famous intellectual exchanges with Descartes questions the inference from the ability being able to clearly and distinctly conceive of mind and body as separate to their actually being separable.[62] "How does it follow", he asks, "from the fact that he is aware of nothing else belonging to his essence that nothing else does in fact belong to it?" A similar question can be asked of Chalmers today. In fact, the kind of objection that Arnauld directs towards Descartes is a central problem for all conceivability-based accounts of the epistemology of modality (see Stephen Yablo, 1993, p. 2). What Arnauld's objection illuminates is a profound difficulty with modal arguments in coming up with an internally verifiable criterion for conceivability that does not admit of counter-examples. The notion that we can determine the nature of consciousness simply through an exercise of the imagination is wanting, not least because anti-zombie arguments can also be constructed using the same strategy employed in Chalmers' zombie argument (Frankish, 2007).

It has been argued that the central mystery about a hard problem of consciousness supposedly in need of explanation is a reaction to and an artefact

produced by adherence to the analytic/functionalist orthodoxy in which consciousness is reducible to or explicable by a set of functional cognitive processes realised in the brain. Such a view or picture *creates* an artificial explanatory gap between function and phenomenology in the first place. Chalmers argues that the existence of an explanatory gap justifies the ontological conclusion that phenomenal consciousness is a fundamental property *in addition to* the fundamental physical properties. However, this argument from the explanatory gap or epistemological "something is left out" claim to the dualist ontological conclusion is vehemently denied by others. There are therefore two broad camps: those who view "the problem of consciousness" as a tractable question and those who deny that there is anything answering to our conception of consciousness to the extent that it goes beyond structure and function. Daniel Dennett belongs to the latter camp, and just as it has been suggested that Chalmers is the modern counterpart of Descartes, Dennett takes on the role of the empiricist Hume, as a type of scientifically-oriented empiricist, for whom the self does not, in some sense, "exist". Dennett is the self-proclaimed captain of what he terms "the A team" of reductive materialists, (along with Patricia and Paul Churchland, Quine, Rorty, Hofstadter, the Churchlands, Lycan, Rosenthal, and Harman, who also argue against the reliability of introspective "evidence" about the inner workings of the mind, and declares that Chalmers is the captain of "the B team", whom along with Nagel, Searle, Fodor, Levine, Pinker, Harnad and others have a "gut intuition" that the B team leaves something out (Dennett, 2001, p. 2). Dennett is an analytic functionalist, like Chalmers, but regards the self as nothing beyond the various "subagencies and processes" in the nervous system that compose us, and as is typical in functionalist accounts, he thinks of *qualia* as the part of an experience left over once all the objective parts are eliminated, and which are in some sense "illusory". According to Dennett's functionalism, we live in a mechanical universe, and our minds are replicable within reasonable accuracy on a universal Turing machine; we are simply machines, with no place for the mind. Although changing in detail, he has kept true to this understanding of cognition throughout his philosophical career. Since Dennett's anti-Cartesianism is largely Humean in nature, Kant's depth of insight into the sources of our cognition can be of great significance in the debate; specifically, in addressing the impasse between those who claim there is a "hard problem" to be addressed by a science of

consciousness (Chalmers and "the B team") and others who deny there is anything of the sort; consciousness is an "illusion". This eliminativism about consciousness is the topic of the next section.

5.4. The No-Self Theorists

In his 1991 book *Consciousness Explained*, Dennett had taken on the Humean position that consciousness is an illusion, there is no inner self, and what we call mind is in reality the causal underpinnings of behaviour distributed across a collection of autonomous subsystems operating without any centralised supervision. For him, minds are simply "fictions" of folk psychology; the "hard problem" is a theorist's illusion (Dennett, 1996b, 1998), something inviting therapy, not a real problem to be solved with any supposed revolutionary new science of consciousness. Computers explain consciousness perfectly well, in terms of subsystems, each of which possesses its own small and specialised task, rather like a worker on a factory assembly line. He claims that Hume had been the first to suggest impressions, ideas, memories and imagination as "subsystems" that explained consciousness and who believed that "something, somewhere, somehow, just must unite them to explain awareness," and struggled with the problem. Dennett terms this "Hume's Problem", or the homunculus problem, which was how to get ideas "to think for themselves" (Dennett, 1978, p. 122). As Dennett diagnoses it, Hume's trouble was that it did not occur to him that there might be increasingly smaller subsystems, which think their simpleton thoughts. Dennett's answer to Hume's problem is that the sum total of these increasingly smaller subsystems eventually discharges the homunculus or "little man" that is the supposed locus of thoughts. In later writings, Dennett conceptualises consciousness neuroscientifically, in terms of its supposed computational architecture in which each of our hundred billion neurons essentially function like tiny organic robots. He writes:

As I like to put it, *we* are robots made of robots–we're each composed of some few trillion robotic cells, each one as mindless as the molecules they're composed of, but

working together in a gigantic team that creates *all the action* that occurs in a conscious agent (Dennett, 2001, p.13).

Dennett regards himself as Hume's successor in making philosophy pay attention to science, especially brain science. As discussed in Chapter 3, for Hume cognition pertains to "ideas", and not directly to external objects; there is a "veil of perception", i.e. that when one perceives an object one is not immediately aware of the object itself but of one's sensory experience *representing* the object. What is significant in terms of the aims of this thesis is that empiricists adopted this Cartesian view about perception as self-evident, and this notion, of ideas representing or *standing for* something in the mind was again adopted, again uncritically, by scientists at the start of the mid 20th century cognitive revolution, and where they were translated into the mental representations of cognitive science, and conceived of as "mental causes", the interactions of which are said to constitute mental processes that constitute a mind. Dennett, as a successor of Hume, developed the Intentional Systems Theory, in which he holds a *moderate* realism about the nature of representations and claims that beliefs and desires are constituted by patterns of observable behaviour, what he terms "the intentional stance", which is "a theory-neutral way of capturing the cognitive competences of different organisms (or other agents) without committing the investigator to overspecific hypotheses about the internal structures that underlie the competences" (Dennett, 2009, p. 344). In other words, Dennett's theory is an attempt to naturalise the mind and to reduce mental phenomena to simple physical systems. He claims the advantage of the intentional stance is that it can be used "to explore models that break down large, sophisticated agents into organizations of simpler subsystems that are themselves intentional systems, sub-personal agents that are composed of teams of still simpler, 'stupider' agents, until we reach a level where the agents are "so stupid that they can be replaced by a machine" (Dennett, 2007, p. 88). [63] For Dennett, "the self" is merely a metaphor for the unity of distributed neural events, which create the illusion of consciousness with its associated feelings of autonomy and freedom. We have the illusion of agency which simply derives from our genetically and mimetically nuanced ability to "take-up the intentional stance." Thinking is simply the process underlying natural selection among "memes" which he regards as the cultural analogue of genes.

The intentional stance meets the criteria for algorithms - it is a "mechanism" that produces results regardless of the material used to perform the procedure, i.e. it is multi-realisable. He writes that "the power of the procedure is due to its logical structure, not the causal powers of the materials used in the instantiation" (Dennett, 1995, pp. 50-51). Dennett claims that cognitive science is badly in need of philosophy, as cognitive scientists are in the grip of delusions and confusions about human cognition and that cognitive science is a "land of plenty" for philosophers since so many of their questions are "ill thought out". As he writes:

> One of the reasons cognitive science is such a land of plenty for philosophers is that so many of its questions—not just the grand bird's-eye view questions but quite proximal, in-the-lab-now questions — are still ill thought out, prematurely precipitated into forms that deserve critical re-evaluation (Dennett, 2009, p. 233).

In fact, he claims, "cognitive scientists [...] are often just as much in the grip of the sorts of misapprehensions and confusions as outsiders succumb to, and being down in the trenches sometimes makes them even more susceptible" (ibid., p. 231). Dennett is fully aware of the difficulty of accounting for consciousness in an information processing model of cognition and how it can lead to conceptual difficulties, and in particular, what he terms, the Myth of Double Transduction (Dennett, 1996, 1998, 2015).

> The idea [of information processing] sometimes leads to serious confusions. The most seductive confusion could be called the Myth of Double Transduction: first, the nervous system transduces light, sound, temperature, and so forth into neural signals (trains of impulses in nerve fibers) and second, in some special central place, it transduces these trains of impulses into some *other* medium, the medium of consciousness! That's what Descartes thought, and he suggested that the pineal gland, right in the center of the brain, was the place where this second transduction took place–into the mysterious, nonphysical medium of the mind. Today almost no one working on the mind thinks there is any such nonphysical medium. Strangely enough, though, the idea of a second transduction into some special *physical* or *material* medium, in some yet-to-be-identified place in the brain, continues to beguile unwary theorists. It is as if they saw — or thought they saw — that since peripheral activity in the nervous system was mere sensitivity, there had to be some more central place where the sentience was created. After all, a live eyeball, disconnected from the rest of the brain, cannot *see*, had no *conscious visual experience*, so that must happen later, when the

> mysterious x is added to mere sensitivity to yield sentience (Dennett, 1998, p. 72).

Dennett rightly dismisses what he terms the "Cartesian Theatre", the place in the brain where "everything comes together" for "central processing". There is no fixed point in the brain where this happens; this is simply one of the most seductive delusions about consciousness, the idea that somewhere in the brain there is a place where a picture of the world is displayed for a "control centre" to deal with. Dennett's alternative theory is that the idea of self-consciousness is a product of a "centre of narrative gravity" or to use his later metaphor "fame in the brain" - there is no such thing as a conscious self. He is eliminative about consciousness; it is an illusion in the sense "that it is not what it is supposed to be". *A fortiori* Dennett denies there is "a problem of consciousness"; that is, consciousness characterised in terms of "what it's like for the subject," and no need to account for *qualia*, the private, incommunicable redness of red or indescribable sight and smell of a beautiful rose. As a functionalist, he regards the self as nothing beyond the various "subagencies and processes" in the nervous system that compose us and *qualia* as the part of an experience left over once all the objective parts are eliminated. For instance, when we look at a deep scarlet, there are two things that occur: one is the sensory *information* that there is deep scarlet, (access consciousness or A-consciousness according to Ned Block, 1995) and the other is that we also *see* deep scarlet (Block's phenomenal consciousness or P-consciousness). And that, the phenomenal experience of seeing the deep scarlet colour, is the part that Dennett, against Chalmers and others in the "B team" dismiss as illusion.

Dennett conceptualises consciousness neuroscientifically in terms of its computational architecture in which each of our 100 billion neurons essentially function like little organic robots. There is nothing a human being can do that a computer cannot. His model of the mind is as software parallel processing in the "virtual machine" of the mind, with massive memory capacity to store programs for any imaginable purpose. There is no locus of consciousness, he claims, and this is because we just know empirically that there is no such place. If you look at the brain all you see are spike trains and nerve pulses, no colours, or sounds, no homunculus.

Dennett's critique can be taken on two levels, dualistic (mind-body interactionism) and materialistic (a place in the brain). At the dualist level all the information is supposed to somehow be taken to a point in the brain and then mysteriously communicated to the incorporeal mind of the subject, which then directs the body, which entails that there must be a material place in the brain responsible for the observer/judge/act process of thought, a homunculus, or brain organ (the Cartesian Theatre). Since mind and body are distinct things there is the problem of interaction in that the sense organs, via the brain must inform the mind; it must present it with data of some kind and the mind must then direct the body in the appropriate action. The locus of interaction for Descartes was the pineal gland or *epiphysis cerebri*. On noting that information from the eyes, as well as the other sense organs, seems to have merged somehow, he hypothesised that there must be a single place in the brain where it all comes together before entering consciousness. According to Dennett, this unquestioning commitment to the Cartesian theatre is what theorists either explicitly or implicitly retain. In *Consciousness Explained* Dennett writes:

> Let's call the idea of such a centered locus in the brain *Cartesian materialism*, since it's the view you arrive at when you discard Descartes' dualism but fail to discard the imagery of a central (but material) Theater where it "all comes together" (Dennett, 1991, p. 107).

Dennett, with his strong background in computer science and artificial intelligence, continues to seek a "computational", information processing model of the mind based on software parallel processing in the hardware or "virtual machine" of the mind. Like modern computers the mind is nothing but hard wired processing, with "memory" for storing programs for almost any imaginable purpose. In his earlier work, *Brainstorms* (1978), he had described and defended the classic GOFAI (Good Old Fashioned AI) strategy that came to be known as "homuncular functionalism", or the task of replacing the little man in the brain with "a committee" of simple agents. Later he used the term "self-organizing systems", of which the termite colony is a prime example. These are systems in which apparently coordinated activity arises from the joint operation of autonomous subcomponents. He writes:

> The behavior of a termite colony provides a wonderful example of it. The colony as a whole builds elaborate mounds, gets to know its territory, organizes foraging expeditions, sends out raiding parties against other colonies, and so on....Yet, in fact, all this group wisdom results from nothing other than myriads of individual termites, specialized as different castes, going about their individual business—influenced by each other, but quite uninfluenced by any master-plan (Dennett, 1998, pp. 39–40).

He recently changed his mind and this "micro agency" is now at the level of neurons which he calls "neuronal subagents". Neurons have agency, in the form of metabolic selfishness, in the sense that there is competition going on between them for resources. Although he had been expecting that his chain or hierarchy of homunculi would end up with the kind of simple switch that a neuron was then widely taken to be; he now thinks that he underestimated the complexity of neurons and their behaviour, and holds that neurons should be considered agents in their own right, competing for control and resources in a kind of pandemonium. "[E]ach neuron, far from being a simple logical switch, is a little agent with an agenda, and they are much more autonomous and much more interesting than any switch" (Dennett, 2013). In other words, intelligent control of behaviour is still a computational, functional process. A mind is still merely a metaphor for the unity of distributed neural events, which create the *illusion* of consciousness. Computation eliminates the self by providing an account of how separate yet interconnected subsystems or neurons in the brain generate coordinated behaviour without the need for any central supervision, or "a place where it all comes together". The basic idea is that there is a level of abstraction where the brain can be described in terms of hundreds, thousands, or even millions of little modules, more or less independent of each other, each with its own functional purpose or goal. This is the hallmark of functionalism, as has been discussed: a commitment to multiple-realisability and to the insistence that what matters are the workings of the mind, what it can *do* rather than how it is constituted, or in Dennett's well known phrase "handsome is as handsome does".

Dennett rightly criticises the Cartesian conception of consciousness, which he calls Cartesian Materialism, in which the mind is pictured as a tiny theatre in the brain "where it all comes together", and acknowledges the seductive conceptual appeal of the "I" as the locus of thoughts, as does Kant. However, he is not blessed with the richness of Kant's insights. The locus of functionalism is not mind or

consciousness but mind's infrastructures or the causal mechanisms underlying mental phenomena. Properly interpreted, results on infrastructural processes are supposed to enhance our understanding of mind; but it is clear that when Dennett (and other functionalists) talk about computation or the computational mind, they refer to causal processes, not the mind itself. For Dennett, the mind is merely a metaphor for the unity of distributed neural events, which create merely the illusion of a punctuated consciousness. If we relate this back again to the discussion in Chapter 3, Kant agrees with Hume (and, by implication, Dennett) that when we introspect we find no unitary self. "No fixed and abiding self can present itself in this flux of inner appearances" (A107). For Kant too, the way to model the mind is on how it works. The categories can be viewed as the most generic logical constraints on the formation of concepts and the "I" of transcendental apperception is that which unifies them. This is an abstract description of the mind formed by reflecting on its conceptual activities, in the sense that they are logically different from descriptions of actual composition. Kant says that nothing can be known from them about the actual nature of that which realises these states; not even something as basic as whether it is simple or complex (A353). As was discussed in Chapter 4, Kant's insistence on the unknowability of the self and his emphasis on abstract "transcendental" and "necessary conditions" for all cognitive pursuits does imply an agreement with functionalists' assertions. However, there are also significant differences. For Kant, the elements in Hume's bundle of perceptions has a unity, a unity which results from the amalgamating activity of a subject, (an x) of which we can say nothing but which plays a certain role in our cognitive abilities and which necessarily exists and provides the explanatory power missing in the Humean account. The crucial point is that for Kant "judgment" requires "a transcendental subject of thought = *x*", a unified, identical self that is the locus of such judgment. The same self that is affected by sensibility must also synthesise and judge, and therefore must possess all the faculties and receive all of the data that contribute to that subject's experience. The transcendental unity of apperception is this ground or *grund* in which all of the necessary components of having experience come together. Kant calls the transcendental unity of apperception the "first principle of the human understanding" (B139). This is the transcendental "I" which implies "identity" and which lies behind the ever changing flux of experience as something which remains

unchanged throughout. Judging empirical objects also requires that the subject have a faculty of self-ascription.

Dennett, following his "favourite philosopher" Hume, holds a particular form of determinism called *compatibalism* (Dennett, 2003). This is the view that free will should be redefined so that it no longer involves a free choice among alternatives and can be made compatible with the mechanistic/reductionist model of science. Thinking is the process underlying natural selection among memes in the same way that biological reproduction is the process underlying natural selection among genes (He takes this idea of memes from Richard Dawkins, 1976, 1989). Dennett's aim is to show how evolutionary thinking can account for everything "from senseless atoms to freely chosen actions". However, he assumes that causality is simply a relation between two events, such as when the motions of atoms at one moment cause their motions at the next, or when the firing of axons in the brain and nervous system causes the arm to move, etc. That is to say, his analysis of causality is mostly restricted to relations where one event can be said to have caused another, chemical reactions, nuclear fission and fusion, magnetic attraction, hurricanes, volcanoes, etc. An alternative view is that causality is a relationship, not simply between one event and another, but between a cognitive agent and their actions in the world. Actions are self-generated or in Kantian terminology, *spontaneous*. Although the contraction of a muscle can, on one level, be regarded as being caused by the nature of the muscular and nervous system, at another level it transcends it, through the ability to weigh up alternative courses of action and the capacity for genuine choice. However, this is totally in accordance with the principle of causality and does not contradict it in any way. Such agency is an expression not just of our embodiment in nature but is also definitive of the capacity to transcend it. It is an expression of an agent's existence not simply as a natural, biological creature, but also as a historical, political, culturally embedded human being. Dennett does not see this - for him agent causation is a "mysterious doctrine". In *Freedom Evolves* he writes:

> How does an *agent* cause an effect without there being an event (in the agent, presumably) that is the cause of that effect (and is itself the effect of an earlier cause, and so forth)? Agent causation is a frankly mysterious doctrine, positing something unparalleled by anything we discover in the causal processes of chemical reactions, nuclear fission and fusion, magnetic attraction, hurricanes,

> volcanoes, or such biological processes as metabolism, growth, immune reactions, and photosynthesis (Dennett, 2003, p. 100).

As discussed in Chapter 4, for Kant causality and freedom refer to two different ways of viewing things. Although theoretical philosophy cannot establish that we are free from spatio-temporal causation, Kant frequently mentions that we do not regard ourselves as agents whose actions are to be explained this way. Although under the "phenomenal perspective", mental activities are causally determined, to engage in cognitive activities means to consider ourselves not as "objects", but "transcendentally", as "subjects" of experience. This means that we must therefore consider ourselves *apart* from the conditions to which all objects of experience are subject, as rational free agents. Here the definition of freedom is purposive, the power to *spontaneously* originate behaviour and generate a new causal series. Transcendental freedom thus refers to how a rational agent can, "from itself" (*von selbst*) (A533/B561), be the spontaneous mental cause of certain natural events or processes. In other words, it is how certain natural events or processes in physical nature are, as Kant says in the second *Critique, in meiner Gewalt*, literally, "under my control" or "in my power" (*CPrR* 5:58). It is this crucial status of the transcendental subject, that adds richness to the Humean account and it is one which also stands in stark contrast to the tendency of contemporary functionalism to minimise the differences between minds and other things. It is clear from Kant's later work, particularly his *Critique of Judgement* and the *Opus Postumum* that for him human persons are also rational human *animals*, and that the capacity for free will is fully metaphysically continuous with their animality:

> The human being, as animal, belongs to the world, but, as person, also to the beings who are capable of rights—and, consequently, have *freedom* of the will. Which ability essentially differentiates [the human being] from all other beings; *mens* is innate to [the human being] (*OP* 21:36).

This will be expanded upon in Chapter 6, on embodied, enactive cognitive science.

Dennett is one among many in adhering to the fundamental notion that the mental and neural are one; the conscious mind is a solely biochemical phenomenon, hence consciousness is an illusion and can be eliminated. In his magnus opus *Being*

No One (2003), Thomas Metzinger referred to "the science of phenomenology" as an impossible theoretical endeavour. This is because, "[f]irst-person access to the phenomenal content of one's own mental states does not fulfil the defining criteria for the concept of 'data'" (Metzinger, 2003, p. 591). In fact, he avers, it is a contradiction in terms. Instead of the transcendental subject, what literally observes or experiences the world is a whole conscious system, a system that incorporates sensory information into a constantly updated overarching self-world representation. What makes this a phenomenal, *conscious* representation, is that it possesses a certain minimum set of representational, informational, and functional properties, e.g., global availability of information for action control and the integration of representations. What we take to be selves are simply the results of ongoing computational processes that satisfy certain conditions and produce a self model. The "no-self thesis", as he calls it, follows from a specific neurobiologically grounded theory of consciousness, which Metzinger terms the Self-Model Theory of Subjectivity or SMT. No such things as selves exist in the world. For all scientific and philosophical purposes, the notion of a self, as a theoretical entity, can be safely eliminated. In fact, not only the self, but also the world it perceives are illusions; neurophysiological processes are all that really exist. Metzinger asserts that interdisciplinary empirical work must replace "armchair" *a priori* intuitions into the nature of reality; nonetheless, his own position rests upon numerous, unquestioned *a priori* assumptions about the nature of reality.

Metzinger's approach is one based on a teleofunctionalist and naturalist view of consciousness which has been viewed as a neurological updating of Kant's project. Instead of the "conditions" of possible experience, Metzinger deduces the "constraints" constitutive of phenomenal consciousness and self-consciousness. Unlike Kant, however, he does not describe consciousness in a single, unitary way. His view is that with the advances in psychology, neuroscience and phenomenology, we can now achieve a much more detailed and nuanced view of consciousness, and dismisses transcendental notions as unnecessary. The time for pure philosophy has gone, what is now needed is an interdisciplinary approach to consciousness, in which the empirical sciences reign supreme and philosophy takes a back seat. He argues that phenomenal selves are "appearances" produced by the ongoing operations of a "phenomenal self-model" that simulates, emulates, and represents aspects of the

system's states to itself (Metzinger, 2003, pp. 160-162). In short, Metzinger's thesis is that the self is nothing beyond a special kind of dynamic representational content. His conclusions about the nature of the self are solidly founded on a set of inferences derived from the scientific method, and on scientific data, which "are things that can be extracted from the physical world by technical devices like telescopes, electrodes or functional MRI scanners" (Metzinger, 2003, p. 606) whereas first person access to one's own mental states does not fulfil the inter-subjectivity criterion of data, since group mediation of independent verification does not exist. Thus described, the brain is to be viewed as a system which constantly hallucinates at the world and creates the content of phenomenal experience (ibid., pp. 51-52).

One of the most significant aspects of Metzinger's teleofunctionalism is the integration of deficient forms of *phenomenal* consciousness, ones where a particular constraint may be absent as the result of neurological damage. The end result is a set of necessary and contingent constraints for consciousness to occur. The addition or subtraction of a particular constraint leads to entirely different forms of consciousness. The upshot is, as is typical of functionalist accounts, consciousness has been eliminated. No such things as selves exist in the world: nobody ever had or was a self. All that exists are phenomenal selves or "appearances" as they appear in conscious experience (Kant calls them appearances also). Metzinger argues that the phenomenal self is not a *thing* (as did Kant) but an ongoing process. What we think of as a self is simply the content of a "transparent self-model" and we are confused if we regard this model as a genuine self. Metzinger writes:

> This phenomenally transparent representation of invariance and continuity constitutes the intuitions that underlie many traditional philosophical fallacies concerning the existence of selves as process-independent individual entities, as ontological substances that could in principle exist all by themselves, and as mysteriously unchanging essences that generate a sharp transtemporal identity for persons. But at the end of this investigation we can clearly see how individuality (in terms of simplicity and indivisibility), substantiality (in terms of ontological autonomy), and essentiality (in terms of transtemporal sameness) are not properties of selves at all" (ibid., p. 626).

On Metzinger's view, the self, the feeling of identity, of being a subject in charge of the physical body, is simply a module within the mind activated by the brain's neural processing. The self is categorically *not* some substantial, essential invariant entity, like a soul, spirit or homunculus. As he emphasises, there are *no*

such things as substantial selves. Metzinger's point is similar to Kant's in many ways. All three properties of the self (i) the substantial "I," (ii) the simple soul and (iii) numerical identity over time are expressly similarly described by Kant as being "illusions" However, Metzinger does not seem to be familiar with the Paralogisms where Kant discusses the self. This is evident from the following passage where he claims that Kant:

> [C]onclude[d] from the fact that, in standard situations, all of us experience ourselves as initiators of our own thoughts or that the "I think" can, in principle and in the large majority of phenomenal configurations, accompany all states of consciousness, that some kind irreducible *entity* (e.g., a transcendental subject) must exist (ibid., p. 446).

However, just the opposite is true of Kant; the unity of apperception does *not* exist as an empirical reality or entity, but as an "x" that necessarily precedes experience yet also encompasses it. In fact, as discussed in chapter 3, the Paralogisms comprised pages of detailed, explicit arguments to that effect. Self-consciousness as inner sense or receptive consciousness of what we passively "undergo" (as we are affected by the play of our own thought) is differentiated from consciousness of our activity, i.e. of what we are "doing" in synthesising or unifying our experience. Nevertheless, the Kantian element of his work is noted and praised by eliminative materialists Paul and Patricia Churchland who write in an editorial of his book:

> *Being No One* is Kantian in its scope, intelligence and depth. Steeped in contemporary neuroscience, psychology and philosophy, the book gives the unsolved Kantian problems of inner self and outer world a new look, a new life, and a new route to solution. Metzinger's story is understandable, compelling, and, quite simply, very very smart.

But they misconstrue Kant: It is important to realise that Kant's transcendental philosophy sought to reflect on the conditions of possibility of experience and cognition, which is epistemic, not phenomenological, i.e. to do with the study of the structure of various types of experience. Thus, it is no coincidence that Kant categorically rejects the attempt to equate the notion of "intuition" with a type of inner experience or introspection. The unity that Kant calls the transcendental unity of apperception is stable identity. Empirical apperception, however, is the "[c]onsciousness of self according to the determinations of our state in inner

perception [which is] (...) always changing". He also calls this "inner sense", and its meaning is most adequately rendered by the Cartesian "*ego cogito*", the preceding condition of which is transcendental apperception, which he describes as "that unity of consciousness which precedes all data of intuitions, and by relation to which all representation of objects is alone possible" (A107).

Kant claims that this type of apperception has the nature of pure, original and unchanging consciousness, the deepest "transcendental ground of the unity of consciousness in the synthesis of the manifold of all our intuitions, and consequently also of the concepts of objects in general" (A106). Through his epistemic analysis he brings forth the idea of the deepest principle of being and knowledge, by means of which we are able to be responsible and free. Metzinger pours scorn on freedom of the will and is concerned to find the cognitive mechanisms that underlie action at an abstract level, or how "mental processing" occurs among representations at the neural level. His view is that the phenomenal self is nothing more than the ongoing operations of a complicated information-processing system. On Metzinger's model, our experience of "being someone" is simply a paradox: we have the experience of "being someone", yet there is no self having this experience. It can all be explained away in representational terms, and what we call "self" can be substituted by "phenomenal self-model". As he puts it on the first page of his book: "No one ever *was* or *had* a self. All that ever existed were conscious self-models that could not be recognized *as* models. The phenomenal self is not a thing, but a *process* (ibid., p. 1).

It is clear that Metzinger's work is the integration of deficient forms of *phenomenal* consciousness only, of inner sense. He writes "the phenomenal self is not a thing, but a process, and the subjective experience of *being someone* emerges if a conscious information-processing system operates under a transparent self-model… You don't see it. But you see *with* it'" (ibid., p.1). In an interview with Being Human journalist Michael Taft he states, "Ultimately, it's a physical process. Today, the best way to describe self-consciousness still is as a representational process: an image that is sometimes generated in the brain, an internal placeholder for the system as a whole, a neuro-computational tool" (Metzinger, 2012).

Thus, Metzinger regards apperception as redundant and unnecessary. In his eyes, Kant's views of subjectivity have been falsified by recent empirical discoveries in neuroscience. Neuroscientists today have attempted to define the fundamental features of the brain and its information processing capabilities in terms of (i) mean firing rates at points in the brain cortex (neurons) and (ii) computations, so that today, the prevailing view in neuroscience is that neurons can be considered fundamentally computational devices. Metzinger applies this metaphor of computation to the physical organism as a whole. He both reduces things down to micro-foundations and expands things up into a final functional effect. There is no epistemic justification for our mental states. Rather, "we have those states because they were functionally adequate from an evolutionary perspective" (ibid., p. 115). The "self illusion" is nothing but a causally generated cognitive phantasm generated by the physical organism as a whole. There are two major assumptions at play here: i) that phenomenal experience is generated by inaccessible sub-personal process, and ii) that something generated by dynamical processes must be an illusory reification. His purpose is to eliminate metaphysics in favour of his own preferred domain of neuro-pathological reflections.

It is suggested that many ideas are touted as those we need in order to be scientifically rigorous, when, in fact, they are simply the dogmatic acceptance of certain philosophical premisses. Metzinger's view is that the self is nothing more than the ongoing operations of a complicated information system, an "illusion" causally generated by the physical system as a whole. But a physical organism is not just a collection of parts; it *is* a whole, and it is this whole that we ordinarily ("folk psychologically"?) call a self. As he views it, conceptual confusions can only be avoided if we stop referring to the self in a naive realistic way. As with Dennett, Metzinger's conclusions are based on the scientific dogma that if something has causal antecedents, then only those antecedents can have independent reality. Moreover, as is often the case, his reasoning is question begging, he makes use of the latest scientific results to justify an ontology that was in large part already *presupposed* in his interpretation of the scientific data.

Metzinger's approach dissolves the conscious self into what is embedded within the whole brain. Although "I" control my actions, the "I" is reconceived as the

coalition of my brain processes. Instead of saying that my consciousness (me) is making the decisions, it turns out that we should say that "I" am conscious of the parts of my brain that are making the decisions. Metzinger claims that "*all* selves are either hallucinated (phenomenologically), or elements of inaccurate, reificatory phenomenological descriptions' (ibid., p. 462). "There is no single entity in or outside the system that directly corresponds to the primitive, pre-reflexive feeling of conscious selfhood" (ibid., pp. 564-565). As discussed, this notion of a non-reifiable sense of self had already been prefigured by Kant over two hundred years ago. However, it did not follow from this that we ourselves do not exist or were merely "appearances". While it is clear that believing the self is an immaterial entity or that it has an anatomical counterpart or core in the brain where everything comes together (homunculus) is incorrect, as Dennett points out, it is not so clear that the self itself can be so eliminated.

It is significant that Metzinger claims that a disembodied but appropriately stimulated brain in a vat could, phenomenologically, enjoy exactly the same kind of conscious experience as an embodied one. "In principle, it would even suffice to properly activate just a subset of this brain, the minimally sufficient neural correlate of your present state, to make a "phenomenological snapshot" of exactly the same kind of conscious experience emerge" (ibid., p. 547; see also pp. 295, 335, 462). [64] For a real sense of consciousness, however, there needs to be an attached body with sensorimotor and autonomic (sympathetic and parasympathetic) systems, and the release of hormones and neurotransmitters, producing a sophisticated interplay of bio-chemicals that give rise to gut feelings, emotions and a sense of embodiment and agency. Metzinger claims that it is possible that "attentional and cognitive agency can functionally be de-coupled from the process of autonomic self-regulation and the spatial self-representation necessary for generating motor behaviour" (ibid., p. 499). But, this is where the argument is extremely weak; the activation of neural correlates of experience would simply be a "phenomenological snapshot" and would not describe what it is like to be conscious self, in which a necessary defining feature is continuity and change. Neither is it simply a matter of "stimulating neurones to create an explicit body image". If there were not a biological lived body attached to the functioning brain, there would be no conscious self at all. Consciousness is not an illusion or hallucination, since it is itself generated by real embodiment, and by a

brain that is part of a real bodily system, which, in turn, is part of a real world. The brain and the body are, structurally and dynamically, so deeply entangled that they are explanatorily inseparable. It is therefore highly questionable whether consciousness *per se* can be understood by considering the actions of neurones apart from the body, for even if it were even possible to set up a brain in a vat, there could be no synthetic apparatus as sophisticated as a real human vascular system in its structural features and functional capacities. Experiments have shown that the very shape and design of the body contribute directly to sensory and bodily awareness. For instance, the position and design of the eyes and of the ears constrain sight and hearing respectively. Hearing experience, for example, is constrained by the shape and location of the ears which direct sound through the amplification and filtering of specific inputs (Chiel and Beer, 1997).

"The self", therefore, far from being an "illusion created by the operations of a complicated information system", is understood more cogently as intimately related to the agency of an organism and given a biological or social-psychological definition that "makes sense of the data". For if one considers the fact that an organism can distinguish itself from both the environment in which it dwells and from other organisms that exist, then the case is strong for a naturally-occurring and biologically evolved "sense" of self. This insight is to be found in the philosophy of Kant. Kant believes that all the threads of his transcendental philosophy come together in this "highest point" (B134), the transcendental unity of apperception, which in Kantian epistemology is the deepest principle of being and knowledge. It is clear that although Kant writes that the self of apperception involves the *notion* of the existence of a unitary subject, he also thought it was more than this. He writes in the Paralogisms, for example: "The proposition "I think" insofar as it amounts to the assertion "I exist thinking" determines the subject (B429), and also that the "I think" is "something *real* that is given (...) something which actually *exists*" (B423n) [my italics]. In the Transcendental Deduction he states "my existence is not mere "appearance" (much less mere illusion)" (B157). For Kant, we have a *sense* of self, which forms an integral and ubiquitous part of our experiential life. And the mind that senses is the same as the mind that possesses or intuits. And that mind must be the same as the mind that employs the table of categories, that contributes empirical concepts to judgment, and that synthesises the whole into knowledge of a unified,

empirical world. Most importantly, for Kant, "the unity of consciousness" by which we judge that such and such is not the ultimate end point. Rather, it is the human (or animal) organism's spatiotemporal perspective or "unique point of view" which is fundamental to the subjectivity of its experience.

There has recently begun a growing trend in Kantian literature, which brings to light his thesis on embodiment where *awareness* can be regarded as pre-figuring the spontaneity involved in the "formal intuition" of the I think.[65] This has been termed "transcendental embodiment" and it is seen as providing "the unifying thread of Kant's epistemology, moral philosophy, aesthetics and teleology of living nature" (Nuzzo, 2008, pp. 8, 9). His views on embodiment can already be seen in his pre-critical essay entitled *Concerning the Ultimate Ground of the Differentiation of Regions in Space*, among others, where he refers to the ability to distinguish between our left hand and right hand which he calls incongruent counterparts, objects that are conceptually identical apart from being mirror images of each other. He also refers to incongruent counterparts in *The Prolegomena to Any Future Metaphysics* (1783) which he describes as a paradox resolvable by his own theory of space as mind-dependent. His argument is that there can be two conceptually indistinguishable objects that we can nonetheless tell apart because they differ in spatial orientation, i.e. we can tell them apart not because of any difference in the objects themselves but rather because we represent space as a form of intuition that renders such differences palpable. A human observer experiences himself as intersected by three planes and as having three sets of "sides", which he describes as up and down, back and forward, and right and left. But which direction is right and which is left can only be established by a conscious, embodied being. This suggests that the body for Kant is the "transcendental ground" for our cognition, the locus or site for our "sensibility. In other words, the first person perspective-ness or me-ness of conscious cognitive content is not simply the apperceptive self awareness of "judgement" that is reliant on the categories, and divorced from the world, but is also necessarily bound up with spatial orientation and temporal asymmetry in an embodied biological system.

Although he does not say too much about it in the first *Critique*, Kant's "logical" or "apperceiving" subject, can be viewed as necessarily embedded in the world, an idea that is also found in the later work, published after his death, the *Opus postumum* where he writes that the organism plays an essential role in the process of

acquiring knowledge, in the sense that the actions performed by the organism belong to the universal principles of the possibility of experience, so that the subject affects itself and "makes itself an object of experience" (*OP* XXII 373: 30–3). He also refers to it in the *Critique of Judgment* (1790) where he presents a strikingly modern self-organisational account of life. There he claims that life cannot be derived at all from the mechanistic laws of Newtonian physics, which merely postulate "efficient causes and not "end causes". This is a teleological explanation of cognition involving active agency: our knowledge of the world is connected with the purposiveness of that world via the organism. Metzinger's self-model theory of subjectivity is a sparse and limited view where selves and subjects are the illusory outcomes of biological processes and do not form part of the ontology of the world. His claim that "attentional and cognitive agency can functionally be decoupled from the process of autonomic self regulation and the spatial self-representation necessary for generating motor behaviour" (ibid. p., p. 499) is a case in point. But this cannot explain how it is that *this* particular body that sits at this desk, typing this thesis right now, gives rise to my first person phenomenological, subjective experience, my unique point of view. Metzinger imagines I could have been someone else, because the self that is the subject of my experiences could have been paired with a different body, and that the "contingency intuition [that I could have been someone else] is not even based on a phenomenal possibility" (ibid., p. 612). But even if it is granted that it is phenomenally *possible* for me to coherently imagine that I could be someone else, it remains true that my actual existence *qua* subject (and hence *qua* self) is *not* contingent in the sense that it necessarily depends on the existence of this particular body. He describes the process of generating experience in terms of the activation of a set of minimally sufficient NCCs, which as he admits, would generate only a snapshot of experience. But lived experiences are more than mere snapshots and the reductionist picture cannot capture the irreducible nature of continuity and change of lived, embodied experience. Metzinger begs the question by assuming that the self model theory is true and the selves are fictitious illusions. According to him the feeling of self is simply caused by a representational self model that we mistakenly confuse to be selves. Selves are such an artifact that they can be eliminated by Occam's razor, which is the favoured tool of analytic philosophy and positivism, for "under a general principle of ontological parsimony it is not

necessary (or rational) to assume the existence of selves, because as theoretical entities they fulfil no indispensable explanatory function" (ibid., p. 337). As aforementioned, his reductionism is founded on the materialist dogma that if something has causal antecedents, then only those antecedents can have independent reality. Perhaps this materialist, and some would argue, "scientistic" stance has blinded him to the fact that even if the self can be described functionally and causally, as generated by the physical-organism-as-a-whole, this does not lead to the conclusion that it is *nothing but* a group of disconnected nerves and cells.

In summary, cognitive science, in the main, has adopted this view of the mind as functioning according to a set of deterministic mechanical laws, a fact that is inevitable given the presupposition of classical physics on which it is based. However, this is to ignore the fact, as Kant was the first to show, that judging something is inexorably linked to human action. In the third antimony of *The Antinomy of Pure Reason*, part of the Transcendental Dialectic, Kant addresses the question of whether mechanistic determinism is reconcilable with the ascription of rational free agency to human beings. In it he states that human actions differ categorically from events, the latter which could be given a causal explanation, differ from the former. This is because descriptions and explanations of human actions are part of what Wilfrid Sellars has termed "the space of reasons" - they are actions insofar as they are part of exchanges in a social world which involve giving and asking for reasons (in the sense of *rationale*, or *justification*), i.e. it is intrinsically connected to *meaning*. Therefore, perhaps the correct conclusion to draw from the fact that "the self" is not found in experience, is neither to accept Dennett's nor Metzinger's mechanistic/eliminativist view, nor to resort to Chalmers inflated ontology concerning a "hard problem". The reason thought experiments about qualia-free zombies have gained so much traction among philosophers of mind is due to the pervasive "the mind is a computer" metaphor. According to the standard computational approach, mental processes do their work by manipulating symbols with algorithms that take perception as input and produce behaviour as output. On this account what supports the spectrum of cognitive activities is a diverse collection of mechanisms in the brain sharing a common representational system. If one rejects this picture, however, one can come to understand that our everyday experience of

the world as being meaningful is inseparable from our experience of it as looking (sounding, smelling, etc.) particular ways, ways in which the idea of *qualia* do not even enter the picture. What is needed is a theory that captures this and is not beset by the various worries which result from the computational framework in which meaning is constructed through the input from an information laden word to the inside of an organism's head.

There is little empirical evidence supporting the presence of computational symbols in cognition; functionalism was simply eagerly adopted because it was able to provide elegant and powerful formalisms for representing knowledge, also perhaps, because it captured a theorist's intuitions about the symbolic character of cognition, and the fact that they could be readily implemented in exciting new Artificial Intelligence programs. A problem with traditionalist functionalist theories is that they cannot readily explain how cognition thus defined interfaces with perception, action, and meaning, what is also termed the meaning or symbol grounding problem. This was largely triggered by Searle's Chinese Room thought experiment (1980) in which he showed that if the Turing Test were conducted in Chinese then he, in a room, could carry out the very same program as a computer, of manipulating symbols, without knowing what any of the words he was manipulating meant. This was meant to show that abstract symbols such as words need to be grounded in something other than relations to more abstract symbols if any of those symbols are to be meaningful. According to Searle, an adequate explanation for meanings or intentionality can only be material, biological one, and the solution to the meaning-grounding problem is sensorimotor grounding.[66] The abstract functionalist, symbol view of meaning is pervasive in cognitive theories and yet it is arguably "one of the most remarkable misunderstandings in the history of science" (Edelman, 1992. p. 228). As Steven Harnad formulates the problem:

> How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads? How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols? (Harnad, 1990, p. 335). [67]

Fortunately, an alternative approach to traditional functionalist cognitive science is to be found in recently emerging research into embodied cognition, which understands cognition as highly dependent on the physical capacities and actions of a cognitive agent. According to this new embodiment perspective, it is impossible to understand the self or consciousness without grounding it in action. In grounded cognition theories, cognitive processing, even in abstract domains such as language acquisition and mathematics, are dependent on sensorimotor skills and bodily resources. More specifically, it is a way of viewing meaning as the coordination of action in order to achieve certain goals. There is no need to look for the hooks that ground the reference and referents, the speech acts and what they are about. Instead, cognition concerns the matching of goals and affordances. Moreover, the hard problem is solved because cognition is naturalised, but it is also *dissolved*, in that its initial motivation is shown to be ill-founded. Affordances (Gibson, 1979) are ways in which a perceiver with a particular type of body can interact with an object, or are to do with what an agent wants to achieve and the opportunities afforded by a situation in order that the agent achieve that goal. For example, a cup affords grasping, lifting, and tilting, allowing the user to drink from it. Such motor affordances are proposed to be central to conceptual knowledge (Borghi, 2005; Glenberg, 1997, 2008). In fact, Glenberg proposed that cognition *evolved* to coordinate effective action; that is, action that enhances survival and reproductive success given the constraints of a particular type of body. In this way cognition is naturalistic, and not considered as something fundamentally different, but rather as dependent on the body, its sensory motor systems, and on context in the real world (Barsalou, 2008; Clark, 2011; Laakso, 2011; Schubert, & Semin, 2009; Stapleton, 2013). Traditional theories place all the responsibility for generating behaviour in the brain; perception is input into a computational, representational system that mentally transforms the input into output or behaviour. Embodied cognition replaces the representational/functional model with agential activity and emphasises the role of the body and its place in the environment in creating cognition, arguing that even the most abstract of concepts are rooted in characteristics of our bodies and in our embodied interactions with the environment. This idea, of embodied, embedded enaction also has its roots in Kant. This is the topic of the next chapter.

6. Kant and Embodied Enactive Cognitive Science

Enaction was first proposed as a model for understanding cognition by F. J. Varela, J. Thompson, and E. Rosch in their 1991 book *The Embodied Mind*. As two of the originators of embodiment theories, Varela and Thompson had also surmised that "what the organism senses is a function of how it moves, and how it moves is a function of what it senses" (Thompson and Varela 2001, p. 242). Varela, in his ground-breaking paper "A Science of Consciousness as if Experience Mattered" (1997) advocated a dissolution of the hard problem as a result of this new methodological approach. The hard problem, as discussed, consists of finding a place for conscious experience within nature, as it is supposedly described by our best scientific theories. Varela argued, against such traditional accounts that conscious experience is not a thing or feature that one *has*, but what one *lives* and what one *dwells* in as an embodied organism. In addition, he points out, in a manner akin to Kant, that third-person objective science assumes the very fact of experience, in order to extract from nature invariants that can validate hypotheses; therefore one should not expect from it a convincing derivation of what is their most basic condition of possibility, experience itself. Instead, what is conventionally thought of as "subject" and "object" are co-constituted or co-arising, as the mind is embodied and arises out of "an active handling and coping with the world"; that "whatever you call an object (...) is entirely dependent on this constant sensory motor handling", which means that an object is not independently "out there", but "arises because of your activity, so, in fact, you and the object are co-emerging, co-arising" (Varela, 1999, pp.71-72). The mind thus construed cannot be separated from either the entire organism or the outside environment. "Knower and known, mind and world, stand in relation to each other through mutual specification or dependent co-origination" (Varela et al., 1991, p.150)....and also "organism and environment enfold into each other and unfold from one another in the fundamental circularity that is life itself" (ibid., p. 217).

Embodied enactive approaches to cognitive science have become increasingly popular over the past twenty or so years (Varela et al., 1993; Varela, 1997; Glenberg, 1997; Clark, 2011, 2016; Anderson, 2003; Thompson, 2007; Barsalou, 2008; Semin

and Smith, 2009; Shapiro, 2011; Stewart, Gapenne & Di Paolo, 2010; Di Paolo and Thompson, 2014; Noë, 2016). Although there are different empirical and theoretical strands to each approach, they all share in common the challenge to traditional approaches to cognition, in which mind is comparable to the software of a computer. From the standpoint of traditional cognitive science, if an agent is to successfully perceive and engage with the world, he/she must operate on "internal representations" of a world that is considered external and separate. Here the central idea is that to perceive is to be in an internal state that just *happens to be* caused by an external world. However, according to the embodiment thesis, this is an inadequate and one-sided understanding of cognition. The mind is already embedded in the world in a deep way; in fact, it is *constitutionally* so through the body's situatedness and active engagement with its environment. Accordingly, if we are to understand cognition, it is more fruitful to think beyond the inner and outer distinction of mind and world found on traditional accounts and to consider the natural unity that already exists between an organism and the environment in which it is actively engaged. This is because any purported "internal representation" that occurs in an actual embodied brain is *a priori,* as a direct result of its relationship to the environment through its embeddedness within an active organism. There is neither subjective supremacy over the objective nor is the subjective absorbed into the objective realm of traditional mechanistic science. Rather such science returns to the experiential realm from which the very dichotomy between subjectivity and objectivity arises, and then establishes *within it* a system of *mutual constraints.*

Contemporary functionalist theories tacitly assume a profound difference between consciousness and biological life (hence the penchant for assuming as, Chalmers does, that the right kinds of computation are sufficient for the possession of a conscious mind). The basis of mind shrinks to "representations" which are in principle multi-realisable and the body is reduced to a kind of input–output device. This kind of short-circuit between mind and brain, as they are thus considered, leads into a conceptual and methodological impasse, for it misses the essentially embodied, relational, biological and biographical character of mindedness. By way of example, let us take the existence of "place cells", neurones discovered by John O'Keefe, (O'Keefe, 1976) which are located in the CA1 and CA3 regions of the hippocampus and which fire when the organism occupies a specific location within

the environment. They are called *place* cells because their firing is primarily location specific (Muller et al, 1987) and may be independent of other descriptions of the organism's behavioural state, such as head direction. Different place cells have different "place fields" and together they provide a cognitive map which enables the organism to navigate and orient itself in space. O'Keefe also thought the function of the map was not simply to permit spatial navigation, but to also act as a memory framework upon which the significant items and episodes of experience could be superimposed (O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., & Maguire, E. A., 1998). Studies on place cells have shown that the organism's active locomotion and the specific environmental constraints on that locomotion have a fundamental bearing on the various facets of place cell activity. For instance, the firing frequency of a place cell is highest when the organism enters into the location equated with the centre of its place field and gets progressively lower as it moves towards the periphery: conversely, the firing rate of the cell increases as the organism moves from the periphery and back again towards the centre of its place field. Thus, the activity of the place cell is dependent upon and indicative of the active movement of the organism through the environment. In other words, the organism does not merely passively perceive the "external" environment so as to generate an "internal" model or map that governs action, as on traditional accounts, but rather its embodied action coupled with its species specific sensory capacities enables place cell activity to be sensitive not only to the environment, but also to the unique trajectory of action within the environment. Other cells, for example, Head Direction (HD) cells fire rapidly only when the head of a freely moving organism points in a restricted range of angles in the horizontal plane. They therefore share with the hippocampal place cells the property of signalling an aspect of the spatial relationship between an animal and its environment. O'Keefe specifically refers to Kant's understanding of cognition and, especially his understanding of incongruent counterparts in his book "*The Hippocampus as a Cognitive Map*", and, in particular, as necessarily involving absolute Newtonian space:

> The major evidence Kant used in support of the necessity for a notion of absolute space concerned objects which were similar but incongruent, such as left and right hands, left and right screws. The parts of these objects and their internal relations had exactly the same description and yet they could not be

> superimposed on each other because the three-dimensional space they occupied was different. In other words, part of the description of a left hand involved a reference to the space in which it was set. If there were only one hand in the universe, it would be impossible to say whether it was a left or right hand without recourse to an absolute spatial framework. At this point Kant was still thinking about absolute space in the Newtonian way, as a property of the physical world. His shift to a psychological interpretation of absolute space was due to Leonhard Euler's influence (O'Keefe, 1978, p. 20).

Kant's mature notion of absolute space was not as a property of the external world, but part of the human mode of perceiving. O'Keefe makes the point that this *a priori* mode of perception must have some correspondence to the physical if it is to be useful to the organism.

Evidence in evolutionary biology suggests that the first neurones evolved in order to organise and coordinate action. As the mind evolved it was not merely an improved reaction to stimuli from the environment that mattered, rather, it required the grasping of complex wholes or situations. A living being situated or embedded in its environment grasps a situation, and in so doing grasps itself in relation to it. In this way an attenuated self-hood or at least "sense making" goes all the way down in that there is a tenuous first person perspective in all biological organisms, and not simply human ones. The mind is formed bi-directionally, between organism and environment, creating our sense of an embodied being in the world, and involves an integrated evaluation of the meaning of a situation in order to weigh up options available for action. In later evolutionary stages, the development of tools and language would have enabled symbolic representations of meaning to be constructed, such as self, world, etc. On this understanding the mind creates wholes, which allows the organism to represent its relationship to the environment, (sense-making) and thus to act not merely in an automatic manner, but instead in a meaningful or purposeful way. It is interesting that Kant attributed the capacity for objective perceptual awareness to non-human animals, despite their lack of conceptual capacities. In Section 6.2 it will be argued that there is room in the Kantian system for the possibility of objective conscious awareness in non-rational animals, a form of awareness that is perceptual without being essentially conceptual in nature. That is to say, the underlying nature of cognitive content is not exhausted

by its logico-conceptual components. This is because the formal, intuitional, spatiotemporal structure of conscious cognitive content just *is* its subjective or "first person" character, and this is achieved prior to the use of the categories in apperception. This kind of *a priori* spatial and temporal cognition first hypothesised by Kant, is supported by the discovery by O'Keefe (et al) of the distinct class of neurons discussed above, whose firing is tuned to an organism's location, position and orientation in space. These exist independently of and prior to experience of the spatial world and form the framework or scaffolding that underpins the coherent organisation of experiences. This neural representation of "allocentric" space, supported by cells in the hippocampus and other brain areas can thus be seen as vindicating a Kantian synthetic *a priori* framework,[68] in that they encode a type of abstract spatial structure which is imposed on the environment without regard for the features of the environment. That is, they required no empirical or sensory experience for their validation.

Thus, the discovery of these place cells and grid cells can be viewed as a neurological updating and vindication of Kant's original theory that Euclidian space constitutes a synthetic *a priori* structure that is constructed by the mind without information from the external world. It is important to note that the organism's active locomotion and the contextual specificities of the environment as it relates to that have a direct influence on various facets of place cell activity. So, although place cells may be regarded as some kind of internal representation that acts as a cognitive map, enabling coordinated navigational activity of an organism, they are also modified and altered by the activity of the organism through a two-way reciprocal relationship. That is, although one could indeed speak of the coordinated activity of place cells in the brain as an internal representation, in the sense that they are brain processes that covary and carry information about the environment, this misses the following two points: Firstly, place cells do not encode a true depiction of a separate environment, but rather encode how relevant features of the environment relate to the organism and its embodied action. Secondly, the perceptive processes that enable the formation of place fields are not *passive* but intrinsically and fundamentally indebted to embodied *action*. Significantly, there is a Kantian *reversal of approach* to traditional cognitive science: internal place representations are not the condition for the possibility of experience, but human action is the condition for the possibility

of representations. [69] There is a two-way reciprocal relationship - action enables the formation of stable place fields (perception), which in turn orients and enables the organism's further action with regard to where it has been and where it can continue to move and act. Buzsáki and Moser have also recently proposed that mechanisms of memory and planning have evolved from mechanisms of navigation in the physical world and hypothesise that the neuronal algorithms underlying navigation in real and mental space are fundamentally the same. In general agreement with Kant, space perception has an *a priori* nature in the sense that place cells and grid cells in the entorhinal cortices determine how we perceive and remember our position in the environment, as well as the events we experience in the environment (Buzsáki, G., and Moser, E.I., 2013).

If we follow this line of thinking, we cannot regard consciousness as something merely epiphenomenal to the "real" processes that underlie it. On the contrary, subjective experience plays an absolutely central role in the systemic interaction of an organism and its environment. For this is what enables the organism to enter into a relationship with the environment at a higher level of meaning at all. Traditional functionalist accounts posit internal processes or representations as a starting point for explaining cognition and frame questions accordingly: How does the mind (or brain) use internal models to perform various cognitive acts? By framing the question this way, however, theorists are forced into explanations that ultimately *depend* upon the existence of internal representations for their coherence. However, a Kantian reversal of approach would amount to the question: How does embodiment provide the necessary conditions for the possibility of experience? The answer is that there is a coupling between organism and environment, instead of a one-way process of external impingement on the senses from an external world as on traditional accounts. An organism constructs and picks out its environment, just as the environment picks out the traits of an organism. Both form a *constitutive* feedback loop such that each is constantly shaping and defining the other, it is *constitutive* because the organism is partially the result of this two-way process of picking out. On the embodied enactive perspective organisms and environments are regarded as constitutionally intertwined and integrated, rather than considered as separate from each other. Internal representations can be a useful heuristic in cognitive science, and many valuable contributions have been made, but using them

as the foundation for cognitive science is limited. This is because human beings are biological organisms, and all biological organisms evolve through interaction with the environment. In other words, "their being is their doing", and as has been pointed out (Di Paolo, 2005, 2009), the circularity of the living, in the sense that their very being is their doing, is at odds with the functionalist concern of providing a substrate- independent account of the operations of representations.

Evolutionary biologists hypothesise that acts and action emerged prior to any purported internal representations, since behaviour evolved before nervous systems. For example, according to Gaspar Jékely, precursors of nervous systems arose to improve control of ciliary locomotion by means of division of labour and economies of scale, and that as a result, the first nervous systems arose, consisting of combined sensory-motor neurons directly translating sensory input into motor output, initially on locomotor ciliated cells and eventually steering muscle cells (Jékely, G. 2011). [70] If indeed the first neurones evolved in order to organise and coordinate action, then it follows that higher level cognitive functions *must* have evolved out of and be dependent upon the processes of embodied action. As evolutionary biologist Peter Godfrey-Smith avers:

> The basic pattern found in the evolution of cognition is a pattern in which individual organisms derive an advantage from cognitive capacities in their attempts to deal with problems and opportunities posed by environmental complexity of various kinds. Cognitive capacities confer this advantage by enabling organisms to coordinate their behaviour with the state of the environment. Cognition itself should be thought of as a diverse "tool-kit" of capacities for behavioral control, including capacities for perception, internal representation of the world, memory, learning, and decision-making" (Godfrey-Smith, 2001, p. 24).

The "hard problem of consciousness" of how mind emerges from the "wetware" of the brain is a result of a functionalist view which tacitly assumes a profound difference between consciousness and biological life. However, the problem only arises if mind and life are conceptualised this way, as separate or separable and intrinsically excluding one other. The underlying problem in traditional cognitive science, then, is one of relating a level of system or process to one of meaning. The conventional approach of isolating and splitting off these two levels leads directly to

the mind-body or mind-brain problem which is irresolvable when viewed through the orthodox lens of analytic philosophy and the associated Cartesian-mechanistic framework in which it is entrenched. To find our way out of this impasse, and to release ourselves from the perplexing difficulties that arise from this dualism, the problem needs to be reconceptualised and this is possible by taking into account the phenomenon of life, including the organism, the lived body and the world in which both mind and brain are embedded. If mind is construed as essentially embodied in the living organism then no longer will reductionist claims such as, ‘‘You are but a pack of neurons”, ‘‘You are a set of functions", or “You are a brain” be relevant. Both statements would be seen as biologically unsound and as prime examples of a category mistake.

As discussed earlier, according to traditional accounts the first main task for a “science of consciousness” is to find the NCCs, specifically, the minimal neural correlates for the phenomenal correlates of consciousness. Once these brain states are discovered there will be isomorphism between them and conscious states, even without the existence of body. As functionalist Ned Block put it “if the relevant brain state were to come about—*somehow*—the experience would be instantiated” (Block, 2005, p. 265). On this picture, it is possible that you could have a functioning brain in a vat (see earlier discussion of Metzinger, Chapter 5.4). Philosophers are fond of engaging in thought experiments and the brain in a vat scenario is one that is easy to imagine on the functionalist model. However, this overlooks the importance of the physiology of an organism in which the functioning brain is subordinate to the maintenance of bodily homeostasis. Moreover, a functionalist methodology such as Marr’s (see chapter 4) is said to make available a complex map of the “localisation of function” that are within the brain, along with a set of mathematical/ computational functions that formally describe the inner “mechanics” of the correlated brain regions. This is a problematic understanding of mentality as the brain activity underlying specific cognitive processes are distributed across many brain areas, making it practically impossible to localise functions to specific brain areas in the first place. There is the further confounding fact that the same brain area may be responsible for many different functions (see Uttal, 2001, later in this chapter). On the embodiment thesis, however, conscious experience is part and

parcel of a life regulation processes involving neural mappings of the body, which constitute a core self that grounds not only that neural activity overall but also the specific neural activity relevant to subjectivity. This forms an invariant basal awareness that remains constant across the constant flux of sensory changes and alterations. Or, in Kantian terms, this invariant basal awareness is the "I think", the original unity of apperception, the *grund* or ground of self-reflective awareness that endures through change. Self-reflective awareness implies an understanding of a self as enduring through time and through changes in states. But, crucially, it is the fact of *embodiment* that makes it possible to experience in the first place. Kant stresses that we intuitively relate things to the orientation of our body in order to argue that space can neither be perceived by the senses nor grasped by the intellect. Therefore, it is argued, he must have conceived of the oriented human body as a "transcendental ground for our cognition" in its own right. On the embodiment paradigm, consciousness is the result of the life regulation processes brought about by the nervous system. Brain and body co-evolved through a history of species specific adaptive behaviour co-constituted and bound by environmental constraints (See Aboitiz 1990, 1996; Chiel and Beer, 1997). Therefore, there is already a strict correlation between cognition and consciousness on the one hand, and real, biological and neurologically animated bodies-in-the-world on the other. Here the central idea is that to perceive is to be in an interactive, co-constitutional, dynamic relationship with the world, not to be in an internal state that happens to be caused by the external world. There is also the closely related "grounded" theory of cognition (Barsalou, 2008), which whilst employing the same constraints and characteristics of embodied situated cognition, posits the inclusion of the higher cognitive abilities: categorising, reasoning, numeric and linguistic processing, as well as theory of mind, all of which are grounded in the brain's modal sensory systems. For Barsalou a concept can be viewed as a dynamical distributed system in the brain that represents a category in the environment or experience, and which controls interactions with the category's instance, whether in the actual environment or in thought.

6.1. On the Tension Between Traditional and Enactive views.

Despite growing support for this enactive view, there is a tension within the philosophy and science of consciousness as most cognitive scientists still adhere to the traditional "brainbound" view, and the philosophical assumption that the head is where consciousness resides (See Evan Thomson and Diego Cosmelli (2011) for a discussion of this). They prefer to reassert a sense of mystery about the emergence of conscious experience from matter, and champion the quest for a science of consciousness that would close the seemingly unbridgeable gap between the physics of the brain and phenomenal experience (See especially Koch, 2012). There are those who declare that present science has already an explanation in store, e.g. in some exotic interpretation of quantum mechanics (Penrose and Hameroff, 1996, 2014; Stapp, 2009, 2014). Others express their faith in some future, but unforeseeable, scientific advance that will dispel the riddle (Chalmers, 1996). Kant's transcendental enquiry continues to be relevant to this ongoing debate. He held a fundamentally different view of human nature in marked contrast to the reductionist, deterministic, and mechanistic picture that is painted by contemporary accounts. As discussed in Chapter 4, although like modern functionalist cognitive scientists, Kant was concerned with the "functions" or conditions needed for functions to work, his philosophy transcends this in ways that are significant in addressing the ensuing "hard problem of consciousness" that arose from functionalism itself. This had led to the ongoing quest for a scientific understanding of consciousness, where it was presupposed that the brain itself is the ultimate seat of consciousness and where in some peculiar sense, it *resides*. Whereas there is intuitive plausibility in such a position, a quite natural assumption, given that we obviously need a brain to think and also that changes to the brain through mechanical manipulation have been shown to cause changes in phenomenal experience, this is a problematic understanding of consciousness, as has been discussed thus far. As noted in Chapter 2.1., recent empirical work in cognitive neuroscience claims to be able to support this continuing scientific endeavour to find "consciousness in the brain" through its program to discover the NCCs (Crick and Koch, 1995, 1998; Chalmers, 2000; Metzinger, 2000, 2003, 2011; Koch 2004; Block 2005; Bayne 2007, 2010; Tononi and Koch 2000, 2008, 2015; Hohwy 2007; 2009; Kiverstein 2009. Oizumi, Albantakis, Tononi

(2014), Hohwy & Bayne (2015). The notion of NCCs is founded on some early work in neuroscience on epileptic patients. Direct stimulation to the cortex in conscious subjects had been shown to bring about experiences with a very particular phenomenology (Penfield 1954, 1958). The fundamental idea of finding the NCCs is that this can be applied more generally; once more is known about the brain, scientists will discover precisely how consciousness comes about. This endeavour to find the neural correlate of consciousness has become a sustained, intense focal point for scientific research on consciousness. Francis Crick and Christof Koch, on presenting their conception of an NCC stated: "Whenever some information is represented in the NCC it is represented in consciousness" (Crick and Koch, 1998, p. 98).

The search for NCCs is specifically a search for the neural correlates of transitive contents of consciousness, as when looking back and forth between red and blue on a screen, a conscious percept is correspondingly shifting from a red percept to a blue percept. The goal of the NCC paradigm is to find the minimal set of neuronal events that correlates with this subjective shift from redness to blueness. That is, the main purpose is to uncover the neural representational systems whose contents systematically match the contents of phenomenal consciousness. Indeed, as discussed earlier, Francis Crick avers that philosophy is superfluous and has gone so far as to proclaim that "No longer need one spend time attempting (…) to endure the tedium of philosophers perpetually disagreeing with each other. Consciousness is now largely a scientific problem" (Crick, 1996, p. 486). However, the question of what it means to be a neural correlate of consciousness is actually *far* from straightforward, and as has been a consistent fundamental premise throughout this thesis, the study of consciousness involves not only empirical, and methodological issues but also *philosophical* questions about the nature of consciousness and its relationship to the brain. Even if it were discovered that, say oscillations at 40 Hz in a certain brain area, correlate perfectly with behavioural measures of consciousness, the problem simply regresses a further stage: the question would now become, why and how should coherent oscillations ever generate consciousness? After all, coherent oscillations are observed in many other branches of science, where they do *not* give rise to consciousness. The same would apply to Roger Penrose and Stuart Hameroff's Orch Orr theory of quantum computation in microtubules, where

consciousness is said to be based on non-computable quantum processing performed by qubits formed collectively within cellular microtubules, and amplified in the neurons. Despite it being a new physical theory of quantum computation, the problem still remains as to what exactly it is about microtubules that allows them to generate consciousness, when other physical mechanisms do not. Moreover, if the quest for the NCC involves searching for elusive notions like *qualia, phenomenality,* or *experience,* it is not really clear what that would entail. The subjective privacy implicit in the concept of phenomenal experience renders it difficult to even operationalise, let alone explain. The usefulness of operational definitions in science is, after all proportional to the extent they have a solid basis in physical reality. The main worry about operationalising a notion as vaguely defined as "consciousness" is that different theorists have come up with radically different operational definitions. There would therefore seem to be a fatal problem at the heart of the quest for the NCCs; this is because there is no clear and distinct idea of what *exactly* is being sought. As discussed in Chapter 5.2. it is doubtful whether there is even any sense to the notion that "consciousness" actually "picks out" a well defined natural kind, as Chalmers and others claim. Although the search for NCCs sounds reasonable, as it concerns taking the usual, well worn scientific route of starting with correlations before moving on to causal ones, it is fully dependent on the intuition that consciousness is something "extra" and different from the physical processes on which it depends. On the one side of the correlation is the measurement of neural activity using E.E.G. fMRI or other brain imaging technology, on the other "subjective experiences" or "consciousness", and it is not at all clear how the two domains can be connected.

Daniel Dennett is well aware of the lure of Cartesian dualism that beguiles even the most tough-minded and relentless scientific reductionists, and is able to explain clearly and perspicuously, in his own inimitable style, the redundancy of the Neural Correlates of Consciousness research program, that begins from the assumption that consciousness is somehow "produced" or "generated" in the brain. As he notes, scornfully, and with more than a hint of sarcasm:

> An area of philosophical naivete´ in the cognitive science of consciousness concerns the quest for the Neural Correlate of Consciousness (NCC). It has seemed obvious to quite a few scientists aspiring to solve the mystery of consciousness that there has to be an NCC, the necessary and sufficient

> conditions, characterized in terms of locatable neural activity, for conscious experiences. How indeed could there not be one, if materialism is true?" (Dennett, 2009, p. 232).

He makes the observation that the quest for the NCC is in all probability "a wild goose chase"; for although states of consciousness obviously *do* causally depend on states of the brain, one can nevertheless wonder in what sense there is, or could be, such a thing as a NCC and how this could occur at all. Firstly, there is a difficulty in the view that there can be an isomorphism of mental states to brain states and secondly and most significantly, there it is questionable assumption that the cognitive mechanisms supporting conscious experience are located completely within the skull. In *The New Phrenology* (2001) psychologist Dr William R Uttal, Professor Emeritus at Arizona State University and Professor Emeritus (Psychology) at the University of Michigan, addresses this question of *localisation*, i.e. whether psychological processes can be defined and isolated in a way that permits them to be associated with particular brain regions. He likens neuro-imaging research to the practice of phrenology of the 19th century and argues that the reason it would be difficult to localise functions to specific brain areas is that the brain activity underlying cognitive processes is widely distributed, much of the brain responding to any stimulus, making it difficult to localise functions to specific brain areas. Also, there is the problem that every brain area participates in multiple functions, which means that asking *where* a given function occurs in the brain is simply the wrong question.

Others have also questioned the "brainbound" view in ways that render doubtful whether there is any sense at all to the notion of a minimal neural substrate or NCC sufficient to produce experience. Susan Hurley, in her 1998 book *Consciousness in Action*, as well as in many articles, asked why it should be considered that the boundary of skin and skull is significant when it comes to explaining consciousness. Hurley defended "vehicle externalism", the view that cognition does not necessarily have to be explained in terms of internal processes. "Brains are in continuous causal interaction with their bodies and their environments" so "[w]hy should dynamics distributed within a pre-specified boundary be capable of explaining qualities, while those beyond not? (Hurley, 2010, p. 116). Similarly, Alva Noë, following Thompson and Varela, brings out the insight that perceptual content depends on the skilful

activity of the whole animal or person acting in time and spatial orientation, making use of its capacities for eye, head, and whole body movements and for directed attention (Noë, 2010). He claims that although it is trivially true that states of consciousness causally depend on states of the brain, for everyone needs a brain to function, there is no reason to think that the neural states that have been shown experimentally to be correlated with conscious experiences match those experiences' content. It is perfectly obvious that something has to be happening in neurons every time an animal has an experience, but this something is neither identical to nor necessary and sufficient for the experience.

According to Chalmers, "the neural subsystem N will be sufficient for the occurrence of the conscious state under conditions C. The NCC will be the neural state that is minimally sufficient for the corresponding conscious content" (Chalmers, 2000, pp. 24-5). Metzinger, in the same volume, also assumes that empirical claims about NCCs will be formulated as laws stating which conscious contents will follow "with nomonological necessity" from the activation in subsystem N (Metzinger, 2000, p. 285). However, if we consider the notion of egocentric space, the space within which the perceiving organism acts in terms of vision, (the study of NCCs focuses on the study of visual perception), one of the most obvious things to note is that people are generally free to move about when viewing an object. The structurally coherent content with which we are concerned in visual perception is that as of a figure-on-a-ground that is *located within egocentric space* (see Merleau-Ponty, 1962, p. 101). This egocentric space, within which we visually perceive an object, is defined by one's whole body and the possible ways it can move. It is significant that Merleau-Ponty acquired this idea from Kant and Varela and colleagues built on Merleau-Ponty's work to develop their enactive approach.[71] In fact, Merleau-Ponty gives an argument for this thesis which corresponds closely to Kant's argument regarding space and time in the metaphysical expositions of the Transcendental Aesthetic: the body's permanence cannot be "a necessity of fact, since such necessity presupposes it", and "factual situations can only impact upon me if my nature is already such that there are factual situations for me" (ibid., pp. 91). Thus, Merleau-Ponty, in extending the notion from Kant of spatial consciousness, brought out the implicit notion that to have a spatialised self-consciousness is to experience oneself as a body, i.e. an agent.

In the Metaphysical Exposition of Space of the first *Critique* Kant stated that our experience is necessarily spatial, as of a world presented outside of us:

> In order that certain sensations be referred to something outside me (*auf etwas außer mich*) (…) and similarly in order that I may be able to represent (*vorstellen)* them as (*als*) outside and alongside one another (…) the representation (*Vorstellung*) of space must be presupposed (A23/B38).

In his *Phénoménologie de la Perception* Merleau-Ponty takes this notion of spatialised self-consciousness and extends it as the self consciousness of a body, and his discussion heightens this Kantian notion: as bodies, we are not just "in" space, but are "of" it; that is, self-consciousness is the self-consciousness of a body, something inherently extended in space.

In everyday experience, although we are prone to separate spatial experience and self-consciousness, rendering the former "the world of things", the other the world of consciousness, these are not two different realities that might somehow come together, but at a fundamental level, they cannot be defined separately from each other. When we think of ourselves as embodied agents, we have a rather different sense of "self-consciousness" than when we think of ourselves as mere introspectors. If we take this bodily agency as the primary form of self-consciousness, then we are able to see self-consciousness not as a clear and isolated imaging or objectifying of the self, but as a lived sense of practical engagement. We can see that the embodied experience of an agent in his or her environment plays a fundamental role in human thought, which in turn sheds light on the cognitive abilities that we display every day. Also, we can appreciate that many of the abilities that enable us think about the world, for example, language use and gesture, actually originated in the bodily experience of individuals as members of a species. Merleau-Ponty views the body as not simply an object in the world like other objects but, following Kant, as the very condition of the possibility for understanding the world. Without the body providing the *centre* or *situatedness* for experience and agency there would be no perception or conceptualisation of the world. The body as a whole is *a priori* in the sense that it precedes and upholds all experience. Noë, following Varela and Merleau-Ponty/Kant, questions whether there is any coherent sense in the

idea of a neural representation having, as part of its "content", a stimulus with which it usually causally covaries. As he writes "Although neural systems causally enable the animal as a situated agent to orient itself in its egocentric space, they themselves do not inhabit this space, nor do they have any access to it as such" (Noë, 2004. p. 15). Perceptual content is intrinsically first person *experiential*, in the sense that the content of an experience is always the content as represented from a particular point of view. Thus, to have a visual experience as of a book on a table is to have an experience as of the book as standing in a certain egocentric spatial relation to the perceiver and as standing out against a background relative to them. In other words, although the brain and brain processes are necessarily involved in the interactive relationship that constitutes awareness and perception, they are not sufficient for them. The brain and body are so dynamically entangled in the causation and realisation of consciousness as to be explanatorily inseparable from it. Equally, this knowing of what we might term "the self-as-body" is inherently spatial. When we eat a meal, we know *where* our knives and forks are in our hands with respect to the food on the plate and how to manoeuvre it into our mouths. The act of eating is necessarily spatial navigation, and the spatiality involved does not just characterise the relationship of me to what is external to me, but equally characterises the relationship of me to myself, as an embodied agent. The upshot of this way of thinking is that our experiences are not things that happen in our heads, but are what happen in agent-environment systems. That is, embodied cognition construes conscious experiences in terms of when situated agents pick up information about "affordances" (Gibson, 1979) through coupled and dynamic interaction with the environment in which they are embedded.

The NCC research programme, as heir of the Cartesian tradition, in which consciousness and brain-processes are considered as separable from each other, rests on the assumption that something can be found out about cognition through assimilating consciousness to snapshot-like phenomenal episodes in the brain (Metzinger, 2003). Yet mental states heavily depend on information from the body and on its developmental history within its own lifetime as well as on fluid interaction with its physical and social environment in the present. It is not possible to individuate conscious states without at least tacitly taking these past or present interactions into consideration. The fundamental message from embodied enactive

cognitive science is that in order to understand consciousness it is not sufficient to simply consider the brain. There is need to consider the embodied, situated life; the biological substrate of consciousness is the whole organism in its dynamic interaction with the environment, not the brain taken in isolation from the bodily situatedness in which it finds itself. Neuroinformatics and computational neuroscience makes use of increasingly sophisticated technology, from neural modelling and numerical data analysis to study the brain, as well as ultra sophisticated imaging techniques: electroencephalography, fMRI (functional magnetic resonance), near-infrared spectroscopic and chemical shift imaging. Nevertheless, measuring devices or brain imaging technology, no matter how cleverly constructed, will not reveal consciousness because that is not where consciousness is. This is the wrong level of analysis. What we call consciousness is not something that happens in the head but unfolds in the dynamic coupling or relationship between embodied action and the world.

It is often said that Kant was not at all concerned with *qualia*, since his main enterprise was concerned with the question of how knowledge is possible, i.e. the objective aspect of the *Critique* which is epistemological, and that he does not in any way commit himself to an ontological claim regarding the nature of consciousness. However, according to Steven Palmquist, this is wrong. Kant "appeals throughout the Aesthetic to various examples of perceived objects, as being *externally given* to the human subject in the process of experiencing them, and he suggests that "intuition" refers to the requirements of our *bodily functioning*" (Palmquist, 2013, p. 8) which is another term for what we might nowadays call qualitative sensation processing, so that Kant's "forms of intuition" are an account of the structure of our embodied perspective on the world. On this reading, Kant's basic claim in the Aesthetic is that "the forms of intuition are not mental operations performed on sense data but are the formal structure of spatio-temporal relations in which objects stand in relation to the body. As such, this first stage of Kant's theoretical system sets out the basis for a rudimentary physics that is primarily *ontological* and thus *physical* in its emphasis, not merely epistemological and mental" (ibid., p. 9). Matthew Rukgaber also claims for the significance of embodiment in Kant's thought: "[t]he ideality and subjectivity of space is concluded to be an account of the perspective

relative nature of the figure-ground relationship or how it is that objects emerge for us in empirical experience as being orientated in a spatio-temporal field" (Rukgaber, 2009). There are several passages in the *Critique* where it is implied that a discussion of the senses is to be understood as a discussion of the body, such as when Kant describes a representation of sense as "a force of nature" (A294/B350). Viewed this way, the transcendental psychology in the *Critique* can be regarded as first stage of Kant's theoretical system that sets the foundations for a theory of mind that is also *ontological* and thus *physical* in its emphasis, not merely epistemological. In other words, Kant's theory of the "forms of intuition" can be considered as the beginning of an account of the structure of our embodied perspective, which was what Merleau-Ponty understood from his reading of Kant and subsequently taken up by Varela and others.[72] Weber and Varela (2002) also cogently note that in Kant's third critique, the *Critique of Judgment* (1790) he presents a strikingly modern self-organisational account of life. There he claims that life cannot be derived at all from the mechanistic laws of Newtonian physics, which merely postulate "efficient causes" and not "end causes".

One of the aims of this work is to defend the point that Kant was the first post-Newtonian philosopher to attempt to fully address the basic philosophical problem of reconciling free-will with universal natural determinism. Kant, from the beginning, was a convinced Newtonian in physics and natural philosophy, and his project in the *Critique of Pure Reason* was to provide it with a philosophical foundation (See Friedman, 2013). [73] Of particular significance is a small treatise written by Kant, the *Metaphysical Foundations of Natural Science*, in which he presents the principles that are specifically required for experience of material nature. These principles include what Kant calls the "laws of mechanics", which determine how a material body communicates motion to another by means of its moving force. According to Michael Freidman (Freidman, 2013) it was Kant's aim in this treatise to provide a deeper philosophical understanding of Newton's work by providing it with a metaphysical foundation using a radically transformed version of Leibnizian metaphysics. Kant clearly shows considerable interest in various attempts to reconcile certain aspects of Leibnizian metaphysics with the Newtonian view of nature. But, Kant claimed, we cannot envision explaining generation or organic

growth mechanistically and decried Leibniz' idea of a monad as "spiritual automaton", as no more than a "freedom of the turnspit" (*CPrR* 5: 97) which reduces man to a "marionette" (*CPrR* 5:101), and morality to a figment of the imagination. We can indeed view ourselves as machines, responding to the environment in predetermined ways. However, we are not compelled to do so, and are able to regard ourselves as agents who initiate trains of events. Kant was concerned throughout his lifetime with the question of how to make sense of living things and came to regard biology as a non-mechanistic life science that supplements Newtonian, determinist, mechanistic science with the teleological concept of a natural purpose (*CPJ* 5:369 415). This is intimately connected to his thesis that there exists an irreducible explanatory gap between inert "mechanical" nature and living nature (*CPJ* 5:369 415).

He writes:

> It is quite certain that we can never adequately come to know the organized beings and their internal possibility in accordance with merely mechanical principles of nature, let alone explain them; and this is indeed so certain that we can boldly say that it would be absurd for humans ever to make such an attempt or to hope that there might yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws (*CPJ* 5:400).

By this is meant that since our human mental lives entail our biological lives, since in order to think there needs to be a biological body, there can also never be a Newton of the human mind. So again, our psychological lives; including, importantly, our power of choice, is free from the determining influence of mechanistic causality. Kant's insight into the "transcendental subject" is his way of reconciling a mechanistic-reductionist, hence deterministic theory of mind with the idea of human freedom. Kant rejects scientific or reductive naturalism, which says that science is, as Wilfrid Sellars formulates it, "the measure of all things" (Sellars, 1963, 1991). Newtonian classical physics had inspired a belief in a deterministic universe external to the human observer who has a passive role to play. In the second *Critique*, the *Critique of Practical Reason,* Kant attempts to unite the idea of human freedom with the mechanistic laws of nature. In so doing he articulates a law of a possible order of nature that is beyond the realm of experience. This super-sensible aspect of nature provides not only the determining grounds for the existence of objects, but also at the same time becomes the freedom of a rational being. Kant is

here committing to the thesis that even allowing for the existence of universal, transcendental laws of nature, and also for the existence of general mechanistic laws of nature, it does not automatically follow that there are specific empirical laws of nature "all the way down". Robert Hanna has made the point that for Kant transcendentally free rational human choices produce what he calls "natural causal singularities" (Hanna, 2006b) and one-time laws, and thereby freely complete nature. As he writes in an exchange with A.W Moore: "T]ranscendentally free agents thus create new unique empirical causal-dynamic laws of nature that fall under, and are permitted by, but are not *compelled or necessitated by,* the general laws of natural mechanism" (Hanna, 2007, p. 121). This is because, he claims, Kant regards biological life and spontaneity of the will as conjunctive and intrinsic structural properties of personhood and rational agency.[74] At the end of the first *Critique,* in the Canon, he emphasised the point that reason is immanently self-developing (A835/B863).[75] This suggests that human cognition involves a necessary relation between the elements of cognition and its overall grasp by a human subject. This is an integral relationship, and one which is often described by Kant as "purposive". Kant took it be the case that such "purposive" relations were not available to machines and this is an important topic within the *Critique of Teleological Judgment* in the third *Critique*. Also in the section of the first *Critique* entitled *The Exposition of the Cosmological Idea of Freedom in Harmony with the Universal Law of Natural Necessity*, Kant discusses the notion that rational human agents are necessarily also rational human living organisms, i.e. biological animals capable of intentionality whose rational mindedness and rational directedness towards objects in the world, other real persons, and themselves, is fully continuous with this.

After writing the first *Critique* Kant realised that not everything could be so neatly subsumed under the *a priori* principles of pure reason. This fact apparently concerned Kant more than many of his followers, who did not see beyond the theoretical frame of the *Critique of Pure Reason*. But for Kant himself it was a concern. It was especially the empirical and not *a priori* character of biology that posed a grave problem. In his last philosophical writings, the *Opus Postumum*, he refers to the work undertaken in the *Critique of Pure Reason*. Without invalidating the *a priori* categories that had been the possibility of all knowledge, he finds an entirely new foundation for them: the lived body. On the Kantian picture of physical

nature presented in the *Opus postumum*, the complete set of general causal laws provides a skeletal causal-dynamic architecture for human nature. His main point is that we cannot understand ourselves as having experiences in the first place if we were not organisms equipped with our various senses. Here Kant emphasises the fact that transcendental philosophy cannot depend on the object given in intuition but must *produce* it in order for it to be "for-a-subject"; the organism mediates between the object and the faculty of the understanding, but the mediating organism does not receive its organisation by being affected by the object, but rather through the spontaneous self-affection or action by which the subject produces its own object of knowledge. He writes: "The subject (object in the appearance) affected by empirical intuition is, insofar as it affects itself according to concepts, an organic body intuited according to the five senses" (*OP* XXII 388: 3–6). The organism does not receive its actuality from the affecting object, nor from sensual intuition, but from the fact that the organism is necessary *for the sake of experience*. Kant is very clear that for the sake of a doctrine of natural science "in the subject an organic principle of the moving forces" is presupposed "in [the form of] universal principles of the possibility of experience" (*OP* XXII 373:1–4). This does not contradict the strictures of his transcendental idealism as he makes clear that we can neither prove nor postulate the possibility of our organisation (*OP* XXII 481:8–9) and that we simply know ourselves "in experience as an organic body" (*OP* XXII 481:10).

For Kant mind is explanatorily and ontologically continuous with life, in the sense that whatever is metaphysically required for a human mind is also present in its biological life. [76] The thesis of epigenesis in biology says that biological material is initially unformed and that form gradually emerges through the non-predetermined or relatively spontaneous operations of an innate endogenous organisational or processing device in interaction with its environment. As mentioned on pp. 40-41, Kant would endorse this and explicitly defends the theory that biological life is epigenetic.[77] He extends this theory analogically to his theory of cognitive innateness (*CPJ* 5: 424) (B167). There is also in Kant's writings a strong continuity between biological life and the spontaneity of action, combined with an emergentist and non-reductive approach to biological life: In the third *Critique, the Critique of the Power of Judgement,* in order to explain the behaviours and natures of living organisms, including the behaviours and natures of rational human animals, Kant claimed that

there is a theoretical obligation to posit the existence of causally efficacious emergent properties that naturally arise from self-organising complex dynamical systems.

> Life without the feeling of the corporeal organ is merely consciousness of one's existence, but not a feeling of well- or ill-being, i.e., the promotion or inhibition of the powers of life; because the mind for itself is entirely life (the principle of life itself), and hindrances and promotions must be sought outside it, though in the human being himself, hence in combination with his body (*CPJ* 5: 278).

> For a body to be judged as a natural purpose in itself and in accordance with its internal possibility, it is required that its parts reciprocally produce each other, as far as both their form and their combination is concerned, and thus produce a whole out of their own causality, the concept of which, conversely is in turn the cause (in a being that would possess the causality according to concepts appropriate for such a product) of it in accordance with a principle; consequently the connection *of efficient causes* could at the same time be judged as an *effect though final causes*. In such a product of nature each part is conceived as if it exists only *through* all the others, thus as if existing *for the sake of the others* and *on account* of the whole, i.e., as an instrument (organ), which is, however, not sufficient (for it could also be an instrument of art, and thus represented as possible at all only as a purpose); rather it must be thought of as an organ that *produces* the other parts (consequently each produces the others reciprocally), which cannot be the case in any instrument of art, but only of nature, which provides all the matter for instruments (even those of art): only then and on that account can such a product, as an *organized* and *self-organizing* being, be called a *natural purpose* (*CPJ* 5: 373-374).

> Strictly speaking, the organization of nature is (…) not analogous with any causality that we know (*CPJ* 5: 375).

> It might always be possible that in, e.g., an animal body, many parts could be conceived as consequences of merely mechanical laws (...) Yet the cause that provides the appropriate material, modifies it, forms it, and deposits it in the appropriate place must always be judged teleologically, so that everything in it must be considered as organized, and everything is also, in relation to the thing itself, an organ also (*CPJ* 5: 37).

The computationalist, or functionalist conception of cognition is necessarily reductionist; mind is akin to some form of computational mechanism, i.e. the mind is

the software and the brain the hardware. The basic tenet here is that all cognitive functions are at bottom a set of rules for handling symbolic entities that represent items of the world. Computationalist/ functionalist or neuro-reductionist approaches generally lead to a paradoxical eliminativism, i.e. the elimination of consciousness as the domain of our subjective experience during the very process of explanation. From this point of view, the mind is necessarily "in the head", but is reduced to representational mechanics, a system of inputs and outputs. This is problematic as consciousness itself is seen as an illusion or a fiction. The "I" that is supposed to be the proper subject of experience, thought, and action doesn't exist. Functionalists like Dennett and Metzinger claim that cognitive neuroscience eliminates the self as bearer of consciousness and controlling agent by revealing the complex, decentralised functional activity in the brain that is actually responsible for behaviour. Chalmers, on the other hand, avers that consciousness cannot be thus eliminated, it is an irreducible feature of reality; hence the "hard problem" that arose in cognitive science of explaining how the amazing, private world of consciousness emerges from neuronal activity in the brain. However, from this new embodied perspective, cognition appears as a dynamical process (and not a syntactic one) of real time variables with the capacity for self-organisation (and not as representational machinery). Moreover, the mind is not "in the head" since its roots are in the body as a whole and also in the extended environment where the organism finds itself. This means that the constitution of a mind is always concurrent with the extended presence of other minds in a social network and the environmental world. In fact, what is present in Kant, finds a convergent development from current philosophy of biology and the scientific notion of autopoiesis. On this view instead of describing consciousness as an illusion, as is commonplace in reductionist science, there is a new understanding of mind in the form of immanent teleological "presence" with truly biological features, inevitably intertwined with the self-establishment of a conscious identity which is at the same time the living process. It is argued that this understanding of cognition has its beginnings in Kant, who was committed to an active, "sensorimotor" view of consciousness, realising the spontaneity of a rational agent through acts of "synthesis" or binding in an embodied biological system. He views the human mind as metaphysically continuous with biological life, recognising that rational human agents are necessarily also rational human living organisms, i.e.

biological animals capable of intentionality whose rational mindedness and rational directedness towards objects in the world, other real persons, and themselves, is fully continuous with this (see Hanna and Maiese, 2009).

One pertinent philosophical question that has risen from enactivism is whether or not a machine could be conscious. Questions about artificial consciousness have important social and moral ramifications as more and more robots and artificial human-like agents are produced. This is the topic of the next section.

6.2. Conscious Machines

The previous section discussed the modern theory of enactive, embodied cognitive science, which is anti-dualist; perception is inherently active and cognition is embodied and situated. According to the original founders of enactivism (Varela, Thompson, & Rosch, 1991), the body is not only the living structure to experiences, but also the setting for cognition, no longer to be viewed as the manipulation of symbols, but as the regulation and coordination of emergent autonomous animal agency. Consciousness is defined in terms of an autonomous agent which, through a self-producing network of processes, constitutes its own identity (Ziemke, 2007b, Thompson, 2007). This means that consciousness arises as part of the process of an embodied entity interacting with the environment in precise ways that are determined by its physiology. The agent does not represent the world "in the head" or brain, but produces it through the nature of its unique way of interacting with its environment. The seeds of this theory can be readily discerned in Kant's *Critique of Judgement* in the way in which he also views the mind as an autonomous system whose interaction with the world furnishes it with an intrinsically meaningful perspective on its environment. Since the enactive framework incorporates both biological agency and phenomenological subjectivity, it allows the traditional mind-body problem that is a direct descendant of Descartes to be recast in terms of what has recently been called the "body-body problem" where "[t]he scientific task is to understand how the organizational and dynamic processes of a living body can become constitutive of a subjective point of view, so that there is something it is like to be that body" (Thompson, 2007, p. 237).

However, importantly, enactivism is pursued in two different styles: *sensorimotor* enactivism (Alva Noë et al) and *autopoietic* enactivism (Evan Thompson et al.). Both types take it to be a fundamental commitment that cognising agents are to be viewed as situated in an irreducibly meaningful world, where meaning is "enacted" or constituted through the tightly-coupled dynamic relationship between an agent's brain, body, and environment. Sensorimotor enactivism, however, makes a *constitutive* claim; that perceptual consciousness is *constituted* by the exercise of sensorimotor capacities (O'Regan & Noe, 2001, O'Regan, 2011; Denegaar & O' Regan, 2015). This makes use of the notion of "sensing" capacities of the sort that can be displayed by simple artifacts; for example, as with artificial sensing devices such as missile guidance systems that are said to "sense" their target. Human beings, in a similar way, through the use of the bodily senses, also have the capacity to engage with the environment and to express their "sense capacities" in a broad range of behaviours. The claim of sensorimotor enactivism is that perceptual capacities depend on the implicit grasp of sensorimotor dependencies. However, it seems evident that full-blown consciousness requires, in addition to this, that the perceiver should be able to potentially use information from the environment in rational planning, goal directedness and deliberation on deciding a course of action. In short, it seems that consciousness involves additional criteria or constraints beyond mere perceptual capacity. The stronger theory of autopoietic enactivism harnesses the notion of autonomous agency that proposes additional criteria related to the origin of the system's tendencies. As Evan Thompson notes, sensorimotor capacities must be the capacities of an agent or self and "agency and selfhood require that the system be autonomous" (Thompson, 2007, p. 417) which he spells out in autopoietic terms. Moreover, conscious experience or "first person givenness" involves self awareness for experiences to be "phenomenally manifest as mine" (ibid., p. 420).

As discussed earlier, Chalmers claims that artificial consciousness might be possible, since the right kind of computations are sufficient for the possession of a conscious mind. At the crux of Chalmers' non-reductive functionalism is the Principle of Organizational Invariance, which asserts that "given any system that has conscious experiences, then any system that has the same fine-grained functional organization will have qualitatively identical experiences" and also that "[s]ystems

with the same causal topology (...) will share their psychological properties" (Chalmers, 1995, p. 232.) This kind of argument supports strong AI, that artificial silicon creatures with the right functional organisation could be in possession of a conscious mind. The idea of strong AI ignores the essential role of the body and its dynamic, goal directed movement in giving rise to the feeling of consciousness. Sensorimotor enactivism goes beyond this and emphasises the constitutive role of action for "consciousness in general" based on three principles: (i) the body is not an object that can be represented; (ii) the presence of the body is the presence of the body in the world, and (iii) the body we experience is the body in action. In other words, that in order for a machine to have first person phenomenal experience, it would require a body with a sensory system, a nervous system, the ability to move and sense its environment, react to it and bring about further changes within its world as a result (see Stuart, 2007, 2007b). It is claimed, by Jan Denegaar and Kevin O'Regan, that sensorimotor enactivism forms a default position for an enactive account of perceptual consciousness, and that even "the particular quality of experience e.g. what makes an experience the experience of red, lies in particular patterns of sensorimotor engagement" (Degenaar & O'Regan, 2015, p. 2). This is said to offer a framework for thinking about perceptual consciousness that applies both to living organisms, as well as, potentially, to artificial systems. On this sensorimotor view, autopoietic organisation, although relevant in living organisms, is not itself essential for perceptual consciousness. But one difficulty with this is that sensorimotor loops alone cannot provide the conceptual means of distinguishing between the intentional action of an autonomous agent and mere accidental movement. Perceptual consciousness is linked to purposive behaviour that is intrinsically meaningful for the system. If we refer back to the example of the target seeking missile it is clear that, unlike a living organism, it lacks purposive agency.

As Hans Jonas notes in his critique of cybernetics, autonomous agency depends on "whether effector and receptor equipment – that is motility and perception alone – is sufficient to make up motivated animal behavior" (Jonas, 1966, p. 117). This further depends on "whether the mechanism is a "whole", having an identity or self-ness that can be said to be the bearer of purpose, the subject of action, and the maker of decisions" (ibid., p. 118). This is because in order for effector and receptor to constitute intrinsically purposive action there must be interposed between them a

centre of "concern". An artificial system consisting only of sensors and motors that are coupled together in some manner, and which for reasons of design does not have to continually bring forth its own existence under precarious conditions, cannot be said to be an individual subject in its own right in the same way that a living organism can. To quote Ziemke "despite all biological inspiration today's adaptive robots are still radically different from living organisms. In particular, despite their capacity for a certain degree of self organisation today's so-called "autonomous agents" are actually far from possessing the autonomy, and consequently the embodiment of living organisms" (Ziemke, 2001, see also Ziemke, 2016). He points out that when we refer to both robots and living organisms as autonomous agents it is important to keep in mind that their respective "autonomy" and "agency" is fundamentally different. We should adopt a position of "caveat spectator" and not take similarity of behaviour for similarity of underpinning. The underpinnings, the biology, the internal constitution and regulation are crucial because "the way an organism constructs itself also shapes the way it constructs its self" (Ziemke, 2007a).

At a 2001 workshop called *Can a Machine be Conscious*, organised by the Swartz Foundation, it was declared that their mission statement was "to explore the application of mathematical physics and computer engineering principles to traditional neurobiology, as a path to better understanding the brain/mind relationship". The universal consensus of philosophers, neurologists and computer scientists attending this workshop was that "we know of no fundamental law or principle operating in this universe that forbids the existence of subjective feelings in artifacts designed or evolved by humans" and "that one day computers or robots could be conscious" (Koch 2001), which is currently considered an "open question". In fact, Professor Steve Torrance, founder member of the Centre for Research in Cognitive Science at Sussex University (COGS), avers:

> It is a matter of some dispute whether the defining properties of autopoiesis can be found outside the realm of the truly biological, and it is thus an open question as to whether there is any sense in which computationally based constructs could ever be seen as being assimilable to an autopoietic framework – that is as original self-enacting loci of meaning and purpose, or indeed of consciousness (Torrance, 2005).

Since then, although the possibility of conscious machines as well as the ethical ramifications of such an eventuality has been frequently discussed, nobody has claimed that anything close to consciousness has occurred in an artefact or machine. As Merleau Ponty theorised in the *Phénoménologie de la Perception*, there is a sense of an *a priori* that belongs to living existence itself, and this is prior to scientific reflection, in fact, prior to all reflection. This is the sense of embodiment, often referred to as the body schema. Merleau-Ponty, extending the notion from Kant of spatial consciousness, brought out the implicit notion that to have a spatialised self-consciousness is to experience oneself as a body, i.e. an agent. A body schema is a pre-attentional, sub-consciously monitored "real time representation of the body in space, generated by proprioceptive, somasensory, vestibular and other sensory inputs (Schwoebel et al., 2001) which form a core consciousness or self. The body schema is the embodied understanding of the whole of the organism's parameters and the constraints of embodiment as such. This idea can be seen in Kant where he infers that mind is inexorably linked to the life of the organism. Although he discusses the notion of the representation of biological life in *The Critique of Judgement* in mechanistic terms, as having a kind of semantic content of a kind that might be seen to be capable in principle of being instantiated in a Turing machine or of being *functionally describabl*e, he also emphasises the first personal phenomenal character of experience, which as we have seen, he calls "the feeling of life" and which is what gives rise to the purposive, goal-directed or teleological behaviour of an organism. Kant's idea appears to be that *both* the semantic or mechanistic content of the representation of biological life and the phenomenal character or consciousness of the feeling of life are necessarily and mutually bound up with each another; that consciousness and intentionality are mutually intertwined *via* the neurobiological life of an embodied animal mind. The first person perspective-ness or "me-ness" of conscious cognitive content is not simply the apperceptive self-awareness of a "judgement" that is reliant on the categories, as in the first *Critique* but is also necessarily bound up with the spatial orientation and temporal asymmetry of a self regulating biological organism. In other words, the formal intuitional spatiotemporal structure of conscious cognitive content just *is* its subjective character, and this is achieved *prior* to the use of the categories in apperception. Moreover, it seems likely that cognition is coextensive with homeostatic metabolic and visceral processes and

that a non-homeostatic, artificial silicon creature, even one with "the right kind of functional organisation" could not be said to truly be in possession of a conscious mind. Mechanistic robots are simply cognitive tools or models rather than entities endowed with cognition and true agency. Furthermore, human intentionality is intrinsic, that of mechanistic artefacts, derived. [78] This has a bearing on how we are to understand enactivism and the benefit of this "Kantian" construal is that it would also help explain consciousness in non-human animals. This is the topic of the next section.

6.3. Animal Consciousness and Non-conceptual Content

Although Kant wrote that intuitions and concepts are cognitively complementary and semantically interdependent for the specific purpose of constituting objectively valid or empirically meaningful judgments (A50-51/B74-76), it is argued that it does not follow that there cannot be "empty" concepts or "blind" intuitions outside that. [79] Kant also held the view that receptive experience is necessarily conditioned by bodily orientation in space and time in a manner which is necessarily non-conceptual, therefore the content of perception is also partially sub-rational. In other words, it is the human or animal organism's spatiotemporal perspective or "unique point of view" that is fundamental to the subjectivity of its experience, not "the unity of consciousness" by which we judge that such and such. It is obvious that non-human animals are conscious and function in the world without having the capacity to make conceptual judgements, and this would account for this capacity in non-human animals, which no one would deny. Construed this way we can see that for Kant the underlying nature of cognitive content is not exhausted by its *a priori* functional components. The pure *a priori* forms of intuition, space and time are involved in both rational cognition in which we make perceptual judgments by means of the categories of the understanding but also in *sub-rational* cognition. Although formal intuition or judgement strictly requires a capacity for self-conscious rational cognition and self ascription, the latter is shared by *all* biological creatures, not simply humans. That is, Kant held a certain non-conceptualism about intuition, which entails that judgmental rationality has a pre-rational or sub-rational cognitive grounding in more basic non-conceptual cognitive capacities that both humans and

animals share. As Robert Hanna avers, non-conceptualism says that "(a) there are certain cognitive capacities which are not determined (or at least not *fully* determined) by conceptual capacities, and (b) that the cognitive capacities which outstrip conceptual capacities can be possessed by rational and non-rational animals alike, whether human or non-human" (Hanna, 2006, pp. 84-85). [80]

Kant's thesis of non-conceptual content is to be located in the Transcendental Aesthetic section of the first *Critique* where he traces back to the forms of space and time, the pure "forms of intuition", which are the necessary *a priori* conditions of every mental representation generated by sensibility. These representations of space and time are not only presupposed by all non-conceptual content but also account for the existence, cognitive significance ("objective validity"), and psychological coherence of our world. In Kant's earlier argument from incongruent counterparts (to be found in his pre critical essay "Concerning the Ultimate Ground of the Differentiation of Directions in Space," (Kant, 1768)) it was clear that some parts of our cognition of objects in the world must depend on how they are given to us in a manner that is necessarily non-conceptual. Also in the Metaphysical Exposition section of the Transcendental Aesthetic, Kant demonstrates that the pure formal intuitions of space and time already underlie any representations, for we cannot even perform geometrical or mathematical calculations in the head without their possibility (A19/B33-A49-B66). That is, there is a second constitutive element of cognition beyond the intellect - we can have direct non-conceptual representations of the forms of intuition, what he calls "pure intuitions" and sometimes "formal intuitions" of space and time. Spontaneous apperception through the application of the pure categories cannot be what exhaustively presents the world to us. This mean that for humans, the for me-ness of conscious cognitive content is essentially bound up with a pre-rational intuition of *spatial orientation* and *temporal asymmetry*. As mentioned, the formal intuitional spatiotemporal structure of conscious cognitive content just *is* its subjective or "first person" character, and this is achieved prior to the use of the categories in apperception. Therefore, Kant is, in effect, saying that in light of the cognitively basic nature of non-conceptual content, a necessarily concept-related capacity for the unification of "a manifold" is not necessary. Rather, it is an animal's spatiotemporal perspective or *unique point of view* that is basic to the subjectivity of its experience. This accounts for animal consciousness; non-

human animals are obviously conscious and function in the world without having the capacity to make conceptual judgements through the application of the categories of the understanding. Kant attributed the capacity for objective perceptual awareness to non-human animals, despite their lack of conceptual capacities, i.e. there is room in the Kantian system for the possibility of objective conscious awareness in non-rational animals that is perceptual without being essentially conceptual in nature. That Kant allows that animals have the capacity to represent their environment can be gleaned from the following passage in the *Critique of the Power of Judgment*:

> Yet from the comparison of the similar mode of operation in animals (the ground for which we cannot immediately perceive) to that of humans (of which we are immediately aware) we can quite properly infer in accordance with the analogy that animals also act in accordance with representations [*Vorstellungen*] (and are not, as Descartes would have it, machines), and that in spite of their specific difference, they are still of the same genus as human beings (as living beings) (*CPJ* 5:464).

Also in the recently translated *Lectures on Metaphysics*, he writes:

> We call an animal alive because it has a faculty to alter its own states a consequence of its own representations. Someone who maintained that in animals the principle of life has no power of representation (vim repraesentativam), but rather that they act only according to general laws of matter, was Descartes, and afterwards also Malebranche, but to think of animals as machines is impossible, because then one would deviate from all analogy with nature... (Metaphysik Volckmann (1784–5) LM 28:449; cf. An 7:212).

For it is the two forms of intuition, space and time that are involved in both rational cognition and in the *sub-rational* cognition of both human and non-human animals. The former constitutes an objective unity of consciousness by virtue of its conceptual logical form, requiring a capacity for self-conscious and self-ascription and by which we make perceptual judgments by means of the categories of the understanding, the latter is shared by *all* biological creatures, rational and non-rational animals alike.

Thus Kant held a certain non-conceptualism about intuition, which entails that our capacity to form "judgements" has a pre-rational cognitive grounding in more fundamental "non-conceptual" cognitive capacities that we share with various non-human animals. As Bermúdez states, "allowing that a creature's representational

capacities can outstrip its conceptual capacities makes it possible for philosophers and cognitive scientists to study aspects of cognition and behavior that remain outside the scope of more traditional approaches" (Bermúdez, 2003).

6.4. On the Role of the Imagination.

As mentioned earlier (Section 3.4.), the role of imagination is a contentious issue in Kantian exegesis but can be interpreted as playing a significant role in the pre-reflective self-awareness and orientation of our bodily selves in the environment. According to Susan Stuart, it is imagination that enables us to build up a pre-reflective bodily expectation about how our experiential world will continue to be, and can be thus interpreted as playing a role in both the intellectual and the non-intellectual pre-reflective awareness of our bodily selves in the world. So, besides the reproductive imagination and the productive or cognitive imagination, there is also a bodily or muscular pre-reflective imagination. In fact, she writes, from the point of view of the agent's conscious resources, it "makes more sense that a bodily or muscular imagination acts in a pre-reflective sensory, that is, visual, olfactory, audial, gustatory, tactile, and kinaesthetic manner. It is an imagination that makes our bodily consciousness possible because it facilitates the experiential interdependence between our thoughts – unity of consciousness – and our world – consciousness of unity, and it is the perceiving and reacting body, the enactive system, which occupies this illusory position being both subject and object of consciousness"(Stuart, 2008, p. 8).[81]

That is to say that imagination in the Kantian framework can be interpreted as playing a significant role in the non-intellectual "grasping", pre-reflective self-awareness, and orientation of our bodily selves in our world. In a similar vein, Angelica Nuzzo proposes that imagination is pre-discursive and embodied: i.e. "transcendental embodiment" refers to the "pure, a priori dimension of our sensibility (cognitive, practical, and aesthetic) – a dimension that is irreducible to purely mental activity and is necessarily embodied" (Nuzzo, 2008. p. 7). According to Nuzzo, Kant grounds the "humanity of reason" in the distinctly human experiences made possible by the *a priori* dimension of human body. Both Stuart's

and Nuzzo's positions tie in with Merleau-Ponty's development of Kant in *Phénoménologie de la Perception*, where he argues that we are not just "in" space, but are "of" it; that is, self-consciousness is the self-consciousness of a body, something inherently extended in space.

Others would dispute the idea that the body plays a role in Kant's theory of the productive imagination, in particular, its schematising activities, since he aligns inner sense with the sphere of the mental (Aquila 1992; Longuenesse 2006, Ginsberg, 2008; Melnick 2009; McDowell, 2009). It is useful here to elucidate Kant's distinction between "productive" and "reproductive" imagination. Productive imagination concerns the possibility of cognition *a priori*, whilst reproductive imagination, whose synthesis is subject to empirical laws, is concerned with psychology, and the laws of association. This involves the retention of earlier intuitions in such a way that certain other representations can "bring about a transition of the mind" to these earlier representations, even in the absence of any current representation of them (A100). The reproductive faculty of the imagination is "merely empirical" (A121). Productive imagination, on the other hand, implies a fundamental grounding of all cognitive capacities in imaginative acts. It does not concern itself with the connection of given intuitions, but rather, with the unified self that is necessary for any experience whatsoever. Kant regards the productive imagination *synthesis speciosa*, "figurative synthesis" or "transcendental function of the imagination" as the bridge between understanding and sensibility (A124). Through imagination, the principle of the synthetic unity of apperception is the principle of figurative as well as intellectual synthesis.

It is in the Deduction that Kant distinguishes between the two kinds of synthesis: "intellectual synthesis" (*synthesis intellectualis),* which is the manifold of an intuition in general, and "figurative synthesis" (*synthesis speciosa*) which is the synthesis of the manifold of sensible intuition, which is necessary *a priori* and depends on the faculty of sensibility. It is the product of the imagination, which, though spontaneous, itself "belongs to sensibility" (B151). Kant talks of figurative synthesis or *synthesis speciosa*, as "an action of the understanding on sensibility and is its first application" (B152). This synthesis is also called the *transcendental synthesis of the imagination* "in order to be distinguished from the merely intellectual combination" (B151). The transcendental use of the imagination resides

in its role in schematising - in producing a schema in order to subsume intuitions given by sensibility under a concept. Combination, however, proceeds prior to imagination and, it is only through the *synthesis speciosa* that "the categories, in themselves mere forms of thought, obtain objective reality…" (B150). Thus, figurative synthesis, *synthesis speciosa*, or transcendental or productive imagination accounts for the possibility of perceptual knowledge of spatio-temporal objects. Figurative synthesis is here presented as a condition for the objective reality of the categories, and, although an exercise of spontaneity in accordance with the synthetic unity of apperception, for the unity of this synthesis must presuppose it, it is a level of spontaneity which, as the "first application", does not yet properly consist in any concepts. Its function is only to "determine the sense *a priori* in respect of its form in accordance with the unity of apperception". That this synthesis is given prior to concepts is confirmed in a footnote at B160-1 which parenthetically directs the reader to §24, and to the outline of *synthesis speciosa*:

> Space, represented as object (as we are required to do in geometry), contains more than mere form of intuition; it also contains combination *[gathering-together [Zusammenfassung]* of the manifold, given according to the form of sensibility, in an intuitive representation, so that the form of intuition gives only a manifold, the formal intuition gives unity of representation. In the Aesthetic I have treated this unity as belonging merely to sensibility, simply in order to emphasise that it precedes any concept, although, as a matter of fact, it presupposes a synthesis which does not belong to the senses but through which all concepts of space and time first become possible. For since by its means (in that the understanding determines the sensibility) space and time are first given as intuitions, the unity of this a priori intuition belongs to space and time, and not to the concept of the understanding (B160-61n).

On this understanding of *synthesis speciosa*, the pre-intellectual syntheses of experience is subordinate to and aligned with the "intellectual" or formal synthesis of understanding, and ensures that conceptual form percolates all the way down. Most importantly, the definition given by Kant ("the faculty of representing in intuition an object that is not itself present" (B151), identifies the imagination's activity with sensibility's capacity to give pure intuitions of space and time as conditions for the possibility of *given* representations (Nuzzo, p. 28). In other words, imagination's spontaneity ultimately derives from Kant's transcendental doctrines on space and time as our forms of receptivity. On this view, we can think of Kant as implicitly holding the view that there is something like the "a priori of the human body", or

that the body is "the transcendental site of sensibility" which "displays a formal, ideal dimension essential to our experience as human beings" (Nuzzo, ibid., pp.8-9). Kant also calls productive imagination a "hidden *art* in the depths of the human soul" [my italics]. It is instructive that the term art is a technical term for Kant. In the *Critique of Judgement*, for example, he defines it as not a product, a beautiful painting, but as the activity of an agent engaged in the act of painting, and is distinguished from theoretical knowledge as the activity of skilled know how (*CPJ* 5:303). [82] This implies a pivotal role of the body in perceptual experience, that perception can take place only if there is imaginative and bodily activity prior to judgement.

In his pre-critical essay *Concerning the Ultimate Ground of the Differentiation of Directions in Space* (1768), Kant discusses the existence of left-and right-handed objects (which instantiate pairs of "incongruent counterparts", objects which are identical in their parts but still different in the sense that they cannot be enclosed by the same surface, such as is the case with a pair of gloves. Kant refers to the sense of orientation that allows us to intuitively distinguish our right hand from the left. He claims that this proved that space does not depend on relations between things, but on the points of view of subjects of perception, or rather on the relationship between perceiving subjects and objects. Kant stresses the fact that we intuitively relate things to the orientation of our body in order to argue that space can neither be perceived by the senses nor grasped by the intellect, but must depend on the transcendental ground for cognition. Kant's account of left/right orientation in this earlier work is relevant to his more or less implicit conception of embodiment in the *Critique*. For orientation is only possible where there is an embodied subject of perception, i.e. it is the fact of embodiment that makes it possible to experience objects in the first place, and it has been suggested that since embodiment can neither be reduced to the forms of intuition nor to a merely empirical fact, Kant must have conceived of the oriented human body as a "transcendental ground for our cognition" in its own right. Angelica Nuzzo's account of embodiment in Kant is that "the body is not the site of the empirical senses but the reference point of our formal sense for spatial orientation" (ibid., p. 36). Nuzzo's interpretation and in-depth analysis of all three of Kant's critiques is guided by what she terms Kant's "transcendental embodiment" which refers to the "pure, a priori dimension of our sensibility (cognitive, practical,

and aesthetic) - a dimension that is irreducible to purely mental activity and is necessarily embodied" (ibid., p. 7). On this reading, the problem of Cartesian mind-body dualism is dissolved since the pure *a priori* dimension of our cognitive, practical and aesthetic sensibility is irreducible to purely mental activities and is necessarily embodied. In other words, phenomenologically, the organism experiences itself as a unified whole acting and interacting in the world to achieve goals: there is an irreducible me-ness involved with the possession, ownership and agency of the bodily starting point for activity. This *zero-point orientation* is not simply bodily position or physiology, but *necessarily* involves the interactive element of sensorimotor activities that require both the tacit and explicit intentionality of a cognising agent.

From the above considerations it is suggested that for Kant the biological substrate of consciousness is the whole organism in its dynamic interaction with the environment; consciousness is inherently intentional, necessarily neurobiologically embodied and engaged with the natural world. He would not have been in agreement with the ideas of analytic/ functionalist cognitive science where the brain is taken in isolation from the non-neural body and environment, and which entails that consciousness could, in principle, be restricted to a brain in a vat rather than a body in a world.

As Stephen Palmquist remarks:

> Neuroscience will begin to dovetail nicely with philosophy when it recognizes that wisdom comes to us once we recognize that the brain itself *trusts the body* for virtually all of its functioning, just as Jung says the ego must trust the Self-archetype (Palmquist, 2013, p.3n).

.

> Kant's ultimate goal in philosophizing was to articulate an idea of human nature as a *unified* whole (ibid., p. 4).

7. Conclusion.

After the demise of the "anti-mentalistic" attitude in cognitive science during the first half of the twentieth century due to the rise of behaviourism, and the restoration of "the mind" to respectability in the second half through the "cognitive revolution", a functionalist-computational mind emerged, which was rendered scientific at the cost of removing from it its most fundamental characteristic: consciousness, or phenomenal, first-person experience. Following from this, when consciousness, or the "what-it-is-like-to-be" problem inevitably reappeared on the scientific scene, it was the "hard problem" of explaining this within a materialistic framework that caught the imagination of scientists and philosophers alike. The quest was on for what had become the holy grail of cognitive science, a materialistic, scientific understanding of phenomenal consciousness. David Chalmers was among the first to oppose such a reductionist scientific theory of mind and proposed a division of the problems of consciousness into the "easy" ones and the "hard" ones, the former being explicable in terms of functional/computational or neural mechanisms and the latter turning on the fact that consciousness and its attendant "qualia" resist any sort of functional definition. One reason for this is the physical causal closure thesis (Kim, 1993). No causal chain involving a physical event will ever cross the boundary of the physical into the nonphysical: If *x* is a physical event and *y* is a cause or effect of *x*, then *y*, too, must be a physical event. The physical causal closure thesis is primarily a metaphysical framework to which materialists and physicalists are committed and which permits no place for what is metaphysically independent of the physical world to have any causal effects. Thus it was that the study of the mind based on the functionalist or information processing model gave rise to the hard problem of consciousness, since it produces a dichotomy of "function" or brain activity, and phenomenal consciousness where conscious properties (*qualia*) are seen as nothing but epiphenomenal, caused by physical occurrences, but themselves causally redundant, having no physical effect on the material world. This led to a proliferation of attempts by scientists and philosophers to solve the deep problems and paradoxes of phenomenal consciousness, such as: What is the evolutionary advantage of consciousness? How can consciousness arise from mere matter? Could

there be a zombie, a creature that looked human in every way but without consciousness? How does one bridge the explanatory gap between brain activity and phenomenal consciousness? The principle tenet of a "science of consciousness", that there is "something it is like" to have an experience, that consciousness is *essentially* characterised by reference to there being "something it is like" to be in a certain mental state is what has given rise to the explanatory gap between mind and matter. This problem seems as intractable as ever, no credible solution has been found despite the vast amount of literature and studies on the subject. It has been suggested that this is because philosophers have attempted to resolve a paradox on its own terms, rather than question the presuppositions that make it unavoidable. The aim of this thesis has been to demonstrate the contemporary relevance of Kant's transcendental psychology in this regard; to show that it is an invaluable conceptual tool for confronting some of the puzzles in contemporary debate. It therefore rails against the orthodoxy of the dominant analytic school of philosophy, according to which Kant had little to say about the mind that was correct or useful, and also against analytic functionalists and "brain bound" theories. Admittedly, Kant's transcendental psychology, in several important respects, can readily be seen as an early form of functionalism or a *proto*-functionalism about the mind, as has been noted by several scholars. However, crucially, it also transcends this in several important ways, ways that can be made fruitful in addressing the puzzling features that arose from functionalism itself.

In order to bring this to the fore, the historical context within which Kant developed his ideas was examined; it was argued that adopting a historical perspective was particularly called for because the many conceptual problems that have plagued the quest for a scientific understanding of consciousness are the result of the lasting influence of this largely forgotten or neglected philosophical heritage. Cognitive science is based on the problematic rationalism of Descartes, where the mind is viewed as some*thing* inner (the brain/ mental processes) disconnected from other people and from the outer world. There is the Cartesian manner of thinking about consciousness as a stream of consciousness in the "theatre" of the mind. A successful science of consciousness must be able to give an account of the contents, the stream, the unity and the continuity. It is also based on the Cartesian/Humean notion that all we perceive are "ideas" (nowadays, representations) which mediate

between us and the external world. On this traditional picture of cognition, the mind just *is* a complex system constituted by mental representations, operating on or processing information, whether in abstract workings or functions of the mind or later, in cognitive neuroscience, in the "wetware" of the brain. It was shown that this model eventually gave rise to the hard problem of consciousness and resulted in a dichotomy of views between those that view the hard problem as a tractable question, a real problem answerable in broadly neuro-scientific terms and those who deny that there is anything answering to our conception of consciousness to the extent that it goes beyond structure and function. This had led to a kind of philosophical and methodological impasse where the arguments go back and forth without solution, all based on the idea that consciousness is kind of super-property added onto mechanistic physical properties of the brain.

A Kantian analysis, however, shows that the modern mind-body problem is the result of a kind of category error, where the *a priori*, transcendental character of subjectivity is confused with psychological "facts" amenable to scientific explanation. It shows that the nature of subjective experience and its phenomenal *qualities* is impossible to understand in terms of dualism of substances or properties and addresses the problem of trying to explain introspective first-person aspects of mental states (the mind) and consciousness in general in terms of third-person quantitative neuroscience (the brain/body). For Kant, there is a deep confusion with the idea of "mind" and "body" as conceptually separable. The nature of the question is misleading and leads us to believe that there is something tractable that can be substantiated in the realm of empirical knowledge. He held that certain mistaken beliefs about the mind and consciousness arise from reification of first person phenomenal experience, because certain philosophers (Descartes) have projected the particularity of private, first person experience onto a third person entity called "the mind" or "thinking substance". This is not to deny that consciousness involves self-awareness, it is to deny that self-awareness can be accounted for on analogy with our consciousness of extra-mental objects, i.e. in terms of a subject-object relationship. Although some contemporary philosophers do recognise and criticise the hidden assumptions underpinning much of cognitive science and neuroscience, it has gone largely unrecognised that the radical critique of this model or understanding of the mind began with Kant, whose powerfully innovative transcendental inquiry into the

necessary *a priori* conditions for cognition was not only a major philosophical breakthrough in his own time, but still surpasses contemporary anti-Cartesian/ quasi-Humean (e.g. Dennett's) efforts today.

Again, with the emergence of the embodied, enactive view of the mind as an alternative to traditional, representational theories of cognition, this idea was prefigured by him over two hundred years ago. On this view, instead of either trying to solve the hard problem of consciousness, (Cartesian) or describing consciousness as an illusion, (Humean) as is commonplace in reductionist, functionalist accounts, a newer understanding of a form of immanent teleological "presence" involving truly biological features is now on the horizon, inevitably intertwined with the self-establishment of a conscious identity which is at the same time the living process. Kant was committed to an active, sensorimotor view of consciousness, realising the spontaneity of rational agent through acts of "synthesis" or binding in an embodied biological system. His views on this, to be found in his later work, the *Critique of Judgement*, depict the human mind as metaphysically continuous with biological life, and display his recognition that rational human agents are necessarily also rational human living organisms, i.e. biological animals capable of intentionality whose rational mindedness and rational directedness towards objects in the world, other real persons, and themselves, is fully continuous with this. In his last philosophical writings, the *Opus Postumum*, Kant refers to the work undertaken in the *Critique of Pure Reason*. Without invalidating the *a priori* categories that had been the possibility of all knowledge, he finds an entirely new foundation for them, and this is the lived body. For Kant, mind is explanatorily and ontologically continuous with life, in the sense that whatever is metaphysically required for a human mind is also present in its biological life. The thesis of epigenesis in biology says that biological material is initially unformed and that form gradually emerges through the non-predetermined or relatively spontaneous operations of an innate endogenous organisational or processing device in interaction with its environment. It has been argued that Kant would endorse this view and in fact explicitly defends the theory that biological life is epigenetic. Importantly, he also extends this theory analogically to his theory of cognitive innateness (*CPJ* 5: 424; B167). Although Kant is often conceived as having offered little attention to the fact that we experience the world in and through our bodies, there is evidence that not only does

Kant, throughout his career and in works published before and after the Critiques, reflect constantly upon the fact that human life is embodied, but the *Critique of Pure Reason* itself may be read as a critical reflection aimed at exploring some significant philosophical implications of this fact. Although there can be little doubt that in the *Critique* reference to consciousness or thoughts is abstract and conceptual, in the *Opus Postumum* he suggests that nothing but the body can be the basis for the *a priori* principles postulated. Thought depends on the proper workings of the senses, and for sensory input to be ordered in such a way so as to make possible the formation of coherent thoughts, an agent must be dynamically coupled to the environment. Accordingly, if we are to understand cognition, it is more fruitful to think beyond the inner and outer distinction of mind and world found on traditional accounts and to consider the natural unity that already exists between an organism and the environment in which it is actively engaged. On this understanding, there is neither subjective supremacy over the objective nor is the subjective absorbed into the objective realm of traditional mechanistic science. Rather such science returns to the experiential realm from which the very dichotomy between subjectivity and objectivity arises, and then establishes *within it* a system of *mutual constraint.*

As Evan Thompson writes in the preface of *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*:

> Where there is life there is mind, and mind in its most complex forms belongs to life. Life and mind share a core set of formal or organizational properties, and the formal and organizational properties distinctive of mind are an enriched version of those fundamental to life. More precisely, the *self-organizing* features of mind are an enriched version of the self-organizing features of life (Thompson, 2007).

This is a fundamentally different perspective of ourselves, in stark contrast to the reductionist, deterministic, and mechanistic portrayal of human nature that is given in contemporary accounts, where a picture of humanity is emerging that threatens to undermine the sense of our own freedom and agency. This new vision of human cognition is potentially liberating, providing room for freedom, morality and reaffirming our understanding of what it is to be human. As was stated at the outset of this work, Kant is not merely a long-deceased philosopher, whose ideas have been superseded, but cognitive science's intellectual godfather and also fellow worker in

the field. He was a revolutionary, with novel and compelling ideas about the mind, human freedom, the place of mankind in nature, and of the irreducible but also non-dualistic mindedness of embodied creatures, whose mental properties are as basic in nature as biological properties, and metaphysically continuous with this. Kant was the first post-Newtonian philosopher to attempt to face up directly and fully to the philosophical problems of consciousness, and also of reconciling free will with universal natural determinism. Scientists and philosophers concerned with the problem of consciousness would benefit greatly from a proper understanding of his views - concerning the mind's synthesising powers, about its various mental unities (in particular, the unity of consciousness), and about consciousness of self.

Bibliography/References:

Aboitiz, F. 1996. Does bigger mean better? Evolutionary determinants of brain size and structure. Brain, Behavior and Evolution 47. pp. 225–245.

Aboitiz, F. 1990. Behavior, body types and the irreversibility of evolution. Acta Biotheoretica 38. pp. 91–101.

Aleksander, I. 2005. The World In My Mind, My Mind In The World: Key Mechanisms of Consciousness in Humans, Animals and Machines, Imprint Academic, UK.

Aleksander, I. & Dunmall, B. 2003. Axioms and Tests for the Presence of Minimal Consciousness in Agents, Journal of Consciousness Studies, 10 (4-5), pp.7–18

Allison, H. 1996. "On Naturalising Kant's Transcendental Psychology." In idem, *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy.*

Allison, H. 1983. *Kant's Transcendental Idealism*, New Haven: Yale University Press; second edition, 2004.

Alvarez, M. 2011. Glock on Analytic Philosophy and History, *Teorema*, 30 (1), pp. 95-102.

Anderson, Michael L. 2003. Embodied Cognition: A field guide, Artificial Intelligence, Volume 149, Issue 1, pp. 91-130

Ameriks, K. 2006. Kant and the Historical Turn: Philosophy As Critical Interpretation. Oxford: Oxford University Press

Anscombe, G.E.M. 1975. The First Person. In Samuel Gutenplan ed., *Mind and Language*, pp. 45-65 Oxford University Press.

Aquila, R. 1992. Personal Identity and Kant's 'Refutation of Idealism'. In *Immanuel Kant: Critical Assessments*. Eds. R. Chadwick/C. Cazeaux. London, pp. 143–167

Aquila, R., 1989. *Matter and Mind*, Bloomington: Indiana University Press.

Armstrong, D., 1968. *A Materialist Theory of the Mind*, London: Routledge; second edition, with new preface, 1993.

Arnauld, Antoine. Pierre Nicole. *Logic: The Art of Thinking*, translated by James Dickoff and Patricia James (Indianapolis: Bobbs Merrill, 1964)

Baars, B.J. 1997. In the Theatre of Consciousness, New York, Oxford University Press.

Baars, B.J. 1996. Understanding subjectivity: Global workspace theory and the resurrection of the observing self. Journal of Consciousness Studies 3: pp. 211-17.

Barsalou**,** L. W**.** 2008**.** Grounded Cognition. Annual Review of Psychology, Vol. 59

Barsalou, L.W. 2003. Situated simulation in the human conceptual system. *Lang. Cogn. Process*. 18:5. pp.13–62.

Barsalou, L.W. 1999 Language comprehension: archival memory or preparation for situated action? *Discourse Processes*, 28(1), pp. 61-80.

Bayne, Tim. 2010. *The Unity of Consciousness*, Oxford University Press.

Bayne, Tim. 2007. M. Marraffa, M. de Caro & F. Ferretti (eds.) *Cartographies of the Mind: Philosophy and Psychology in Intersection*. Dordrecht: Kluwer, 201-10

Beaney, M. (ed.), 2013. *The Oxford Handbook of the History of Analytic Philosophy*, Oxford University Press.

Bennett, M.R. & Hacker, P.M.S. 2003. *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Berkeley, G. 1710/1975. A Treatise Concerning the Principles of Human Knowledge in *Philosophical Works; Including the Works on Vision*. M. Ayers (ed.). London: Dent.

Bermúdez. J. 2015. 'Nonconceptual Mental Content', Stanford Encyclopedia of Philosophy, text online.

Bermúdez, J. 2003. *Thinking Without Words*. New York. Oxford University Press.

Bennett, J. 1974. *Kant's Dialectic*. Cambridge: Cambridge University Press.

Bennett, J. 1966. *Kant's Analytic*. Cambridge: Cambridge University Press.

Bird, G. 2003. Kant and Strawson's Descriptive metaphysics, in Hans-Johann Glock (ed.) *Kant and Strawson*. Oxford: Clarendon Press.

Block, N. 2005. Two neural correlates of consciousness. Trends Cogn. Sci. 9, pp.46–52

Block, N. 1995. On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18: pp. 227–47.

Block, N. 1980 "What is Functionalism?," in N. Block (ed.), *Readings in the Philosophy of Psychology*, 2 vols. (Cambridge: Harvard Univ. Press

Broad, C.D. 1925. *The Mind and its Place in Nature*. London, Kegan Paul, Trench, Trubner & Co.

Brook, A. 2008. Kant's view of the mind and consciousness of the self. Stanford encyclopedia of philosophy.

Brook, A. 2004. Kant, Cognitive Science, and Contemporary NeoKantianism. *Journal of Consciousness Studies* . 11, no. 10-11. pp. 1-25.

Brook, A. 1994. *Kant and the Mind.* Cambridge and New York: Cambridge University Press.

Buzsáki, G. & Moser, E.I., 2013. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience* 16:130-138.

Byrne, A. 2004. 'What phenomenal consciousness is like,' in R. Gennaro (ed.) 2004.

Byrne, A. 1997. 'Some like it HOT: consciousness and higher-order thoughts,' *Philosophical Studies,* 86: 103–129.

Carruthers, P. 2006. The Architecture of the Mind: massive modularity and the flexibility of thought. OUP.

Carruthers, P. 2000. *Phenomenal Consciousness*. Cambridge: Cambridge University Press.

Caves, C. M. Fuchs, C. A. & Schack, R, 2007. "Subjective Probability and Quantum Certainty," Stud. Hist. Phil. Mod. Phys. 38, p. 255

Castañeda, H. 1967. On the Logic of Self-Knowledge. *Noûs, 1*(1), pp. 9-21.

Carnap R. 2003, [1928] *Der Logische Aufbau der Welt*. Leipzig: Felix Meiner Verlag. English translation by Rolf A. George,1967 The Logical Structure of the World/ Pseudoproblems in Philosophy, University of California Press, Berkeley.

Chalmers, D. J. 2013. "Panpsychism and Panprotopsychism." *The Amherst Lecture in Philosophy* 8 : pp.1–35. <http://www.amherstlecture.org/chalmers2013/>.

Chalmers, D. J. 2011. A Computational Foundation for the Study of Cognition. *Journal of Cognitive Science* 12, pp. 325-359.

Chalmers, D. J. 2004. Epistemic Two-Dimensional semantics. *Philosophical Studies* 118: 153–226.

Chalmers, D.J. 2003. Consciousness and its Place in Nature in (S. Stich and F. Warfield, eds) *Blackwell Guide to the Philosophy of Mind* (Blackwell).

Chalmers, D.J. 2002a The Components of Content, in Chalmers, Philosophy of Mind: Classical and Contemporary Readings 608-33 Oxford: Oxford University Press

Chalmers, D. J. 2002b. On Sense and Intension. *Philosophical Perspectives* 16: pp. 135–82.

Chalmers D.J. 2000 *"What is a neural correlate of consciousness?",* in Metzinger, Thomas, Neural Correlates of Consciousness:Empirical and Conceptual Questions, MIT Press.

Chalmers, D.J. 1996. The Conscious Mind: In Search of a Fundamental Theory. Oxford University Press.

Chalmers, D.J. 1995a Facing up to the Problem of Consciousness, *Journal of Consciousness Studies* 2(3):200-19.

Chalmers, D.J. 1995b. The Puzzle of Conscious Experience, *Scientific American*, 273, pp. 80-6.

Chella, A. & Manzotti, R. 2007. Artificial Consciousness, Imprint Academic, UK

Chiel, H.J. & Beer, R.D. 1997. The brain has a body: Adaptive behaviour emerges from interactions of nervous system, body and environment. *Trends Neurosci*.20: 553-557.

Chisholm, R. 1981. *The First Person*, Brighton, Sussex: Harvester Press.

Churchland, P.M. 1988, *Matter and Consciousness*, Revised Edition. Cambridge, MA: MIT Press

Churchland, P. S. 1986. *Neurophilosophy*. Cambridge: MIT Press

Churchland, P.M. 1981. *Eliminative Materialism and the Propositional Attitudes*. Journal of Philosophy 78 pp. 67-9

Clark, A. 2016. *Surfing Uncertainty*. Oxford: Oxford University Press

Clark, A. 2011. Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press, New York.

Cotterill, R.M.J. 1998. Enchanted Looms: Conscious Networks in Brains and Computers. Cambridge University Press.

Cotterill, R. M. J. 1995 'On the unity of conscious experience', Journal of Consciousness Studies, Imprint Academic, Vol. 2, No. 4, pp. 290–311

Cottingham, J., Stoothoff, R., & Murdoch, D., 1992, *Descartes: Selected Philosophical* Cambridge, UK: Cambridge University Press.

Craik, Kenneth, J. W., 1943. *The Nature of Explanation*, Cambridge: Cambridge University Press.

Crick, F. 1994, *The Astonishing Hypothesis*: The scientific search for the soul. New York: Scribner.

Crick, F. & Koch, C. 1998. Consciousness and Neuroscience. In Cerebral Cortex, 8: pp. 97-107.

Crick, F. & Koch, C. 1995. Why neuroscience may be able to explain consciousness. Scientific American 273(6): pp.84-85.

Crick, F. & Koch, C. 1990. Towards a Neurobiological Theory of Consciousness. Seminars in the Neurosciences 2: pp. 263-275.

Damasio, A.R. 1994, *Descartes' Error: Emotion, Reason and the Human Brain*, NewYork: Putnam.

Davidson, Donald 1970. Mental Events. In L. Foster & J. W. Swanson (eds.), *Experience and Theory*. Humanities Press: 79-101.

Dawkins, R. 1976, 1989, 2006. *The Selfish Gene*. Oxford: Oxford University Press.

Degenaar, J. & O'Regan, J. K. 2015 *Sensorimotor theory and enactivism*. Topoi

Dehaene S, & Naccache, L, 2001. *Towards a cognitive neuroscience of consciousness:* Basic evidence and a workspace framework. Cognition 79:1–37

Demarest, B. (2017). Kant's epigenesis: specificity and developmental constraints. *History and Philosophy of the Life Sciences*, *39*(1), 3.

Dennett, D.C. 2015. Why and How Does Consciousness Seem the Way it Seems, open-mind.net.

Dennett, D.C. 2013. http://www.edge.org/conversation/the-normal-well-tempered-mind

Dennett, D.C. 2009. *The Part of Cognitive Science That Is Philosophy* in Topics in Cognitive Science 1 pp. 231–236 Centre for Cognitive Studies, Tufts University.

Dennett, D.C. 2009b. Intentional Systems Theory. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), The Oxford Handbook of Philosophy of Mind, OUP, Oxford.

Dennett, D.C. 2007. Philosophy as Naive Anthropology: Comments on Bennett and Hacker in *Neuroscience and Philosophy*, Columbia University Press.

Dennett, D.C. 2005. Sweet Dreams: Philosophical Obstacles to a Science of Consciousness MIT Press: Cambridge, Massachusetts.

Dennett, D. C. 2003. *Freedom Evolves.* New York, Viking Press.

Dennett, D.C. 2001. The fantasy of first-person science http://ase.tufts.edu /cogstud/ papers/chalmersdeb3dft.htm

Dennett, D.C. 1998. Brainchildren: Essays on Designing Minds. *Representation and Mind)* (MIT Press.

Dennett, D.C. 1996. Facing backwards on the problem of consciousness. Journal of Consciousness Studies 3: pp. 4-6.

Dennett, D. C. 1993. Back from the Drawing Board in Bo Dahlbom, ed., *Dennett and his Critics: Demystifying Mind*. Blackwell.

Dennett, D,C. 1992, The Self as a Center of Narrative Gravity in F. Kessel, P. Cole and D. Johnson, eds, *Self and Consciousness: Multiple Perspectives*, Hillsdale, NJ: Erlbaum.

Dennett D,C. & Kinsbourne M. 1992. *Time and the Observer.: The Where and When of Consciousness in the Brain.* Behavioural and Brain Sciences 15 pp 183-247.

Dennett. D.C. 1991 *Consciousness Explained.* Boston: Little, Brown.

Dennett D,C. 1986 *Brainstorms* Harvester Press, Brighton Sussex.

Descartes, R, 1984-1991. *The Philosophical Writings of Descartes*, trans. John Cottingham, Robert Stoothoff, Dugald Murdoch and Anthony Kenny, Cambridge: Cambridge University Press, 3 vols.1984-1991.

Di Paolo, E. A. 2009. Extended life. *Topoi*, *28*(1), pp. 9–21.

Di Paolo, E. A. 2005. Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), pp. 429–452.

Dretske. F. 1995. Naturalizing the Mind, MIT Press, Cambridge, Massachusetts.

Dreyfus, Hubert L. & Dreyfus, Stuart E. 1999. The challenge of Merleau-Ponty's phenomenology of embodiment for cognitive science. In Gail Weiss & Honi Fern Haber (eds.), Perspectives on Embodiment: The Intersections of Nature and Culture. Routledge. pp. 103-120.

Dreyfus, Hubert. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Dreyfus, Hubert. 1972. "*What Computers can't do – A Critique of Artificial Reason*". Harper & Row, New York, Evanston, San Francisco, London.

Edelman, G.M. 1992. *Bright Air, Brilliant Fire: On the Matter of the Mind.* Basic Books, New York.

Fodor, J. 1983. *The Modularity of Mind* Cambridge, Mass. MIT Press.

Fodor, J. 1981. *RePresentations: Philosophical Essays on the Foundations of Cognitive Science.* Cambridge, MA, MIT Press/Bradford Books.

Fodor, J. 1975. *The Language of Thought.* Cambridge, Mass: Harvard University Press.

Frankish, Keith. 2007. The Anti-Zombie Argument. *The Philosophical Quarterly* 57: pp. 650-666.

Frege, G. 1884. *Die Grundlagen der Arithmetik: eine logisch-mathematische Untersuchung über den Begriff der Zahl*, Breslau: w. Koebner; trans J. L. Austin, 1980, as *The Foundations of Arithmetic: A Logic-Mathematical Enquiry into the Concept of Number*, Oxford: Blackwell.

Förster, Eckart. 2000. *Kant's Final Synthesis*: An Essay on the Opus Postumum. Cambridge, Mass.; London, England: Harvard University Press.

Förster, Eckart. 1989. *Kant's Transcendental Deductions*: The Three Critiques and the Opus Postumum. Stanford: Stanford University Press.

Gibson, J.J. 1979. The Ecological Approach to Visual Perception. Haughton Mifflin. Boston.

Freidman, M. 2013. *Kant's Construction of Nature: A Reading of the* Metaphysical Foundations of Natural Science, Cambridge: Cambridge University Press.

Ginsborg, H. 2008. "Was Kant a Nonconceptualist?". *Philosophical Studies* 137, pp. 65–77.

Glenberg, A. M.1997. Mental models, space, and embodied cognition. Creative thought: An investigation of conceptual structures and processes, (pp. 495-522). Washington, DC, US: American Psychological Association, xv, 56.rg

Glock, Hans-Johann. 2008a. Analytic philosophy and history: A mismatch? *Mind* 117 (468): pp. 867-897.

Glock, Hans-Johann. 2008. *What is Analytic Philosophy?*, Cambridge: Cambridge University Press

Godfrey-Smith, P. 2001. Environmental complexity thesis and the evolution of cognition. In *The Evolution of Intelligence*, ed. R. Sternberg and J. Kaufman. Philadelphia: Taylor & Francis.

Goertzel, B., and Ikle, M. 2012, Mind Uploading (introduction to a special issue on this topic) International Journal of Machine Consciousness.

Guyer, P. 2008. *Knowledge, Reason, and Taste: Kant's Response to Hume*. Princeton NJ: Princeton University Press

Guyer, P. 1989. *Psychology and the Transcendental Deduction* in E Forster (ed) Kant's Transcendental Deduction pp. 46- 68, Stanford: Stanford University Press

Guyer, P. 1987. *The Cambridge Companion to Kant and Modern Philosophy*. (Cambridge Companions to Philosophy) Cambridge: Cambridge University Press

Guyer, P. 1987. *Kant and the Claims of Knowledge.* Cambridge and New York: Cambridge University Press

Guyer, P. 1980 *Kant on Apperception and A Priori Synthesis,* American Philosophical Quarterly 17, pp. 205-212.

Hacker, P.M.S. 2012. The Sad and Sorry History of Consciousness: being, among other things, a Challenge to the 'Consciousness-studies Community'. Royal Institute of Philosophy Supplement, 70, pp 149-168.

Haikonen, P.O. 2007. Robot Brains. Circuits and Systems for Conscious Machines. Wiley.

Haikonen, P. O. 2003. *The Cognitive Approach to Conscious Machines.* Imprint Academic. Exeter, UK.

Hanna, R. 2015. The Limits of Sense and Reason: An Analytic and Critical Commentary on Kant's Critique of Pure Reason (unpublished).

Hanna, R. 2014. Kant's anti-mechanism and Kantian anti-mechanism. Studies in History and Philosophy of Biological and Medical Sciences

Hanna, R. 2013. The Togetherness Principle, Kant's Conceptualism, and Kant's Non-Conceptualism", *Stanford Encyclopedia of Philosophy*.

Hanna, R. & Maiese, M. 2009. *Embodied Minds in Action* Oxford: Oxford Univ. Press

Hanna, R. 2008. *Kantian Non-Conceptualism*, "Philosophical Studies" (137)

Hanna, R. & Moore, A. W. 2007. Reason, freedom and Kant: An exchange. Kantian Review 12 (1): pp. 113-133.

Hanna, R. 2006a. *Rationality and Logic.* Cambridge Mass: MIT. Press

Hanna, R. 2006b. *Kant, Science, and Human Nature*. Oxford: Oxford University Press

Hanna, R. 2004. *Kant and the Foundations of Analytic Philosophy*. Oxford: Oxford University Press

Hanna, R & Thompson, E, 2003. "The Mind-Body-Body Problem," *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies* 7 pp. 23-42

Hardcastle, V.G. 1995. *Locating Consciousness*. Amsterdam & Philadelphia: John Benjamins Press.

Harnad, S. 2003. "Can a Machine Be Conscious? How?", Journal of Consciousness Studies,10. No. 45: pp. 67-75

Harnad, S. 1990. The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, 42: pp. 335-346

Hatfield, G. 1992. Empirical, rational, and transcendental psychology: Psychology as science and as philosophy. In Paul Guyer (ed.), Cambridge Companion to Kant, Cambridge University Press. pp. 200–227.

Held, R. & Hein, A. 1963. Movement-produced stimulation in the development of visually guided behaviour. *J. Comp.Physiol. Psychol.* 56: pp. 872–876.

Hohwy, J. 2009. The neural correlates of consciousness: New experimental approaches needed? *Consciousness and Cognition*, 18(2): pp. 428–438.

Hohwy, J. 2007. The Search for Neural Correlates of Consciousness. *Philosophy Compass* 2/3: pp. 461- 474.

Hohwy, J. & Bayne, T. 2015. The Neural Correlates of Consciousness: Causes, Confounds and Constituents. In *The Constitution of Phenomenal Consciousness: Toward a science and theory*. S. Miller (ed). Amsterdam: John Benjamins, pp.155-176.

Holland and Knight, 2006. The Anthropomimetic Principle, Department of Computer Science, University of Essex.

Hurley, S.L. 2010. *The Varieties of Externalism* In The Extended Mind: MIT Press.

Hurley, S.L. 1998. Consciousness in Acton. Harvard University Press. *Kant on Spontaneity and the Myth of the Given* , Proceedings of the Aristotelian Society pp. 137-164

Hume, D. 1737, *A Treatise of Human Nature* Ed, L.A. Selby Bigge, Oxford University Press. 1962.

Hume, D. 1748 *An Enquiry concerning Human Understanding*, edited by Tom L. Beauchamp, Oxford/New York: Oxford University Press, 1999.

Jackendoff, R. 2007, *Language, Consciousness, Culture: Essays on Mental Structure*. The MIT Press.

Jackendoff, R. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.

Jackson, F. 1986. "What Mary Didn't Know". *Journal of Philosophy,* 83: pp. 291-295.

Jackson, F, 1982. Epiphenomenal Qualia Author(s): *The Philosophical Quarterly*, Vol. 32, No. 127, Published by: Blackwell Publishing for The Philosophical Quarterly

Jamieson and Bekoff, 1992. Carruthers on Non-conscious Experience. *Analysis*, 52: pp. 23-28

Jonas, H. 1966. *The Phenomenon of Life: Toward a Philosophical Biology*. Northwestern University Press.

Jékely, G. 2011. Origin and early evolution of neural circuits for the control of ciliary locomotion. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1707): pp. 914–922.

Kannisto, T. 2013. Modality and Metaphysics in Kant In Stefano Bacin, Alfredo Ferrarin, Claudio La Rocca & Margit Ruffing (eds.), *Kant und die Philosophie in welburgerlicher Absicht. Akden des XI. Kant-Kongresses* 2010. Walter de Gruyter. pp. 633-646.

Kant, I. 1770/2004, Inaugural Dissertation, Concerning the Form and Principles of the Sensible and Intelligible World, Kessinger Publishing, LLC.

Kant, I. 1781/1787 *Critique of Pure Reason,* trans. Norman Kemp Smith, London, Macmillan; New York , St Martin's Press, 1963..

Kant, I.1790/ *Critique of the Power of Judgment*, trans. P. Guyer and E. Matthews, 2000 Cambridge: Cambridge University Press.

Kant, I. 1788, *Critique of Practical Reason*, trans. Werner S. Pluhar, intro. Stephen Engstrom, Hackett Publishing Company, 2002.

Kant, I. 1783 *Prolegomena to Any Future Metaphysics* (trans. P. Carus, rev. with intro. by James Ellington). Indianapolis, IN: Hackett Publishers, 1977 (Ak. IV).

Kant, I. 1786 *The Metaphysical Foundations of Natural Science* (trans. with intro. by James Ellington). Indianapolis, IN: Library of Liberal Arts, 1970. (Ak. IV).

Kant, I. 1768. *Concerning the ultimate ground of the differentiation of directions in space*, In: The Cambridge Edition to the Works of Immanuel Kant: Theoretical

Philosophy 1755-1770. ed. D. Walford, E. Meerbote, 2003, Cambridge: Cambridge University Press.

Kant, I. 1900-. Akademie der Wissenschaften (ed.), Kants gesammelte Schriften. Berlin: Georg Reimer (later Walter De Gruyter

Kant, I. 1798. *Anthropology from a Pragmatic Point of View*. Mary Gregor (ed. And trans.). The Hague: Martinus Nijhoff, 1974.

Kant, I. 1993. Kant's Opus postumum. In E. Förster & M. Rosen (Eds.), *Opus Postumum* (The Cambridge Edition of the Works of Immanuel Kant, pp. 1-2). Cambridge: Cambridge University Press.

Kaplan, D., 1989. "Demonstratives", in *Themes from Kaplan*, J. Almog, J. Perry, and H. Wettstein, eds., New York: Oxford university Press.

Kemp Smith, Norman. 1962. *A Commentary to Kant's "Critique of Pure Reason".* New York: Humanities Press.

Kim, J. 2006. *Philosophy of Mind* (2nd edn., Boulder, CO: Westiew Press.

Kim, J. 2005. *Physicalism, or Something Near Enough*, Princeton University Press

Kim, J. 1998. Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation. Cambridge, MA, MIT Press

Kim, J. 1997. "The Mind-Body Problem," in *The Oxford Companion to Philosophy* edited by Ted Honderich. pp 185-207.

Kim, J. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge University Press.

Kim, J. 1989. The myth of non-reductive materialism. _Proceedings and Addresses of the American Philosophical Association. 63 (3): pp. 31-47.

Kirk, Robert, 2015. "Zombies", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.)

Kitcher, P. 2006. *Kant's Philosophy of the Cognitive Mind.* In P Guyer ed., The Cambridge Companion to Kant, pp. 169-202. Cambridge: Cambridge University Press.

Kitcher, P. 1993. *Kant's Transcendental Psychology*. New York: Oxford University Press.

Koch, C. 2012. *Consciousness: Confessions of a Romantic Reductionist*, The MIT Press.

Koch, C. 2004. *The quest for consciousness: a neurobiological approach*. Englewood, U. S-CO: Roberts & Company Publishers.

Koch, C. 2001. Final Report of the Workshop, *Can a Machine be Conscious*, The Banbury Center, Cold Spring Harbor Laboratory, 13-16.

Kornblith, H. 2006. Appeals to intuition and the ambitions of epistemology. In Herrington, S. (ed), *Epistemology Futures*. Oxford:Oxford University Press.

Kosslyn, S.M. 2005 "Mental images and the brain." *Cognitive Neuropsychology* 22. pp. 3-4

Kripke, S. 1980 "Naming and Necessity", In *Semantics of Natural Language*, edited by D. Davidson and G. Harman. Dordrecht; Boston: Reidel.

Laakso, A. 2011. Embodiment and development in cognitive science. Cognition, Brain, Behaviour: An Interdisciplinary Journal (Special Issue: Embodiment and Development), 15, pp. 409- 425

Lakoff, George, 2008. The Functionalist's Dilemma [Review of Jackendoff's Language, Consciousness, Culture: Essays on Mental Structure, American Scientist, (Jan-Feb).

Lagerspetz, Olli. 2002. 'Experience and consciousness in the shadow of Descartes', Philosophical Psychology, 15: 1, pp. 5-18

Lamy, D., Salti, M., & Bar-Haim, Y. 2009 Neural correlates of subjective awareness and unconscious processing: an ERP study. J. Cogn. Neurosci. 21, pp. 1435–1446.

Landes, Donald A. 2015. "Between Sensibility and Understanding: Kant and Merleau-Ponty and the Critique of Reason". *The Journal of Speculative Philosophy* 29.3: pp. 335–345.

Libet, B. 2006. *"Reflections on the Interaction of the Mind and Brain"* Progress in Neurobiology. 78: pp. 322–326.

Libet, B. 1985. *"Unconscious cerebral initiative and the role of conscious will in voluntary action"*. The Behavioral and Brain Sciences. 8: pp. 529–566.

Leiter, B. 2001. On the Analytic/Continental Distinction, The Lieter Reports, Wiley-Blackwell.

Levine, J. 2011. Review of David Chalmers, *The Character of Consciousness*, Oxford University Press, 2010. Notre Dame Philosophical Reviews.

Levine, J. 2001. Purple Haze: The Puzzle of Consciousness, Oxford and New York: Oxford University Press.

Levine, J. 1983. Materialism and qualia: The explanatory gap. Pacific Philosophical Quarterly 64: pp. 354-61.

Leibniz, G. 1714/ 1988. The Principles of Nature and Grace and Monadology. In G.H.R. Parkinson ed., Leibniz' Philosophical Writings. Guernsey Press Co. Ltd.

Locke, John. 1690/1959 *An Essay Concerning Human Understanding, 2* vols. A. C. Fraser (ed.). New York: Dover.

Longuenesse, B 2006. Self-Consciousness and Consciousness of One's Own Body: Variations on a Kantian Theme. *Philosophical Topics* 34(1 f.) pp.283–309

Lyre, H. 2006. Structural Realism and Abductive-Transcendental Arguments. In: M. Bitbol et al. (eds.): "Constituting Objectivity: Transcendental Perspectives on Modern Physics". Springer, Berlin.

Mandik, P. & Weisberg, J., 2008. Type-Q Materialism**.** In Chase Wrenn, ed. *Naturalism, Reference, and Ontology: Essays in Honor of Roger F. Gibson*, New York: Peter Lang Publishing. pp. 223-246.

Marr, D. 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, MIT press.

Marr, D., Poggio, T. 1976. From Understanding Computation to Understanding Neural Circuitry, Artificial Intelligence Laboratory. A.I. Memo. *Massachusetts Institute of Technology*. AIM-357.

Matherne, S., 2016. Kantian Themes in Merleau-Ponty's Theory of Perception. *Archiv für Geschichte der Philosophie* 98 (2), pp. 193-230.

Maturana, H., & Varela. F. 1980. *Autopoiesis and cognition: The realization of the living*. Boston:Reidel.

McDowell, J. 2009. *Having the World in View*. Cambridge

McGinn, C. 1989. "Can We Solve the Mind–Body Problem?", *Mind*, 98, pp.349-366. Reprinted in McGinn, C. 1991, 1–22.

Melnick, A. 2009. *Kant's Theory of Self*. London. 2013. "Two Charges of Intellectualism Against Kant." *Kantian Review* 18(2), 197–219.

Meerbote, R. 1989. Kant's Functionalism." In: J. C. Smith, ed. *Historical Foundations of Cognitive Science*. Dordrecht, Holland: Reidel. pp 161-87

Mensch, J. 2013. Kant's Organicism: Epigenesis and the Development of Critical Philosophy, The University of Chicago Press.

Merleau-Ponty, M. 1962. Phenomenology of Perception, trans. by Colin Smith, London: Routledge & Kegan Paul; New York: The Humanities Press

Metzinger,T. 2012 http://www.beinghuman.org/article/interview-thomas-metzinger-what-self

Metzinger, T. 2009. *The Ego Tunnel: The science of the mind and the myth of the self*, New York, Basic Books.

Metzinger, T. 2005. Precis of Being No One. The Self-Model Theory of Subjectivity. Cambridge MA: MIT Press. In Psyche 11 (5).

Metzinger, T. 2003. Being No One. The Self-Model Theory of Subjectivity. Cambridge MA: MIT Press.

Metzinger, T. 2000. _Neural Correlates of Consciousness: Empirical and Conceptual Questions. MIT Press.

Metzinger, T. 1995. *Consciousness Experience*. Imprint Academic.

Minsky, M 1985. *The Society of Mind,* Simon and Schuster, New York

Minsky, M. 1980. K-lines: A theory of Memory. *Cognitive Science* 4, pp.117-133

Nagel, T. 2012 , *Mind and Cosmos.* Oxford: Oxford Univ. Press.

Nagel, T. 1974. What is it like to be a bat? *Philosophical Review* 83:435-50.

Noë, A. 2016 Sensations and situations: a sensorimotor integrationist approach. J Conscious Stud 23(5–6):66–79

Noë, A. 2010. *Out of our heads: Why you are not your brain, and other lessons from the biology of consciousness*. New York: Hill and Wang.

Noë, A. & Thompson, E., 2004. Are there neural correlates of consciousness? *Journal of Consciousness Studies*, **11**, No. 1, 2004, pp. 3-28.

Nuzzo, A. 2008. *Ideal Embodiment: Kant's Theory of Sensibility,* Bloomington and Indianapolis: Indiana University Press,

O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., & Maguire, E. A. 1998. Place cells, navigational accuracy, and the human hippocampus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), pp.1333–1340.

O'Keefe, J. & Nadel, L., 1978. *The Hippocampus as a Cognitive Map,* Oxford University Press.

O'Keefe, J. 1976. "Place units in the hippocampus of the freely moving rat". Experimental neurology. **51** (1): pp. 78–109.

Oizumi M, Albantakis L, & Tononi G. 2014. From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0. PLoS. Comput Biol 10(5).

Palmquist, S.R. 2013, Kant's Perspectival Solution to the Mind-Body Problem - Or, Why Eliminative Materialists Must Be Kantians. www. academia.edu

Papineau, D. 2003. Reply to Kirk and Melnyk. SWIF Philosophy of Mind_ 9.

Papineau, D. 2002. Thinking about Consciousness. Oxford: Clarendon Press.

Papineau, D. 1993. "Physicalism, consciousness, and the antipathetic fallacy." *Australasian Journal of Philosophy* 71, pp. 169-83.

Penfield W. 1958. Some Mechanisms of Consciousness Discovered During Electrical Stimulation of the Brain. *Proceedings of the National Academy of Sciences of the United States of America*. 44: pp. 51-66.

Penfield, W. & Jasper, H.H. 1954. *Epilepsy and the Functional Anatomy of the Human Brain*. Little, Brown.

Penrose, R. 1994. *Shadows of the Mind, An Approach to the Missing Science of Consciousness*. Oxford University Press.

Penrose, R. 1989. *The Emperor's New Mind, Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press.

Penrose R. & Hameroff, S., 2014. "Consciousness in the universe: A review of the 'Orch OR' theory," Physics of Life Reviews 11: pp. 39-78.

Penrose R. & Hameroff, S., 2011. *"Consciousness in the Universe: Neuroscience, Quantum Space-Time Geometry and Orch OR Theory"* Journal of Cosmology. 14.

Perry, J. 2009. "Directing Intentions," in Joseph Almog and Paolo Leonardi (eds.), *The Philosophy of David Kaplan*, pp. 187–207. Oxford: Oxford University Press.

Perry, J. 1979. "The Problem of the Essential Indexical", *Nous* 13 pp. 3-21.

Perry, J. 1975. "Introduction: The Problem of Personal Identity" in John Perry ed., *Personal Identity* pp. 3-30. Berkeley: University of California Press.

Piccinini, G. 2015, 'Access denied to zombies', Springer, Netherlands..

Pinker, S. 1997: How the Mind Works. New York: Norton.

Pinker, S. 2005. So How *Does* the Mind Work?' *Mind & Language*, 20/1.

Pippin, R. 2014. The Significance of Self-Consciousness in Idealist Theories of Logic, *Proceedings of the Aristotelian Society*, 114 (2 pt 2) pp. 145-166.

Pippin, R. 1987. "Kant on the spontaneity of mind." *Canadian Journal of Philosophy* 17, pp. 449-476.

Powell, C. T. 1990. *Kant's Theory of Self Consciousness,* Oxford University Press.

Putnam, H.1988 *The Meaning of Meaning*. *Representations and Reality*. Cambridge Mass.: MIT Press.

Putnam, H. 1975. Mind, Language and Reality: Philosophical Papers, Volume 2. New York: Cambridge University Press

Putnam, H. 1975. "The Mental Life of Some Machines", *Philosophical Papers*. 1st ed. Vol. 2. Cambridge: Cambridge University Press. pp. 408-428.

Putnam. H. 1967. The Mental Life of Some Machines, in: Castañeda, H-N. (ed.), *Intensionality, Minds and Perception*. Wayne State University Press.

Putnam, H. 1967. "The Nature of Mental States" (originally published as "Psychological Predicates"), in Captain, W. H. and Merrill, D. D. (eds.), *Art, Mind and Religion*, Pittsburgh: University of Pittsburgh Press, pp. 37- 48. Reprinted in Putnam, H. 1975: pp. 429-440.

Pylyshyn, Z. 1999. *What's in Your Mind* in What is Cognitive Science? Lepore, L. and Pylyshyn, Z. (eds) Blackwell.

Quine, W.V.O. 1969. Epistemology Naturalised. *In Ontological Relativity and Other Essays.* New York Columbia University Press.

Quine, W.V.O. 1951."Two Dogmas of Empiricism" *The Philosophical Review* 60: pp. 20–43. Reprinted in his 1953 *From a Logical Point of View*. Harvard University Press.

Reggia, J. A. 2013. The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks , 44*: pp. 112-131.

Revonsuo, A. & Newman, J. B. (1999). Binding and consciousness. *Consciousness and Cognition* 8 (2) pp. 123-127.

Rosenberg, J, F. 2005, Accessing Kant, Oxford University Press.

Rosenberg, J.F. 1986. The Thinking Self, Philadelphia, P A: Temple University Press, reissued by Ridgeview Publishing Co. 2008.

Rosenthal, D.M. 2005. *Consciousness and Mind*, Oxford: Clarendon Press

Rosenthal, D.M. 1997. "A Theory of Consciousness," in *The Nature of Consciousness: Philosophical Debates.* Editors Ned Block, Owen Flanagan, and Güven Güzuldere, pp. 729-54. Cambridge, MA: MIT Press.

Rosenthal, D.M. 1991. The Independence of Consciousness and Sensory Quality. *Philosophical Issues* 1, pp. 15-36.

Rukgaber, Matthew S. 2009. "The Key to Transcendental Philosophy' Space, Time and the Body in Kant." Kant-Studien 100: pp.166-186.

Rumelhart, D. E. & McClelland, J. L. 1986. PDP Research Group, Parallel distributed processing: Explorations in the microstructure of cognition: Cambridge, MA: Bradford Books/MIT Press.

Russell, B. 1927. The Analysis of Matter. London: Kegan Paul.

Russell, B. 1921, *The Analysis of Mind*, London: George Allen & Unwin.

Ryle, Gilbert. 1949. *The Concept of Mind*, London: Hutchinson.

Schiffer, S. 198. *Remnants of Meaning.* Cambridge, Mass.: MIT Press.

Schrödinger, E. and Penrose, R. 1992. *What is Life?: With Mind and Matter and Autobiographical Sketches*, Canto, Cambridge, Cambridge University Press.

Schubert, T.W., & Semin, G.R. 2009. Embodiment as a Unifying Perspective for Psychology. European Journal of Social Psychology, 39, pp.1135–1141

Schwoebel, J., Friedman, R., Duda, N., & Coslet, H. B. 2001. 'Pain and the body schema: Evidence for peripheral effects on mental representations of movement', Brain, 124 (10), pp. 2098–2104

Searle, John. 1993. The Problem of Consciousness in *Experimental and Theoretical Studies of Consciousness', Ciba Foundation Symposium no. 174*, (2008) Chichester: Wiley.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3, pp. 417-424.

Sellars, W.,1974. "… this I or he or it (the thing) which thinks …," *Essays in Philosophy and its History*. D.Reidel Publishing Company. Dordrecht. pp 62-90.

Sellars. W. 1963. *Science, Perception and Reality*, Routledge & Kegan Paul Ltd; London, and The Humanities Press: New York; reissued in 1991 by Ridgeview Publishing Co., Atascadero, CA.

Sellars, W. 1962 "Philosophy and the Scientific Image of Man", in *Frontiers of Science and Philosophy*, ed. by Robert Colodny (University of Pittsburgh Press; Pittsburgh, PA;

Sellars, W. 1956. *Empiricism and the Philosophy of Mind* edited by Robert Brandom, Harvard University Press.; Cambridge, MA; 1997.

Shallice, T. 1988. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Shapiro, L. 2011. *Embodied Cognition*. New York: Routledge

Shannon, Claude E. & Warren Weaver 1949: *A Mathematical Model of Communication.* Urbana, IL: University of Illinois Press.

Shannon, Claude E 1948: 'A Mathematical Theory of Communication', Part I, *Bell Systems Technical Journal*, 27, pp. 379-423

Shoemaker, S. 1996. The First-Person Perspective and Other Essays,. Cambridge: Cambridge University Press.

Shoemaker, S. 1990. First-person access. Philosophical Perspectives pp. 4:187-214.

Shoemaker, S. 1975. Functionalism and qualia. Philosophical Studies 27: pp. 291-315. Reprinted in *Identity, Cause, and Mind* (Cambridge University Press, 1984).

Shoemaker, S. 1968, "Self-Reference and Self-Awareness," in *Journal of Philosophy*, 65/19: pp. 555-67

Schulting, 2013. *Kant's Deduction and Apperception. Explaining the Categories*, Palgrave Macmillan, Basingstoke

Smart. J.J.C. 1959, Sensations and Brain Processes. *The Philosophical Review*, Vol. 68, No. 2. pp. 141-156.

Stapleton, M. 2013. Steps to a "Properly Embodied" cognitive science. Cognitive Systems Research, 22(2), 1-11.

Stapp, H. 2014. Mindful Universe: Quantum Mechanics and the Participating Observer. Springer.

Stapp, H. 2009. Mind, Matter, and Quantum Mechanics (The Frontiers Collection). Springer.

Sternberg, E. J. 2007. Are You a Machine? The Brain the Mind and What it Means to be Human, Amherst, NY: Prometheus Books

Strawson, P.F. 2003. A bit of Intellectual Autobiography', in H.-J. Glock, ed., *Strawson and Kant* Oxford: The Clarendon Press, pp. 1–14.

Strawson, P. F. 1966. *The Bounds of Sense.* London, Methuen Publishing.

Strawson, P. F 1959, *Individuals*, London, Methuen Publishing.

Stewart J., Gapenne E. & Di Paolo E. A. 2010. (eds.) Enaction: Toward a new paradigm for cognitive science. MIT Press, Cambridge MA

Stuart, S.A.J. 2010. Conscious machines: memory, melody and muscular imagination. *Phenomenology and the Cognitive Sciences*, 9(1), pp. 37-51.

Stuart, S.A.J. 2008. From agency to apperception: through kinaesthesia to cognition and creation. *Ethics and Information Technology*, 10(4): pp. 255-264.

Stuart, S.A.J. 2007. Machine consciousness: cognitive and kinaesthetic imagination. *Journal of Consciousness Studies, 14*(7): 141–153.

Stuart. S.A.J. 2007a. "Unifying Experience: Imagination and Self-Consciousness", book chapter in *The Mind, The Body and The World*, edited by Brendan Wallace & Alastair Ross, Imprint Academic, pp.116-31.

Taylor, Charles. 1984. Philosophy and its history. In: Richard Rorty et al. (eds.) *Philosophy in History*. pp.17-30. [Online]. Ideas in Context. (No. 1). Cambridge: Cambridge University Press.

Thagard, P. 2009, Why Cognitive Science Needs Philosophy and Vice Versa. Topics in Cognitive Science, 1: pp. 237–254.

Thagard, P. 1996 , "Cognitive Science", *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition).

Thompson, E., and F. Varela, 2001, "Radical Embodiment: Neural Dynamics and Consciousness," *Trends in Cognitive Sciences*, 5: pp. 418–425.

Thompson, E & Cosmelli, D. 2011: "Brain in a Vat or Body in a World: Brainbound versus Enactive Views of Experience," *Philosophical Topics* 39 : pp.163-180.

Thompson, E. 2007. *Mind in Life.* Cambridge: Harvard Univ. Press.

Thompson, E. 2005. Sensorimotor Subjectivity and the Enactive Approach to Experience. In Phenomenology and the Cognitive Sciences 4(4), pp. 407-427.

Thompson, Michael. 2012. Embodied Cognition: Kant's Conceptual Architecture, University of North Texas.

Tononi, G; Koch, C. 2015. Consciousness here, there and everywhere. *Philosophical Transactions of the Royal Society, London, Series B, Biological Sciences*.

Tononi, G. 2012. Integrated information theory of consciousness: an updated account. Arch Ital Biol 150: pp.56–90.

Tononi, G. 2008. Consciousness as integrated information: a provisional manifesto. Biol Bull 215: pp. 216–242.

Tononi, G. Koch C. 2008. The neural correlates of consciousness: an update. Ann N Y Acad Sci 1124: pp. 239–61.

Tononi, G. 2004. An information integration theory of consciousness. *BMC Neuroscience*, *5*, 42.

Torrance, S. 2008. Ethics and consciousness in artificial agents. *ArtificiaI Intelligence & Society22*: pp. 495–521.

Torrance, S., 2005, Thin Phenomenality and Machine Consciousness. In: Chrisley R, Clowes, R, Torrance, S. (Eds). Proceedings of the 2005 Symposium on Next Generation Approaches to Machine Consciousness.

Treisman, A. 2006. Object tokens, binding and visual memory. In H. Zimmer, A. Mecklinger, & U. Lindenberge (Eds.), Handbook of binding and memory: Perspectives from cognitive neuroscience. New York: Oxford University Press: pp. 315–338.

Treisman, A. 1986 (Nov) "Features and Objects in Visual Processing" *Scientific American 225.* pp. 114-125.

Treisman, A., and Gelade, G. 1980. "A feature-integration theory of attention." *Cognitive Psychology* 12, pp. 97-136.

Turing, A.M. Computing Machinery and Intelligence. Mind 49. pp. 433-460.

Tye, M. 2005. *Consciousness and Persons*. Cambridge,MA: MIT Press.

Van den Berg, H. (2014). Kant's Organicism: Epigenesis and the Development of Critical Philosophy, Jennifer Mensch. International Studies in the Philosophy of Science, 28(1), 99-101.

Varela, F.J., 1999. *Steps to a Science of Inter-being: Unfolding the Dharma Implicit in Modern Cognitive Science*, in Watson, Batchelor and Claxton, (eds.): pp.71-89.

Varela F. J. 1997. *A Science of Consciousness as if Experience Mattered*. In: Hameroff S., Kazniak A.& Scott A.(eds.) *Towards a Science of Consciousness*, MIT Press. pp. 31-44.

Varela, F., Thompson, E., & Rosch, E. 1991. *The embodied mind: Cognitive science and human experience*. Cambridge: MIT.

Varela, Maturana, & Uribe, 1974. Autopoiesis: The organization of living systems, its characterization and a model, Biosystems, Volume 5, Issue 4, pp. 187-196.

Van Inwagen, P. 1998, "Modal Epistemology", Philosophical Studies 92, pp.67-84,

Vicente, A. 2000. *"On the Causal Completeness of Physics"* . International Studies in the Philosophy of Science. 20: pp. 149–171.

Weber, A. & Varela, F.J. 2002, "Life after Kant: Natural purposes and the autopoietic foundations of biological individuality", *Phenomenology and the Cognitive Sciences*, 1, pp. 97-125.

Wilkerson, T.E. 1976. Kant's Critique of Human Reason: A Commentary for Students. Oxford University Press.

Williams, B. 2002, *Truth and Truthfulness: An Essay in Genealogy*, Princeton: Princeton University Press.

Williams, B. 2002, 'Why Philosophy Needs History', *London Review of Books* 17 October, pp. 7–9;

Wilshire, B. 2002. *Fashionable Nihilism: A Critique of Analytic Philosophy*, State University of New York Press, 2002.

Wittgenstein, L. 1964. *Blue and Brown Books.* Oxford: Basil Blackwell Publishers

Wittgenstein, L. 1953 Philosophical Investigations, Oxford

Wittgenstein, L. 1922, Tractatus Logico-Philosophicus, Routledge, 2001.

Wittgenstein, L. quoted in Hanna, P & Harrison, B 2004, Word and world : practice and the foundations of language. Cambridge; New York: Cambridge University Press.

Yablo, S. 1993, "Is Conceivability a Guide to Possibility?", *Philosophy and Phenomenological Research*, 53: pp. 1–42.

Zahavi, D. 2006. *Subjectivity and Selfhood: Investigating the First- Person Perspective.* Cambridge, MA: MIT Press

Zammito, J. 2007, *Kant's Persistent Ambivalence Toward Epigenesis, 1764–90*, in: Understanding Purpose. Kant and the Philosophy of Biology, ed. by Philippe Huneman, Rochester: University of Rochester Press,

Ziemke, T 2007a. *What's Life Got to do with It*: In Antonio Chella & Riccardo Manzotti (eds.), *Artificial Consciousness*. Imprint Academic: pp. 48-66

Ziemke, T. 2007b. The embodied self: Theories, hunches and robot models. *Journal of Consciousness Studies, 14*(7): pp. 167–179.

Ziemke, T. 2001. The construction of 'reality' in the robot: Constructivist perspectives on situated artificial intelligence and adaptive robotics. Foundations of Science. 6 (1-3): pp. 163-233.

Ziemke, T. 2016. The body of knowledge: On the role of the living body in grounding embodied cognition, Biosystems, Volume 148, Pages 4-11.

Zmigrod, Sharon; Hommel, Bernhard, 2013. Feature Integration across Multimodal Perception and Action in Multisensory research 26 (1-2): pp.143-57

Zmigrod, Sharon; Hommel, Bernhard, 2011. The relationship between feature binding and consciousness: Evidence from asynchronous multi-modal stimuli. *Consciousness and Cognition* 20. pp. 586–593

Zammito, J. 2003, "'This Inscrutable Principle of an Original Organization': Epigenesis and 'Looseness of Fit' in Kant's Philosophy of Science", *Studies in History and Philosophy of Science* 34, 73–109

Appendix 1. Notes

[1] Patrica Kitcher writes "Through his analysis of the prerequisites of cognition, Kant discovers a connection that links cognitive states. Even the most minimal cognitive task….demands a synthesis of states ... in further states, but acts of synthesis both create and presuppose relations among cognitive states. Synthesis creates a relation of dependence. The resulting state depends for its content, and so for its existence as a particular cognitive state, on the existence of earlier states". (Kitcher, 1990, p.117)

[2] This is in reference to a famous essay by Thomas Nagel called "What is it Like to be a Bat?"(1975), which has become almost synonymous with the formulation of "the hard problem of consciousness".

[3] Daniel Dennett regards what he terms "heterophenomenology", (the observation and recording of the mental lives of others as it is publicly expressed or manifested) as a way of bridging the subjective-objective divide. He writes "[Heterophenomenology] is the *neutral* path leading from objective physical science and its insistence on the third-person point of view, to a method of phenomenological description that can (in principle) do justice to the most private and ineffable subjective experiences, while never abandoning the methodological scruples of science (Dennett, 1991, p. 72). But his argument has many flaws.

[4] The clearest account of this distinction is in the *Anthropology* where Kant writes. "Inner sense is not pure apperception, consciousness of what we are doing; for this belongs to the power of thinking. It is, rather, consciousness of what we undergo as we are affected by the play of our own thoughts." (1798, Ak. vii, p. 161)

[5] Importantly, intuitions for Kant are not sense experiences; an intuition "is not as such as can itself give us the existence of objects....[i]t is derivative (intuitus derivativus) not original (intuitus originarius)" (B 72). He argues that our knowledge of objects is grounded in the transcendental unity of apperception: an object is "that in the concept of which a manifold of a given intuition is united" (B137). He describes the "manifold of a priori sensibility" as the "material for the concepts of pure understanding" without which the latter would be "without any content" (A77). This means that before they are synthesised, various given intuitions form a manifold, by being "received" by the a priori forms of sensibility (time and space). This suggests that empirical objects are "constructed" in the synthesis that unifies the manifold (see also A190/B235). The manifold of a priori sensibility is thus the material for synthesis, as opposed to the modes of transcendental unity, the categories, which are the form. Kant characterises an empirical intuition as a representation that is related to the object through sensation, where sensation is defined as a subjective representation that refers to our state insofar as we are "affected by objects" (A19–20/B34, A320/B377).

[6] These two concepts of consciousness, empirical apperception or 'inner sense' and 'transcendental apperception', generate two very different questions about the relation between consciousness and nature. On the one hand, there is the question of how mentality is related to physical nature; on the other hand, there is the question of how "spontaneity" is related to the whole of nature.

[7] The passage reads: "One of the reasons cognitive science is such a land of plenty for philosophers is that so many of its questions—not just the grand bird's-eye view questions but quite proximal, in-the-lab-now questions—are still ill thought out, prematurely precipitated into forms that deserve critical re-evaluation."

[8] In Strawson's own words "Descriptive metaphysics is content to describe the actual structure of our thought about the world, revisionary metaphysics is concerned to produce a better structure". However, the details of "the actual structure of the thoughts about the world" that Strawson wants to reveal through his descriptive metaphysics are cashed out in terms of his psycho semantic linguistics, the structure of which can be revealed by digging "beneath the surface of any natural language" (Strawson, 1959, p. 9).

[9] As Strawson writes: "A major part of the role of the Deduction will be to establish that experience necessarily involves knowledge of objects, in the weighty sense" (1966, p. 88).

[10] Strawson himself regarded *The Bounds of Sense* is "a somewhat ahistorical attempt to recruit Kant to the ranks of the analytical metaphysicians, while simultaneously discarding those metaphysical elements that refused any such absorption". He writes that his intention in writing it was "to preserve and present systematically what I took to be the major insights of Kant's work, while detaching them from those parts of the total doctrine that, if they had any substantial import at all, I took to be at best false, at worse mysterious to the point of being barely comprehensible" (Strawson, 2003, pp. 8-9)

[11] Hanna writes in *Kant and the Foundations of Analytic Philosophy*: "It has been forcefully argued by several leading contemporary philosophers that analytic philosophy has now reached a stage of crisis in its development. This crisis arises from the very unsettling fact that many and perhaps even most analytic philosophers now question the defensibility and ultimate intelligibility of the very idea of analysis. But how can there be analytic philosophy without a cogent and coherent conception of philosophical analysis? In this sense, the analytic consensus in contemporary philosophy—as intellectually vigorous, institutionally secure, and one might even say bull-marketish, as it undoubtedly is— is speeding towards a crash. Michael Friedman has very plausibly traced the origins of this crisis back to analytic philosophy's rejection of Kant, via its intimate but stormy relationship with logical positivism." (Hanna, 2004, p.11).

[12] In *Remnants of Meaning* (1987) Steven Shiffer presents a devastating criticism of analytic philosophy, and claims that the kinds of semantic linguistic projects at the heart of contemporary analytic philosophy are incoherent and impossible, thus undermining the whole enterprise.

[13] "It seems to many people that consciousness is a mystery, the most wonderful magic show imaginable, an unending series of special effects that defy explanation. I think that they are mistaken: consciousness is a physical, biological phenomenon—like metabolism or reproduction. . . ." (Dennett, 2005. p.57)

[14] Intuitions are said to be elicited in response to thought experiments or the description of possible case scenarios. Using intuitions about possible cases is central to the methodology of much contemporary analytic philosophy. The idea that intuitions can provide us with access to the truth is currently a hot topic in analytic philosophy.

[15] In his book *Fashionable Nihilism: A Critique of Analytic Philosophy*, Wilshire criticises the impersonal nature of analytic philosophy, and how it is overwhelmingly accepted by contemporary academia as being beyond reproach.

[16] Brian Leiter is co-editor of The Philosophical Gourmet Report (also known as the Leiter Report or PGR) which is a ranking of graduate programs in philosophy in the English-speaking world.

[17] An excellent resource is to be found in Hatfield's 1992 paper "Empirical, rational, and transcendental psychology", where he discusses the complexities of psychology as a science in Kant's view.

[18] Glock distinguishes stronger from weaker historicist claims, and concludes that analytic philosophy is no more subject to reasonable objections on the basis of historical considerations than any other systematic approach to philosophy. Weak historicism is when a study of the past is useful without being indispensible. However, as Bernard Williams, one of the most venerated practitioners of analytic philosophy avers, the study of philosophical history is not simply an aid to contemporary philosophical analysis, but essential to it since the genesis of certain concepts or beliefs is crucial to their content and validity (Williams, 2002. See also Alvarez, 2011).

[19] The causal closure thesis is the characteristic principle of physicalism or materialism and states that "[n]o physical event has a cause outside the physical domain" (Kim, 1993, p. 280). Also formulated as "all physical effects have only physical causes" (Vincente, 2006, p. 150.

[20] The Paralogisms chapters are concerned with the rationalist philosophers who "succumb to a powerful illusion grounded in the very nature of reason itself (A298/B354).

[21] He writes: "There is a minimally sufficient neural correlate for the content of consciousness at any given point in time. If all properties of this local neural correlate are fixed, the properties of subjective experience are fixed as well. Of course, the outside world could at the same time undergo considerable changes. For instance, a disembodied but appropriately stimulated brain in a vat could – *phenomenologically* -enjoy exactly the same kind of conscious experience you do right now while reading this book" (Metzinger, 2003, p. 547).

[22] Autopoietic systems, whether simple unicellular or more complex organisms act to further their existence, through the appropriate exchange of internal components with surroundings, and via the maintenance of boundary conditions.

[23] In *Kant's Organicism* Mensch argues persuasively that Kant's epistemological reflections should be understood against the background of eighteenth century biology, which significantly impacted his philosophical development.

[24] Wilfrid Sellars is an early, if not the earliest, contemporary functionalist in the philosophy of mind and described mental states as individuated by the inferential roles they play in thought, independently of their physical realization.

[25] It is later argued that the faculty of reason is inherently constrained by the particular contingent conditions of our human animal embodiment.

[26] The key to the difference between functionalist and neural models is that functionalism reduces consciousness to a role, whereas neural models identify consciousness with a physical or biological property that implements or realises that role in humans.

[27] An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient, under conditions C, for the corresponding state of consciousness. An NCC (for content) is a minimal neural representation system N such that representations of a content in N is sufficient, under conditions C, for representation of that content in consciousness (Chalmers, 2000, p. 31).

[28] To bring this out, Kant claims that the unity of thought might be the result of a collective unity of substances acting together just as "the [single] motion of a body is the composite motion of all its parts" (A353) and the continuity of thoughts might be based in a series of substances that pass their states from one to another, just as "[an] elastic ball which impinges on another similar ball in a straight line communicates to the latter its whole motion, and therefore its whole state" (A363n).

[29] In *Kant's Organicism*, Jennifer Mensch makes a case for how the concept of epigenesis, a radical theory of biological formation, lies at the heart of Kant's conception of reason, and that it was not simply a metaphor for Kant but centrally guided his critical philosophy (Mensch, 2013). It should be noted that this view is controversial and that there are objections to Mensch's view. Hein van den Berg (2014), for example, questions the validity of the organicist interpretation, since, according to her, it is unclear what the use of biological terminology would add to the understanding of Kant's deduction of the ideas of reason. Why would Kant take epigenesis, a theory which he often criticised, as a model for his transcendental philosophy? She draws attention to historians of science who emphatically disagreed with this view, since epigenesis would pose significant problems for Kant's philosophy, and argues that Kant was never fully comfortable with it. Epigenesis was a theory of generation giving expression to the fundamental eighteenth-century intuition of hylozoism, the idea of radical spontaneity in matter itself, which, it is claimed, Kant denies (Zammito, 2003, 2007). In defending herself from this charge, Mensch notes that Zammito recounted and re-evaluated this view in his reader's report on *Kant's Organicism* for the University of Chicago Press in the Autumn of 2011, stating "I can still cling to my view that Kant was never quite comfortable with epigenesis, but as a theory of nature, while I will concede with alacrity that he may well have been far more enamoured of it as a basis for metaphysics than I had conceived." Zammito (2003) had previously argued that, in spite of Kant's employment of the term epigenesist, his position contained many traits of the direct rival theory, preformationism. Mensch, however, is of the opinion that historians of science must uncouple Kant's writings on generic preformation from the use he makes of epigenesis with respect to reason, which were written during the so called "silent period", and argues that Kant's letters, lectures, notes, and the marginal notations he made alongside the textbooks he used for his classes should not be ignored in an attempt to understand his views, as they are a valuable source of what lies behind Kant's theoretical endeavours. She notes that they are frequently made use of by many other Kantian scholars: Wolfgang Carl, Paul Guyer, Beatrice Longuenesse, and Patricia Kitcher, to name a few. She also argues, persuasively, that van der Berg appears to ignore Chapter 7 which is the main focus for her position, where she describes Kant's account of "transcendental affinity" as the key to understanding the manner in which an epigenetic reason is ultimately necessary for the success of the Transcendental Deduction. Boris Demarest (2017) has also convincingly argued against Zammito, that Kant's version of epigenesis is more like classical epigenetics than preformationism, and gives a detailed argument as why it was that he clung to preformationist terminology.

[30] In *Naming and Necessity* Kripke argues for the non-identity of pain and C-fibre stimulation on the basis that each can be conceived of as existing without the other. Kripke's intuitions are, in this sense, Cartesian.

[31] Glock identifies Kant as the one who genuinely sets the table for the analytic conception of philosophy. In Kant we find *a priori* metaphysics, the centrality of epistemology, and the vision of philosophy as autonomous from the special sciences while remaining a cognitive discipline.

[32] Whereas there was a certain conceptual muddiness in the A edition version, in the B edition version of the Deduction, the two sides of the A Deduction, the objective side and the subjective side, are

integrated into a *single* line of argument.

[33] See Michael Dummett *Origins of Analytical Philosophy*, where he writes that the rejection of psychologism by Frege leads more or less inevitably to the linguistic turn and analytic philosophy (Dummett, 1993, p. 25).

[34] The problem with this narrow interpretation of Strawson's is that it overlooks Kant's view that analysis in the *Critiqu*e is just part of a much broader argumentative framework. As Graham Bird has written of a putative Kantian descriptive metaphysics: "Kant's descriptive metaphysics is a descriptive metaphysics of science, including psychology, and of ordinary experience" (Bird, G. 2003, p. 77).

[35] As Robert Pippin points out in a recent paper, for Kant logic had no content of its own, but was the "form" of the thought of any possible content. In addition, the unit of meaning, the truth-bearer, or judgment, was essentially apperceptive. Judging was implicitly the consciousness of judging. This was for Kant a logical truth (Pippin, 2014).

[36] Hanna claims the Critique is first and foremost about the nature of human *Erkenntnis* or "cognition" and "its anthropocentric, real metaphysics, transcendental idealism, it is *not* a treatise in what has come to be known in the neo-Kantian and Analytic traditions as *Erkenntnistheorie* or "epistemology," except as a secondary by-product." See Robert Hanna, The Limits of Sense and Reason: An Analytic and Critical Commentary on Kant's Critique of Pure Reason (2015, unpublished).

[37] Kant writes: There are only two ways in which we can account for a necessary agreement of experience with the concepts of its objects: either experience makes these concepts possible or these concepts make experience possible. The former supposition does not hold ... There remains, therefore, only the second supposition – a system, as it were, of the *epigenesis* of pure reason – namely, that the categories contain, on the side of the understanding, the grounds of the possibility of all experience in general (B 166-67).

[38] Notably, of the early analytic philosophers, Wittgenstein, in the *Tractatus* did *not* make this interpretive mistake, and even went to so far as to assert that *logic is transcendental*: 6.13 Logic is not a theory but a reflexion of the world. Logic is transcendental" (*TLP* 169).

[39] Rosenberg claims that transcendental logic "is a species of pure *specialized* logic" because it is "concerned with the most general principles of our thinking about objects experienced as in space and time" (2005, p. 90).

[40] In the *Principles of Philosophy* (1640) Descartes wrote: "By the word 'thought' ('*pensée*') I understand all that of which we are conscious as operating in us".

[41] It should be noted that Descartes does not consistently adhere to classical substance dualism. For instance, he writes:The soul is really joined to the whole body, and . . .we cannot properly say that it exists in one part of the body to the exclusion of the others. For the body is a unity which is in a sense indivisible because of the arrangement of its organs, these being so related to one another that the removal of any one of them renders the whole body defective. And the soul . . . is related solely to the whole assemblage of the body's organs. (Descartes, *Passions of the Soul*, 339, AT 351).

[42] For example, Peter Hacker writes: "[T]he larger part of the multitudinous philosophical writings of the consciousness studies community, and a considerable number of neuroscientific writings, rest on

fundamental conceptual confusions"… "The so called "hard problem of consciousness" and the plethora of related puzzles that arise from this mistaken understanding are simply conceptual confusions masquerading as empirical questions" (Hacker, 2012). Similarly, Paul Thagard claims that, contrary to commonly held views, philosophy is extraneous to cognitive science, it has a crucial role to play (Thagard. 2009).

[43] Kant writes: In every syllogism I first think a *rule* (the major premise) through the *understanding*. Secondly, I *subsume* something known under the condition of the rule by means of *judgment* (the minor premiss). Finally, what is thereby known I *determine* through the predicate of the rule, and so *a priori* through *reason* (the conclusion). The relation, therefore, which the major premiss, as the rule, represents between what is known and its condition is the ground of the different kinds of syllogism. Consequently, syllogisms, like judgments, are of three kinds, according to the different ways in which, in the understanding, they express the relation of what is known; they are either categorical, hypothetical, or disjunctive (A304/B360-1).

[44] It is useful here to elucidate Kant's distinction between "productive" and "reproductive" imagination. Productive imagination concerns the possibility of cognition a priori, whilst reproductive imagination, whose synthesis is subject to empirical laws, is concerned with psychology, and the laws of association.

[45] See Susan Stuart (2008) who claims that *imagination* in the Kantian framework can be interpreted as playing a significant role in the non-intellectual grasping, the pre-reflective self- awareness, and the orientation of our bodily selves in our world. Also Angelica Nuzzo's idea that imagination is prediscursive and embodied: i.e. it is "transcendental embodiment, which refers to the "pure, a priori dimension of our sensibility (cognitive, practical, and aesthetic) – a dimension that is irreducible to purely mental activity and is necessarily embodied" (Nuzzo, 2008. p.7). In addition, Michael Thompson's in his 2012 paper 'Embodied Cognition: Kant's Conceptual Architecture', puts forward the view that the table of logical judgements are drawn directly from the power of the imagination through a "formal" bodily component in space and time, i.e. they derive from the very structure of embodiment, but in a way that Kant can still insist upon the universality and necessity of the basic architecture of consciousness.

[46] Allison, for example, refers to the following passage in the *Critique*:Understanding is, to use general terms, the faculty of cognitions (*Erkenntnisse*). They consist (*bestehen*) in the determinate relation of given representations to an object: and an object is that in the concept of which the manifold of a given intuition is united. Now all unification of representations demands unity of consciousness in the synthesis of them. Consequently it is the unity of consciousness that alone constitutes the relation of representations to an object, and therefore their objective validity and the fact that they are cognitions (*Erkenntnisse*): and upon it therefore rests the very possibility of the understanding (B137) (trans. P. Guyer and A. Wood).

[47] As he writes in his book *Consciousness Explained* or what critics nickname "*Consciousness Explained Away*", "the self, according to my theory, is an abstraction defined by the myriads of attributes and interpretations (including self-attributions and self-interpretations) that have composed the biography of the living body whose Center of Narrative Gravity it is" (Dennett, 1991).

[48] As Dan Zahavi has noted: The growing disenchantment with higher-order theories made people look elsewhere for a viable alternative, and within the last couple of years quite a few have taken a closer look at Brentano (Zahavi, 2004, p. 71).

[49] It should be mentioned that in his later years Putnam had little patience for either computational functionalism or its underlying philosophical agenda.

[50] Putnam wrote:
Consider what the brain state theorist has to do to make good his claims. He has to specify a physical-chemical state such that any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain . . . etc. At the same time it must not be a possible state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extraterrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may be pain . . . Finally, the hypothesis becomes still more ambitious when we realize that the brain-state theorist is not just saying that pain is a brain state; he is, of course, concerned to maintain that every psychological state is a brain state. Thus if we can find even one psychological predicate which can clearly be applied to both a mammal and an octopus . . . but whose physical-chemical "correlate" is different in the two cases, the brain-state theory has collapsed (Putnam, 1967, pp. 436-7)

[51] Kant speaks of three different selves: the "phenomenal" or "empirical" self, the "noumenal" or completely unknown self, and the self of transcendental apperception (Bxxvii-xxix). However, commentators have noted that there is some ambivalence over his views as he also writes: "The I think expresses an indeterminate empirical intuition, i.e., perception. (...) [S]omething real that is given, given indeed to thought in general, and so not as appearance [phenomenon], nor as thing in itself *(noumenon)*, but as something which actually exists, and which in the proposition "I think " is denoted as such...(B422-23n). I think this ambiguity can be resolved by the thesis of "transcendental designation". Kant also says, in consciousness of self "nothing manifold is given" (B135) This is a reference to self whereby we "denote" ourselves without noting any quality of ourselves. Kant's focus here is on the logical or apperceptive actus of transcendental designation, of reference to oneself as "I". In this kind of reference to self we denote ourselves, purely "intellectually", the empirical representations supplying the material by which to do so.

[52] "Intrinsic qualities of qualia are not functionalizable and are therefore irreducible, and hence causally impotent. They stay outside the physical domain, but they make no causal difference and we won't miss them." Kim claims that "phenomenal mental properties are not functionally definable and hence functionally irreducible" (Kim, 2005, p. 29).

[53] Interestingly, in *Physicalism or Something Near Enough* Kim questions whether the unsolvability of both the problem of consciousness and the problem of mental causation mean that there is some *hidden flaw* somewhere in the functionalist system of concepts and assumptions? He writes "Some philosophers would be willing to take this as a sufficient ground for urging us to abandon our present system of concepts in favor of a cleansed and tidier one, claiming that the conundrum of mental causation and consciousness is reason enough for jettisoning our shared scheme of intentional and phenomenal idioms, with its alleged built-in "Cartesian" errors and confusions" (Kim, 2005 p. 8). He does not abandon functionalism, though, for him there are mental properties that can be captured by a functional definition, whereas qualitative aspects of some mental phenomena cannot be and are therefore irreducible and epiphenomenal, i.e. they lie outside the physical domain. Hence physicalism isn't the whole story.

[54] Descartes' early modal argument for substance dualism is in the *Sixth Meditation* (1641) where he notes that since he can "clearly and distinctly" conceive of himself existing apart from his body (and vice versa), and since the ability to clearly and distinctly conceive of things as existing apart

guarantees that they are in fact distinct, he is *in fact* distinct from his body.

[55] R. M. Hare is credited with being the first to use the term to characterise the dependence of moral properties on natural properties. In the 1970s Donald Davidson used it to describe the relation between the mental and the physical.

[56] He writes: When I use the notion of a rigid designator, I do not imply that the object referred to necessarily exists. All I mean is that in any possible world where the object in question *does* exist, in any situation where the object *would* exist, we use the designator in question to designate that object. In a situation where the object does not exist, then we should say that the designator has no referent and that the object in question so designated does not exist (Kripke 1971, p. 146).

[57] A great deal of scholarship has gone into trying to solve the problem of the Cartesian Circle, and there are probably as many solutions as there are commentators.

[58] What it is to be a logically possible world is to be a "conceptually possible" world, where "conceptual possibility" is defined as: "conceivable on ideal rational reflection," or "ideally conceivable" (1996, p. 35).

[59] "Modal sceptic" Peter van Ingawen regards confidence in intuitions as little more than the product of a philosophical culture that has become used to accepting such claims without question. He accepts we have much basic, everyday modal knowledge but denies we have the capacity to justify modal claims that are far removed from this basic knowledge, i.e. that are far out. (See Van Ingawen, 1998).

[60] See Toni Kannisto's review of Kant's analysis of modality which he says can be applied today to those analytic philosophers that have attempted to derive properly metaphysical results from mere logical analysis of modality.

[61] This kind of argument supports strong AI, which is that artificial silicon creatures with the right functional organisation could be in possession of a conscious mind. Strong A1 is a version of functionalism in which the computational state of a computer is exactly like a functional state in a brain. Mental states are information processing states of (a program implemented in) the brain.

[62] Arnauld offers a counterexample: "Suppose someone knows for certain that the angle in a semi-circle is a right angle, and hence that the triangle formed by this angle and the diameter of the circle is right angled. In spite of this, he may doubt, or not yet grasped as certain, that the square on the hypotenuse is equal to the squares on the other two sides; indeed he may even deny this if he is misled by some fallacy. But now, if he uses the same argument as that proposed by our illustrious author, he may appear to have confirmation of his false belief, as follows: 'I clearly and distinctly perceive', he may say, 'that the triangle is right-angled'; but I doubt that the square on the hypotenuse is equal to the squares on the other two sides; therefore it does not belong to the essence of the triangle that the square on its hypotenuse is equal to the squares on the other sides" (CSM II, pp. 141-142/AT VII, pp. 201-202).

[63] "The idea is that, when we engineer a complex system (or reverse engineer a biological system (like a person or a person's brain), we can make progress by breaking down the whole wonderful person into sub-persons of sorts agent-like systems [*sic*] that have *part* of the prowess of a person, and then these homunculi can be broken down further into still simpler, less person-like agents, and so forth—a *finite*, not infinite, regress that bottoms out when we reach agents so stupid that they can be replaced

by a machine". Daniel Dennett, "Philosophy as Naive Anthropology: Comments on Bennett and Hacker," in *NP*, p. 88.

[64] "All such a system would experience would be the presence of one unified world, homogeneous and frozen into an internal Now as it were. Neither a rich internal structure nor the complex texture of subjective time or the perspectivalness going along with a first-person point of view exists at this point. We could call this 'Selfless Snapshot Consciousness'"(Metzinger, 2005, p. 3).

[65] See Matthew S. Rukgaber *The Key to Transcendental Philosophy, Space, Time and the Body in Kant",* 2009, and Angelica Nuzzo's *Ideal Embodiment: Kant's Theory of Sensibility*, 2008, which gives a thorough exegetical account of embodiment throughout all three Critiques. Also Helge Svare's *Body and Practice in Kant*, 2006, where she argues that Kant's critical philosophy is a reflection on what it means to be embodied.

[66] Searle writes: "Whatever else intentionality is it is a biological phenomenon, and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena" (Minds, Brains, Programs, 1980).

[67] Even though nowadays artificial systems are able to acquire autonomously their own methods for grounding concepts and hence the symbols, it is nevertheless still a human being that supplies the symbol systems and the conceptualisations of the world for the system by drawing them from an existing human language.

[68] Kant maintains that space is a necessary *a priori* form of intuition which, though not derived from the world, gives certain knowledge of that world. It is the pure *a priori* form of "outer sense", i.e. that by which we represent objects "as outside us ", i.e. in space. This is illustrated by the fact that we have synthetic *a priori* knowledge in geometry, the "science that determines the properties of space" (B41) which we could not have arrived at by analysis of concepts alone, for "in a triangle two sides together are greater than the third, can never be derived from the general concepts of line and triangle, but only from intuition, and this indeed a priori, with apodeictic certainty (A25/B38). Likewise time is the pure *a priori* form of inner sense, our awareness of our inner states, as they "appear" to us as "inner determinations". "We cannot think a line without drawing it in thought, or a circle without describing it. We cannot represent the three dimensions of space save by setting three lines at right angles to one another from the same point. Even time itself we cannot represent, save in so far as we attend, in the drawing of a straight line...and in so doing attend to the succession of this determination in inner sense" (B154). At B150 Kant had distinguished between intellectual synthesis, or *synthesis intellectualis* and figurative synthesis or *synthesis speciosa*, i.e. transcendental acts of the productive imagination. Concrete determination in intuition requires figurative synthesisis, a pre-intellectual synthesis of experience, subordinate to and aligned with the intellectual or formal synthesis of understanding. This productive synthesis is the application of the understanding to sensibility to either inner or outer sense, as the above passage illustrates. Moreover, the acts of the productive imagination through which we construct geometric figures are acts of a motor subject, i.e. the motion of the subject is that which renders possible the determination of inner sense: "Motion, as an act of the subject (not as a determination of an object), and therefore the synthesis of the manifold in space, first produces the concept of succession - if we abstract from this manifold and attend solely to the act through which we determine the inner sense according to its form. The understanding does not, therefore, find in inner sense such a combination of the manifold, but produces it, in that it affects that sense" (B 155). These passages illustrate Kant's a priori framework and the interdependence within determination in the synthesis of outer sense (space) on the determination of inner sense (time).

[69] Theories of active perception also emphasise that a full and exact adaptation to the world requires action produced sensory feedback. Sensory stimuli are not experienced passively but gathered actively through interaction with the environment - action deployment structures the way humans develop sensory awareness. For example, Held and Hein (1963) showed that if developing kittens experience the world only passively, they develop suboptimal perceptual awareness.

[70] It is of interest that Jékely and his team discovered that when Platynereis larvae sense different physical environmental conditions such as changes in light, temperature or the chemical composition of the water they alter their movement accordingly through the triggering of neuronal signals that regulate ciliary movement. The researchers found that the nervous circuitries responsible for this are built in an unusually simple way: the sensory nerve cells also have motor function, which means that they send the motion signal directly to the ciliary band.

[71] That Merleau Ponty derived the ego-centric space idea from Kant is highlighted by this passage: "The idea of a single space and a single time, being grounded upon that of a summation of being, which is precisely what Kant subjected to criticism in the Transcendental Dialectic, needs in particular to be bracketed and to produce its genealogy from the starting point of our actual experience" (Merleau-Ponty, 1962. p. 256). There is ample evidence in that it was through the influence of Lachièze-Rey's interpretation of Kant that Merleau Ponty came to see the discussion of schematism as a resource to develop his own theory of the body schema (Merleau Ponty, pp. 403–8/443–8). One of the main features of Lachièze-Rey's interpretation was the claim that, for Kant, schematism, perception, and embodiment are intimately interconnected (See Matherne, 2016)

[72] Merleau-Ponty's Phenomenology of Perception can be viewed as a phenomenological rewriting of the *Critique of Pure Reason* from *within* the paradoxical structures of lived experience, effectively merging Kant's Transcendental *Aesthetic* and Transcendental *Analytic.* This is not surprising, given the neo-Kantian context of Merleau-Ponty's education (Husserl and Heidegger) and the fact that he aimed to establish the phenomenal ground between empiricism and intellectualism.

[73] Kant's theory of the exact sciences has quite recently been rediscovered and re-evaluated, and is enjoying a revival of interest in contemporary scholarship. See, e.g., Brittan, *Kant's Theory of Science*; Butts (ed.), *Kant's Philosophy of Physical Science*; Edwards, *Substance, Force, and the Possibility of Knowledge: On Kant's Philosophy of Material Nature*; Friedman, *Kant and the Exact Sciences*; Friedman, *Kant's Construction of Nature;* Plaass, *Kant's Theory of Natural Science*; Warren, *Reality and Impenetrability in Kant's Philosophy of Nature*; and Watkins (ed.), *Kant and the Sciences.*

[74] In fact, according to Hanna, Kant in the third *Critique* points out that "in order to explain the behaviours and natures of living organisms, including ... the behaviours and natures of rational human animals, we are theoretically obliged to posit the existence of causally efficacious emergent properties that naturally arise from self-organizing complex dynamical systems"(Hanna, 2007, p. 121).

[75] "Systems seem to be formed in the manner of lowly organisms, through a *generatio aequivoca* from the mere confluence of assembled concepts, at first imperfect, and only gradually attaining to completeness, although they one and all have had their schema, as the original germ, in the sheer self-development of reason" (A835/B863).

[76] Hanna makes the point that according to Kant we have teleological inner sense intuitions of our own biological lives. He writes: for Kant "biology adds the notion of what I will call *natural causal singularities*, and correspondingly the concept of the non-linear non-equilibrium thermodynamics (also known as ''complex systems dynamics'') of self-organizing systems to the familiar classical

notions of mechanistic natural causal regularities and the linear equilibrium dynamics of inertial physical systems" (Hanna, 2006, p. 437). After Varela, he understands the Kantian thesis as meaning that biological life is literally identical to non-conscious or conscious mind. So that "non-conceptual phenomenal affective-emotional consciousness in inner sense entails embodied biological life: conscious beings are necessarily also living organisms" (ibid, p. 435).

[77] Kant explicitly adopted epigenesis as a scientific theory of generation in the *Critique of Judgment*. However, he specifically followed Johann Freidrich Blumenbach's conception of epigenesis, with his separation of the realm of the living from the non-living, which lent Kant the tools he required to demarcate his metaphysics from theories of the natural world. Kant thought Blumenbach's theory provided an appropriate scientific explanation for generation, one which he considered as having a "great advantage (…) on experiential grounds" over preformation (Kant. 1790 *CPR* 5. p. 424). The attraction of this theory was that it did not require an appeal to divine or supernatural forces in nature, but emphasised the notion of immanent development, or in Kant's words, "self producing*" (selbst hervorbringend)* nature. It has been cogently argued that he adopted this specific model of epigenesist due to his disagreements with Herder regarding the independence of reason and nature; naturalisation and corresponding elimination of metaphysics, especially bothered him (see D. Helbig, D. Nasser, 2016, pp. 98-107). Also, as Mensch writes: "It was the unity of purposes within organic life, the fact that organisms could be both self-sustaining and vigilant regarding the need for repair, that made natural products amazing, not the mechanical operations themselves. For Kant it was thus the principle of life, the capacity for a being's generation and self-organization that needed explaining, and recourse to neither supernatural nor purely mechanical grounds of explanation could satisfy that need" (Mensch, 2013, p. 64)

[78] Dennett claims that human intentionality is also derived, because it is explained by the way it was designed, that is, by natural selection "which is just as real- but just less easily discerned because of the vast difference in time scale and size". His arguments are not convincing (Dennett, 1987, p. 318, see also Dennett, 1990, p. 62).

[79] Kant writes, for example: Appearances might very well be so constituted that the understanding should not find them to be in accordance with the conditions of its unity... But since intuition stands in no need whatsoever of the functions of thought, appearances would nonetheless present objects to our intuition (*A90*-91/B123).

[80] Conceptualists claim that Kant's slogan "Thoughts without content are empty, intuitions without concepts are blind" shows the necessary requirement of concepts for intuitions in such a way that sensible representations would lack representational content without the guidance of understanding. However, non-conceptualists claim that, in Kant's view, concepts are required only "*for the specific purpose of constituting objectively valid judgments*" (Hanna, 2013, p. 93).

[81] Susan Stuart writes in a footnote that although it is clear that for Kant thoughts are conceptual things "their underpinning is very definitely the proper working of the senses, and for sensory input that can be ordered and unified in such a way that makes possible the formation of a *posteriori* concepts, the agent will need to be dynamically coupled to her environment."

[82] *Art* as a skill [*Geschicklichkeit*] of human beings is also distinguished from *science* [*Wissenschaft*] *to be able* [*Können*] from *to know* [*Wissen*]), as a practical faculty is distinguished from a theoretical one, as technique [*Technik*] is distinguished from theory (*CPR 5:303).*