’The Cognitive View in Cognitive Science: An Investigation in the Context of Wittgenstein’s Later Philosophy’

A Thesis Submitted for the Degree of Doctor of Philosophy in the Ludwig-Maximilians-University Munich

Dr. Wolfram B. Schmitt MBA

Jesus College
University of Oxford

March 2006

Primary Examiner: Prof. Dr. Wilhelm Vossenkuhl

Secondary Examiner: PD Dr. Richard King

Examination Date: 24/07/2006

Dedication

To my Mum and Dad, and to my brother Klaus with love and gratitude.

Acknowledgements

I decided to write this thesis about 10 years ago, just after I had left school. By this time my fledgling interests in human nature had matured and developed into a serious interest in the human mind and philosophy. The journey I have started back then will undoubtedly continue, and the submission of this thesis is but a *"pit-stop"*. I could not possibly have come this far, however, had it not been for a great deal of support provided by others. First and foremost, I would like to thank my mum and dad, who continuously nurtured my curiosity and love of books, and who taught me the importance of a 'humane perspective'. Above all, I would like to thank them for their limitless love and support, which protected me through so many good times and some darker ones too.

On the academic side, I would first of all like to thank Prof. Dr. Wilhelm Vossenkuhl, for guiding my first steps in philosophy, for providing council and support during my time as a student of medicine and philosophy at Munich, when the university wouldn't let me sit my philosophical examinations, and for taking me on as a doctoral student whilst studying and working at Oxford. Few others would have had sympathy for my cause, and even less would have gone through such extraordinary lengths to help me succeed. Thank you so much.

At Oxford, I would like to thank Prof. Dr. PMS Hacker, who helped me discover and understand Wittgenstein's ideas, who has always guided, in one way or other, my interdisciplinary as well as my philosophical interests proper, and who saved me from error on numerous occasions.

Furthermore, I would like to thank Prof. Dr. TJ Horder, for bringing me to Oxford in the first place, and for guiding my interests and development at Oxford as mentor and friend throughout the years.

I would also like to express my deepest gratitude to Prof. Dr. Nick Rawlins and Dr. David Bannerman, who provided me with help and support whilst exploring the neuroscientific approach to mind, and who never wavered in their encouragement whilst being faced with my tantrums and onslaughts against their chosen field of endeavour. Thanks for always being there for me as mentors and friends.

I would also like to express my thanks and my love to some of my friends, who have turned my stint at Oxford (7 years and counting…) into the best time of my life. First, my co-residents and fellow dance aficionados at *Dance Central* (past and present) Jumin, Nick, Pete and Robin (as well as our regular and not so regular part-time residents Harriet, Jane, Sarah, Dave and Tim) for creating a home full of love and laughter,- the perfect refuge from the turmoil encountered both in the Ivory Tower and the *real world*. Next I would like to thank Emmy, for all those fabulous moments on the dance floor (as well as the not so fabulous ones, which due to her unique talent (read: sex, sizzle and style) were few in number), for her love, support and understanding, and for her patience in enduring my occasional tantrums and caprices. There is no one like you! Further thanks go out to Cécile who has been (rather unbeknownst) a continuous source of inspiration and happiness. There is not much to say other than this, - you're magic!

Finally, I would like to say thanks to my bulldogs Achilles, Velvet, Giacomo and Giacomino, for providing much needed distraction and a wet kiss when my spirits where flagging. Love you loads!

Contents

Chapter II. On the Significance and Implications of Wittgenstein's Later Philosophy - An Overview

Chapter III. Crypto-Cartesianism, Brain-Body Dualism and Related Confusions – Solving the Mind-Body Problem: A Prelude

Chapter IV. The Cognitive View of the Mind

Chapter V. Thinking and the Language of Thought

I. Works by Wittgenstein and their Abbreviations

a) In German

BB_D Das Blaue Buch. Eine Philosophische Betrachtung (Das Braune Buch) übersetzt von Petra von Morstein. In: *Werkausgabe , Bd. 5, Frankfurt am Main, Suhrkamp (1997)*

BPP_I Bemerkungen über die Philosophie der Psychologie (Band 1). In: *Werkausgabe , Bd. 7, Frankfurt am Main, Suhrkamp (1997)*

BPP_{II} Bemerkungen über die Philosophie der Psychologie (Band 2). In: *Werkausgabe , Bd. 7, Frankfurt am Main, Suhrkamp (1997)*

BT_D The Big Typescript – TS 213: German-English Scholar's Edition. *Aue M & Luckhardt G (eds.), Oxford, Blackwell Publications (2005)*

LPP_D Letzte Schriften über die Philosophie der Psychologie. In: *Werkausgabe , Bd. 7, Frankfurt am Main, Suhrkamp (1997)*

PB Philosophische Bemerkungen. In: *Werkausgabe , Bd. 2, Frankfurt am Main, Suhrkamp (1997)*

PG Philosophische Grammatik. In: *Werkausgabe , Bd. 4, Frankfurt am Main, Suhrkamp (1997)*

PU Philosophische Untersuchungen. In: *Werkausgabe , Bd. 8, Frankfurt am Main, Suhrkamp (1997)*

TB_I Tagebücher 1914-1916. In: *Werkausgabe , Bd. 1, Frankfurt am Main, Suhrkamp (1997)*

TB_{II} Ludwig Wittgenstein – Denkbewegungen, Tagebücher 1930-1932/1936-37 (MS 183). Ilse Somavilla (Hrsg.), Teil 1: Normalisierte Fassung. *Innsbruck, Haymon Verlag (1997)*

TLP_D Tractatus-Logico-Philosophicus. In: *Werkausgabe , Bd. 1, Frankfurt am Main, Suhrkamp (1997)*

ÜG Über Gewissheit. In: *Werkausgabe , Bd. 8, Frankfurt am Main, Suhrkamp (1997)*

Z_D Zettel. In: *Werkausgabe , Bd. 8, Frankfurt am Main, Suhrkamp (1997)*

b) In English

AWL Wittgenstein's Lectures, Cambridge 1932-35, from the Notes of Alice Ambrose and Margaret MacDonald. *ed. Alice Ambrose, Oxford, Blackwell (1979)*

BB_E The Blue and Brown Books. *Rush Rhees (ed.), Oxford, Blackwell (1958)*

BT_E The Big Typescript – TS 213: German-English Scholar's Edition. *Aue M & Luckhardt G (eds.), Oxford, Blackwell Publications (2005)*

CV Culture and Value. *G.H. von Wright (ed.), P. Winch (trans.), Oxford, Blackwell (1980)*

LPP Wittgenstein's Lectures on Philosophical Psychology 1946-47 (Note by Geach PR, Shah KJ, Jackson AC). Ed. Geach PT, Harvester Wheatshief, Hamel Hampstead (1988)

LW_I Last Writings on the Philosophy of Psychology, *vol. 1, 1982, G.H. von Wright and H. Nyman (eds.), trans. C.G. Luckhardt and M.A.E. Aue (trans.), Oxford: Blackwell.*

LW_{II} Last Writings on the Philosophy of Psychology, *vol. 2, 1992, G.H. von Wright and H. Nyman (eds.), trans. C.G. Luckhardt and M.A.E. Aue (trans.), Oxford: Blackwell.*

LWL Wittgenstein's Lectures in 1930-33. In G.E. Moore, Philosophical Papers, *(Allen and Unwin, Londo, 1959)*

LSD The Language of Sense Data and Private Experience (Notes taken by Rush Rhees of Wittgenstein's Lectures, 1936). *Philosophical Investgations 7 (1984), pp.1-45, 101-140*

NB Notebooks 1914-1916, *1961, G.H. von Wright and G.E.M. Anscombe (eds.), Oxford: Blackwell.*

OC On Certainty, 1969, G.E.M. Anscombe and G.H. von Wright (eds.), G.E.M. Anscombe and D. Paul (trans.), Oxford: Blackwell.

PG Philosophical Grammar*, 1974, R. Rhees (ed.), A. Kenny (trans.), Oxford: Blackwell.*

PI Philosophical Investigations (PI)*, 1953, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.*

PO Philosophical Occasions*, 1993, J. Klagge and A. Nordmann (eds.), Indianapolis: Hackett.*

PR Philosophical Remarks*,* 1964, R. Rhees (ed.), R. Hargreaves and R. White (trans.), Oxford: Blackwell.

RPP$_I$ Remarks on the Philosophy of Psychology, 1980, vol. 1, G.E.M. Anscomb and G.H. von Wright (eds.), G.E.M. Anscombe (trans.),

RPP$_{II}$ Remarks on the Philosophy of Psychology, 1980vol. 2, G.H. von Wright and H. Nyman (eds.), C.G. Luckhardt and M.A.E. Aue (trans.), Oxford: Blackwell.

PLP The Principles of Linguistic philosophy. *Waismann G; Ed. Harré R, London, Macmillan and St. Martin's Press,*

Z Zettel*, 1967, G.E.M. Anscombe and G.H. von Wright (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.*

Nachlass

MS, TS, MSS Collected Manuscripts of Ludwig Wittgenstein as edited on Facsimile CD Rom*, (1997), The Wittgenstein Archives at the University of Bergen (ed.), Oxford: Oxford University Press.*

II. Works by Russell and their Abbreviations

ML Mysticism and Logic. *London, Routledge* (1994)

OK Our Knowledge of the External World as a Field for Scientific Method in Philosophy. *London, Open Court* (1914)

PD My Philosophical Development. *London, G. Allen&Unwin Ltd.* (1969**)**

PP The Problems of Philosophy. *Oxford, Oxford University Press* (1989)

III. Abbreviations of Classical Philosophical Works

DA De Anima. *Aristotle; Lawson-Tancred H (Translator), London, Penguin Paperbacks (1987)*

LN Leviathan. *Thomas Hobbes (1651), London, Methuen (1951)*

E An Essay Concerning Human Understanding, John Locke (1690). *London , Dent&Sons* (1973)

THN *A Treatise on Human Nature, David Hume (1749). Selby-Bigge LA & Nidditch PH (2nd. Ed.), Oxford; Oxford University Press*

CrPR Critique of Pure Reason, Immanuel Kant (1781). *Pluhar W, Transl.; Cambridge, Hacket Publishing (1996)*

KrRV Kritik der Reinen Vernunft, Immanuel Kant (1781), *Felix Meiner Verlag, Hamburg (1998)*

Preface

I believe that there are only a select few topics, which arouse a similar level of interest and curiosity among academics and laymen alike, as does the study of mind and brain. Although mind and brain have been capturing the attention of philosophers for centuries, it is the *"scientific investigation"* of age old philosophical queries by so-called *cognitive scientists,* which is distinctive of the developments of the last few decades and which, in times to come, may well be considered the hallmark of the study of mind in the 20th and early 21st centuries. In the past, advances in the natural sciences underlay or boosted a plethora of developments in the technological, economic and political spheres that not only improved the standard of living and prolonged the average life span for a vast number of people, but also fuelled hopes that a new and improved understanding of *the nature of man* was also within reach. Despite all the benefits mankind derived from scientific and technological progress past and present, the success of the natural sciences also helped spread and foster a virulent and nowadays *quasi-ubiquitous* and unquestioned believe in the omnipotence of science and technology. In the context of the study of mind, in particular, it fostered the widespread (mis-)conception that in order for an investigation leading to insight and understanding a "scientific" approach is a *sine qua non*.

This thesis highlights the danger of such an approach. By investigating the framework of explanation adopted by cognitive scientists, the *cognitive view,* and by examining its inherent conceptions of mind and thought, it will be shown that the scientific, or rather *scientistic*, approach inherent in the views depicted above is highly questionable, and in the case of the study of mind and brain, does not further insight and understanding, but rather prevents it. In the 20th century, Ludwig Wittgenstein has been one of the few philosophers who recognized the fallaciousness of these ideas and who opposed the inherent scientism. His later philosophy, which provides the philosophical *mise en scène* for the following investigations, provides a much-needed antidote against the misconceptions common among cognitive scientists today. Although, many of the shortcomings of the ideas and views of cognitive scientists have, with a certain regularity, been discussed by philosophers working in the tradition of the later Wittgenstein (e.g. Kenny, Hanfling, Hacker, Hyman and Glock) their criticisms, more often than not, fell on deaf

ears. I believe that one of the main reasons underlying the imperviousness of the community of cognitive scientists to the criticism of the kind of analytic philosophy inspired by the later Wittgenstein, is that it mainly focused on select misconceptions inherent in the cognitive view, but regularly failed to point out their place in the overall framework of thought for cognitive scientists. As a result, it was easy to dismiss these criticisms with excuses of the sort *"but that does not bear any direct relevance to the work I do…"*.

For this reason, this thesis aimed to portray and examine the cognitive view in its entirety, i.e. to depict the intricate interconnections existing between the premises that provide the foundation of the cognitive view, and to point out their disastrous consequences for our understanding not only of a select view aspects of mental phenomena, but to our understanding of the mind (and consequently of human nature) in its entirety. In order to provide such a "big picture", and to describe the numerous often very subtle interconnections between the various ideas making up the cognitive view in all their breadth and depth, a lot of well-trodden ground had to be revisited and reviewed. Thus, Anthony Kenny's and Peter Hackers' discussions of the mereological fallacy (Chapter 3, Chapter 4 first half) will provide the "base camp" from which we will visit less familiar aspects of the cognitive view via novel routes on our way to the peak of this "philosophical mountain". Where required, my indebtedness to the works of these philosophers will be highlighted throughout the text.

Zusammenfassung

Mit dem Aufstieg der Kognitionswissenschaften in der zweiten Hälfte des 20. Jahrhunderts, traten die klassischen Fragen und Probleme bezüglich der Natur des menschlichen Geistes aus dem Schatten reiner philosophischer Betrachtung heraus und rückten ins Zentrum naturwissenschaftlicher Untersuchungen vor. Das breite Interesse welches diesen Fragen von Seiten verschiedener Psychologen, Linguisten, Computerwissenschaftlern, sowie Medizinern und Neurowissenschaftlern entgegengebracht wurde, führte schließlich zur Erschließung eines neuen interdisziplinären Wissenschaftszweiges, - den Kognitionswissenschaften. Die Kognitionswissenschaften bilden ein Amalgam verschiedener geisteswissenschaftlicher und naturwissenschaftlicher Disziplinen, die gemeinsam das Ziel verfolgen das Mysterium des menschlichen Geistes und geistiger Phänomene wie Denken, Sprache, Gedächtnis, Wahrnehmung etc., durch die Anwendung vermeintlich komplementärer und synergistischer Methoden zu enträtseln. Seit ihrer Geburtsstunde in den fünfziger Jahren, wurde diese neue *Wissenschaft des Geistes* als der Schlüssel zur Erklärung und zum Verständnis der Grundlagen intelligenter Handlung und intelligenten Verhaltens angesehen. Es ist allgemein Anerkannt dass die menschliche Fähigkeit sich in einer komplizierten und sich ständig wandelnden Umwelt zurecht zu finden auf die bemerkenswerte geistige Leistungsfähigkeit des Menschen zurückzuführen ist. Die Fragen denen die Kognitionswissenschaften nachgehen, sind somit aufs innigste mit unserem Selbstverständnis als menschliche Wesen verknüpft, und die Beantwortung dieser Fragen verspricht ein tieferes Verständnis der menschlichen Natur und des vernünftigen Handelns.

Das Ziel der vorliegenden Arbeit war es die Erklärungsgrundlage, d.h. das wissenschaftstheoretische Erklärungsmodel, der modernen Kognitionswissenschaften, die so genannte *cognitive view* , einer kritischen begrifflichen Analyse zu unterziehen. Die Grundlage dieser Analyse bildete dabei Ludwig Wittgensteins Spätphilosophie. Nach eingehender philosophischer Untersuchung der der cognitive view zugrunde liegenden Prämissen, und in Folge der Darstellung und Diskussion der mannigfaltigen begrifflichen Unzulänglichkeiten

und Missverständnisse, wurde diese als für die Untersuchung und das Verständnis des menschlichen Geistes und geistiger Phänomene als unzureichend und defizitär entlarvt. Das durch die *cognitive view* vermittelte Bild des menschlichen Geistes als einen biologischen Informationsprozessor, erweckt nur den Schein Aspekte des menschlichen Geistes verstehen und erklären zu können, was vor allem durch den im Rahmen dieses Bildes verwendeten modernen Technologiejargon erweckt wird. Der Grossteil der Kognitionswissenschaftler ist sich dieser Tatsache nicht bewusst, und arbeitet im Schatten einer gewaltigen Illusion. Die Bedeutung der vorliegenden Arbeit entspringt somit nicht zuletzt aus der ungeheueren Popularität und dem Respekt welche den Kognitionswissenschaften von Akademikern und Nichtakademikern gegenwärtig entgegengebracht werden. Die Ideen und Vorstellungen von bekannten Kognitionswissenschaftlern wie z.B. Daniel Dennett, David Chalmers, Antonio Damasio, Francis Crick, Gerhardt Roth oder Wolf Singer, üben einen enormen Einfluss auf das Verständnis des menschlichen Geistes und der menschlichen Natur der breiten Öffentlichkeit aus. Themen wie Bewusstsein, Denken, freier Wille, das Wesen des Ichs, oder auch die Rolle von Emotionen dominieren die Diskussionen vieler akademischer Fachzeitschriften. Über deren Stand und Ergebnisse wird jedoch mittlerweile regelmäßig in den Feullietons verschiedener Tageszeitungen und Nachrichtenmagazine berichtet wird. Es war ein Anliegen dieser Arbeit, durch das Aufzeigen der der Erklärungsgrundlage der modernen Kognitionswissenschaften zugrunde liegenden Missverständnisse, und der Rückbesinnung auf ein Aristotelisches Verständnis des Geistes als der Fähigkeit Intellektuelle Fähigkeiten zu erwerben, der Verbreitung verschiedener Neuromythologien entgegenzuwirken.

Im ersten Kapitel, wurde nach einer allgemeinen Erläuterung der Ziele und Absichten der vorliegenden Arbeit (s.oben), eine detaillierte Darstellung der fundamentalen Prämissen, welche dem Erklärungsmodell der Kognitionswissenschaften zugrunde liegen, erarbeitet. Besondere Aufmerksamkeit wurde dabei der kognitionswissenschaftliche Darstellung des menschlichen Geistes als einen biologischen Informationsprozessor, der Information empfängt, speichert, umwandelt und übertragt, geschenkt. Diese Darstellung hat Ihre historisch philosophischen Wurzel in den repräsentationalistischen Vorstellungen von Descartes und Locke, und kann als eine moderne Version dieser klassischen Ideen

betrachtet werden. Die begrifflichen Probleme und Unzulänglichkeiten der modernen *cognitive view*, welche in den darauf folgenden Kapiteln diskutiert worden sind, sind zum Einen als eine Manifestation des anhaltenden Einflusses dieser klassischen Ideen, zum Anderen als Folgeprobleme welche aus der mit diesen Ansichten verbundenen philosophischen Grundlagen entstehen, zu verstehen. Die Darstellung des Einflusses und der Bedeutung klassischer philosophischer Ideen auf die moderne kognitionswissenschaftliche Forschung diente somit einerseits der Erläuterung relevanter historischer Zusammenhänge, als auch der Identifizierung der Ziele der nachfolgenden philosophischen Untersuchungen. Zu diesen Zielen gehörten,

- die Vorstellung des menschlichen Geistes als einen biologischen Informationsprozessor, und das damit verbundene Bild des Geistes als eine nur dem Individuum durch Introspektion zugänglichen privaten Welt.
- die Vorstellung von Denken als einer Interaktion zwischen den vom Gehirn bearbeiteten mentalen (geistigen) Repräsentationen
- die Vorstellung von mentalen Repräsentationen, als Symbolen in *„Mentalesisch“*, einer dem Denken eigenen Sprache (sog. *language of thought hypothesis*), welche als *Medium* des Denkens dient,
- die Ansicht das *mentalesisch* die Grundlage von Intentionalität darstellt und dass durch die kausale Zusammen Wirkung mentaler Repräsentationen rationales Handeln erklärt werden kann,
- die Ansicht das die Struktur und Normativität geistiger und sprachlicher Inhalte durch mentale Repräsentationen erklärt werden kann.

In zweiten Kapitel, wurde eine Übersicht über die Entwicklung und den Inhalt der Wittgensteinschen Spätphilosophie, welche den methodologischen Rahmen und das philosophische Rüstzeug für die im Laufe dieser Arbeit angestellten philosophischen Untersuchungen stellte, verfasst. Besonderen Wert wurde dabei auf eine ausführliche Darstellung des revolutionären Charakters von Wittgensteins Auffassung vom Wesen philosophischer Probleme und dem Zweck der Philosophie gelegt. Viele Missverständnisse und Fehlinterpretationen von Wittgensteins

Spätphilosophie haben in der Vergangenheit an diesem Punkt ihren Anfang genommen. Die Philosophieauffassung des späten Wittgensteins, das Verständnis von philosophischen Problemen als sprachlicher, d.h. begrifflicher, Probleme, sowie Wittgensteins Einführung eines neuen Erfolgsmaßtabs in die Philosophie (*viz.* die Beseitigung und Auflösung begrifflicher Probleme, im Gegensatz zur Beantwortung philosophischer Probleme durch Theoriebildung nach dem Vorbild der Naturwissenschaften), wurden Anhand der für Wittgensteins Spätphilosophie zentralen Begriffe wie z.B. *Übersicht* und *Übersichtliche Darstellung, Therapie,* veranschaulicht. Darüber hinaus wurden sowohl die sprachlichen wie auch die nicht-sprachlichen (kulturellen und historischen) Quellen und Ursachen philosophischer Probleme erläutert. Außerdem wurde das Verhältnis zwischen Wissenschaft und Philosophie kritisch beleuchtet. Dabei wurde gegen die unter Kognitionswissenschaftlern (und vielen Philosophen) weit verbreitete Meinung argumentiert, dass die Fragen und Probleme bezüglich des menschlichen Geistes mit denen sich die Philosophie seit jeher beschäftigt hat, mit den Mitteln moderner naturwissenschaftlicher Technologien im Rahmen eines hypothetisch-deduktiven Verfahrens gelöst werden können. Diese Ansicht beruht jedoch auf einem tiefgehenden Missverständnis des Wesens der Wissenschaft und der Philosophie. Naturwissenschaftliche Methoden können nur zur Beantwortung empirischer Fragen erfolgreich herangezogen werden. Da jedoch philosophische Probleme und Fragen (nach dem Verständiss Wittgensteinscher Spätphilosophie) begrifflicher und nicht empirischer Natur sind, können diese von den Naturwissenschaften unmöglich beantwortet werden. Insofern sich die modernen Kognitionswissenschaften also mit Problemen und Fragen beschäftigen die auf Begrifflichen Missverständnissen beruhen, können diese mit deren Methoden nicht gelöst werden. Die gegenteilige Einschätzung kommt einer Selbsttäuschung gleich. Es ist vielmehr die Aufgabe der Philosophie solche im Kern sprachlichen Probleme durch begriffliche Untersuchungen aufzulösen. Naturwissenschaft und Philosophie stehen somit in einem komplementären Verhältnis zueinander. Das Fehlen dieser Einsicht ist ein grundlegender Fehler der modernen Kognitionswissenschaften, der einem Verstehen des menschlichen Geistes im Wege steht. Erfolgreiche naturwissenschaftliche, d.h. empirische, Arbeit setzt begriffliche Klarheit voraus. Diese Klarheit aber, fehlt den modernen Kognitionswissenschaften oftmals (s.unten).

Im dritten Kapitel, wurden zunächst die als allgemein gültig betrachteten Behauptungen und Erklärungsansätze die von Kognitionswissenschaftlern und Philosophen in jüngster Zeit aufgestellt wurden erläutert, und gezeigt dass vielen dieser Annahmen eine degenerierte Form des Cartesianismus zugrunde liegt. Das Hauptaugenmerk wurde dabei auf die weit verbreitete Zuschreibung psychologischer Attribute zum Gehirn oder zu Teile des Gehirns gerichtet. Durch die Untersuchung der *Grammatik* psychologischer Prädikate (bzw. Prädikatoren) wurde gezeigt dass solche Zuschreibungen eine Verletzung mereologischer Prinzipien mit sich führen. Dadurch wird üblicherweise Übersehen, dass das für die Erhellung geistiger Sachverhalte verwendete Explanans in Wirklichkeit nichts Erklärt, sondern nur den Schein einer Erklärung erweckt. Nach der Erläuterung dieser Problematik, wurden außerdem die Ursachen dieser Problematik und ihre Verbindung mit einer Vielzahl weit verbreiterter Vorstellung des Geistes und des geistigen, die allesamt auf verschiedenen sprachlichen Missverständnissen beruhen, untersucht. Unter diesen Vorstellungen war die analyse des Bildes des Geistes als ein privates nur dem individuellen Subjekt durch Introspektion exklusiv zugänglichem Reich, und die damit verbunden Behauptung epistemischer Privatheit die sich in Aussagen wie z.B. „Nur ich kann wissen dass ich Schmerz empfinde“ ausdrückt. Das mit diesen Ansichten verbundene Verständnis der Sprache wurde ebenfalls diskutiert. Im Laufe der folgenden Diskussion wurde aufgezeigt, dass solche Aussagen irrtümlich als epistemische Behauptungen missverstanden werden, tatsächlich aber *grammatische* Aussagen verkörpern. Somit können verschieden Individuen sehr wohl die gleiche Empfindung verspüren: Innere Zustände und Empfindungen bedürfen äußerer Kriterien.

Im Verlauf des vierten Kapitels wurde der Begriff des Geistes einer philosophischen Untersuchung unterzogen, und gezeigt das entgegen dem Verständnis des menschlichen Geistes als einer biologischen Entität, d.h. als einen Informationsprozessor der Informationen empfängt, speichert, umwandelt und überträgt, der Geist weder eine materielle noch geistige Entität darstellt. Als Alternative zu dieser Vorstellung, wurde für eine Rückbesinnung auf die Aristotelische Vorstellung des Geistes als der Fähigkeit zum Erwerb Intellektueller Fähigkeiten argumentiert. Da diese Position, den menschlichen Geist weder als materielle noch geistige Entität betrachtet, entgeht sie die mit dem klassischen Leib-

Seele Problem verbundenen Erklärungsschwierigkeiten. Im Anschluss an diese Diskussion wird die Vorstellung mentaler Repräsentationen einer genauen Untersuchung unterzogen. Dabei wird u.a. argumentiert, dass dadurch dass Symbole generell einem durch Regeln bestimmten Gebrauch unterliegen , die Vorstellung von mentalen Repräsentation als symbolischen Beschreibungen in einer dem Denken eigenen Sprache unsinnig ist, da Beschreibungen in diesem Sinne nicht im Gehirn vorfindbar sind. Neuronale Aktivitäten selbst können nicht als Symbole interpretiert werden, da Sie keinen durch Regeln bestimmten Gebrauch unterliegen. Um den Status eines Symbols zu erwerben, muss aber mit dem Zeichen die Möglichkeit dieses korrekt oder inkorrekt anwenden zu können verbunden sein. Von den Nervenzellen des Gehirns kann jedoch weder gesagt werden das sie wissen noch dass sie nicht wissen was ein bestimmtes Symbol bedeutet oder nach welchen Regeln es angewandt wird. Ähnliche Schwierigkeiten und Probleme wurden im Zusammenhang von „Hirnkarten“ diskutiert. Obwohl Sinnesreize sehr wohl mit neuronal Aktivität korreliert werden können, und somit eine quasi kartographische Ordnung der Funktion gewisser Hirnabschnitte möglich ist (siehe z.B., Motor-Homunculus), ist die unter Kognitionswissenschaftlern weit verbreitete Vorstellung dass solchen Karten der neuronalen „Kommunikation“ dienen unsinnig. Im Schlussteil des Kapitels, wird dann die Vorstellung der Speicherung mentaler Repräsentationen im Gehirn untersucht, eine ebenfalls Zentrale Vorstellung innerhalb der *cognitive view* und der modernen Neurowissenschaften. Diese abschließende Diskussion zeigte, dass diese Vorstellung nur Sinn machen würde, wenn eine solche gespeicherte Repräsentation einer dem Lesen mächtigen *Person* zugänglich wäre, die einer anderen Person z.B. erklären könnte was eine spezifische Repräsentation darstellt. Kognitionswissenschaftler und Philosophen haben bis dato aber keine Kriterien, im Sinne einer nicht-induktiven Evidenz für eine Behauptung, die den Sinn eines behaupteten Satzes festlegt, bestimmt, die Angeben was als solch eine Repräsentation zählt. Somit bleibt die Vorstellung einer mentalen Repräsentation Sinnleer.

Das fünfte Kapitel behandelte die der cognitive view inhärente Vorstellung des Denkens, *viz.* die Ansicht das Denken einem Zusammenwirken geistiger Repräsentation in mentalesisch entspricht. Zunächst wurde durch eine grammatische Untersuchung des Begriffes denken gezeigt, dass diese Vorstellung den

polymorphen Charakter des Denkens vernachlässigt. Zudem kommt in der kognitionswissenschaftlichen Vorstellung des Denkens auch die zwischen dem Denken und dem Ausdruck des Denkens bestehende innige Verknüpfung zu kurz, die außerdem eng mit den durch normative Bestimmungen zum Denken verwendeten Begriffe sowie der mit deren Anwendung und Zuschreibung verbundenen Urteilen, vernetzt ist. Im Anschluss an diese Diskussion wurde dann für Konzeption des Denkens argumentiert die im Einklang mit der Aristotelischen Vorstellung des Geistes steht. Denken, so wurde argumentiert, sollte als die Fähigkeit verstanden werden sich in ein eng verflochtenes Netzwerk von durch Regeln bestimmten Handlungen einzubinden. Denken besteht nicht in der Zusammenwirkung geistiger Repräsentationen, sondern kommt in den verschiedenen Formen sprachlichen und nicht-sprachlichen Verhaltens zum Ausdruck. Die enge Beziehung zwischen denken, dem Ausdruck des Denkens und durch Regeln bestimmten Verhaltens, bildet außerdem auch die Grundlage des strukturierten und allgemeinen Denkens.

Kapitel sechs untersuchte schließlich die der cognitive view innewohnende Erklärung von Intentionalität, welche sprachliche Inhalte zunächst auf geistige Inhalte zurückführt, und diese dann im Rahmen eines reduktionistischen Ansatzes zu erklären sucht. Zusätzlich zu diesem Erklärungsansatz wurde außerdem die damit verbundene Annahme diskutiert dass geistigen Zuständen eine kausal Rolle in der Genese von Handlungen zukommt, und die Plausibilität der kognitionswissenschaftlichen Erklärung der Normativität und Struktur geistiger und sprachlicher Inhalte untersucht. Im Laufe der Diskussion wurde gezeigt, dass die für den kognitionswissenschaftlichen Erklärungsansatz zentrale Unterscheidung zwischen intrinsischer und hergeleiteter Intentionalität inplausibel ist, da sie entweder zu einem *infiniten Regress* führt oder impliziert das man niemals die Bedeutung eines Zeichens vollständig verstehen könne. Als Kapitalfehler des kognitionswissenschaftlichen Denkansatzes wurde zudem die Vernachlässigung des normativen Aspekts von Sprache und Geist identifiziert. Kognitionswissenschaftler neigen dazu sprachliche Normen als metaphysische Behauptungen misszuverstehen, und fügen somit häufig banale Aussagen und Ausdrücke zu tiefen metaphysischen Bildern und Vorstellungen zusammen.

Die zahlreichen begrifflichen Misverständnisse die im Laufe dieser Arbeit aufgedeckt und diskutiert wurden, zeigen dass die cognitie view auf völlig unzureichenden, defizitären und oft sinnfreien Annahmen und Prämissen basiert. Zudem wird gezeigt dass die cognitive view nicht nur das Produkt verschiedener begrifflicher Missverständnisse und metaphysischer Spekulationen ist, sondern auch selbst eine Quelle philosphischer Misverständnisse darstellt, und dadurch einem unverzerrtem Verständnis des menschlichen Geistes und, wie z.B. die gegenwärtig geführte Diskussion um die vermeintliche Ilusion des freien Willens zeigt, der mensschlichen Natur im Wege steht. Im Gegensatz zur cognitve view, erweist sich die Aristotelische Vorstellung es Geistes als der Fähigkeit Intellektuelle Fähigkeiten zu erwerben nicht nur als die philosophisch unproblematischere Variante, sondern birgt auch das Potential die Grundlage für eine Kognitionswissenschaft zu bilden, die frei von den hier diskutierten begrifflicen Problemen ist.

Chapter I

Setting the Scene:
The Study of Mind, Representationalism and the Cognitive View – Main Currents, Basic Tenets and Fundamental Premises

1. The Contemporary Study of Mind and Thought: Origins and Influences

With the rise of the brain sciences in the second half of the 20th century, questions about the nature of the human mind and its defining features of language and thought transcended the domain of purely philosophical reflection, while attracting the attention and interest of psychologists, neuroscientists, linguists and computer scientists. This diversification of interest provided the crucial momentum for the creation of a new and interdisciplinary field of research, an amalgam that has come to be known as cognitive science. In cognitive science, philosophical reflection and empirical investigation have joined forces in order to unravel the mystery of mind and mental phenomena such as thought, language, memory, perception etc., while using complementary, and it is thought potentially synergistic methods of investigation. Since its inception in the early 1950s, it was hoped that this new *science of the mind* would provide an explanation and understanding of the mechanisms, which underlie intelligent behaviour and agency. It is generally acknowledged that the human ability to adapt to a complex and ever changing world is grounded in the remarkable mental capacities of human beings. The questions which cognitive science pursues are thus intimately tied to the very heart of our human self-conception, as the answers it strives to provide seem to hold the promise of a deeper understanding of human nature and rational agency.

Together with investigations in cosmology and genetics, the study of the (human) mind and mental phenomena has adopted a pivotal position within the program of twentieth century natural science. Historically, the integration of the study of the mind into the program of the natural sciences is partly due to the dissent that developed between the spirit and methods of the natural sciences and the

humanities or *Geisteswissenschaften,* whose humanistic spirit[1] they had hitherto been allied with[2]. The rifts, which ensued in the wake of this feud, had a profound impact not only on the study of the mind, but also on intellectual life and academia in western society as a whole. For the dissent between the natural sciences and the humanistic studies precipitated in the general erosion of humanistic values and the decline of a high culture as evident in the transformation of the conception of the value of education and its harnessing to the needs of a post-industrial society. This development culminated in the devaluation of the role of the humanities in education widely visible in modern societies. In philosophy, this trend precipitated in the promotion of the scientific method in philosophy, and its related spin-offs of logical positivism, naturalized epistemology and contemporary neurophilosophy[3]. It culminated in the establishment of a distinctly scientistic so-called *empirically informed philosophy of mind* as the main form of philosophical reflection informing the thinking of cognitive scientists.

One of the main reasons, which underlay the growing rifts between the natural sciences and the humanistic studies, was the immense success that the sciences enjoyed, particularly in the twentieth century[4]. This led to the nowadays wide spread believe that an understanding of humanity can only be arrived at by applying the

[1] The term 'humanism' is of 19th century origin and was initially used to refer to the spirit and work ethos of the Renaissance humanists. Beginning with Francesco Petrarca, the Renaissance humanists strived for a revival of classical learning (e.g. literature, art, history, philosophy and the law). The advocating and practicing of *studia humanitatis* (e.g. grammar, rethoric, poetry) by Petrarca and his followers, the *umanisti,* aimed for the development of the human excellences in all their forms (e.g. Porter & Teich 1991, p.26ff). In doing so, they created not only a mere intellectual movement but a movement that actively pursued the rebirth of a cultural ideal of life which was to transform the cultural self-consciousness of Europe. This cultural ideal did not only inspire a distinct *vita contemplativa* but also precipitated in the idea of a *vita activa* to whose success self-knowledge and self-understanding derived from the *studia humanitatis* was a *sine qua non*. For the umanisti, the studia humanitatis was at the heart of men´s endeavour to live the life of a morally autonomous being. Today, the precise conception of the humanities is somewhat ambivalent and sometimes extends even to the study of certain branches of economics. The following are generally acknowledged to form the core of the humanities: the classics (ancient Greek and Latin), Literature, jurisprudence, art and art history and criticism, music and musicology and, of course, philosophy.

[2] See: Hacker (2001), p.33f

[3] Neurophilosophy is a project within the analytic philosophy of mind. It has been initiated and promoted by the philosophers Paul and Patricia Churchland. The foundation of neurophilosophy is provided by eliminative materialism. The champions of neurophilosophy believe that a scientifically acceptable theory of ind is only possible if the mentalistic discours of folk-psychology is substituted by models and theories from cognitive scence and neurobiology. (see e.g.: Churchland 1986, 1992, 2002, Metzinger 1995, Pauen & Roth 2001)

[4] In a wider context, the advances in the natural sciences and subsequently advanced scientific technology provided the main momentum for the unleashing of powerful economic forces that furthered the rifts between the natural sciences and the humanities (Sen 2001, p.3f). (See also: Snow 1993)

causal hypothetico-deductive framework of explanation encountered in the natural sciences. Combined with their tremendous economic potential and the promise of financial gain the natural sciences thus continuously undermined the value of humanistic understanding - the distinctive forms of explanation and understanding characteristic of humanistic studies[5]. At the same time, the triumphant increase in popularity, which the natural sciences enjoyed, gave rise to *scientism* - the illicit extension of the methods and forms of explanation of the natural sciences. A prominent (early) example of scientism is the doctrine of the *Unity of Science*. This doctrine, which alleges the methodological homogeneity of scientific and humanistic understanding, was vigorously propounded by the logical positivists earlier this century, but is rooted in the positivism of the 19th century and in earlier post-Cartesian mechanism. At its most extreme scientism is *reductive* taking, for example, the form of *ontological reductionism* (i.e. *physicalism)*. As such it has become the defining credo of much of contemporary philosophy of psychology and cognitive science, affirming the identity of the psychological with the neural and informing the explanatory framework of cognitive science (see section 2 below). In its *non-reductive* form scientism is *methodological* and views the study of man as being tantamount to the study of nature. In accord with this view, it is believed that in order to be of any value, the logical structure of explanation in humanistic studies has to take the form of explanations of the natural sciences (e.g. in the form of a hypothetico-deductive causal explanation of human thought and action as attempted by cognitive science). On this view, a full understanding and explanation of the nature of human mind and thinking would require a knowledge of causes and the laws that describe and determine them. In an attempt to account for mental phenomena, *physicalism* in modern cognitive science, for instance, deems these causal laws to be abstract computational algorithms executed by the neural circuitry of the brain (see sections 2 and 3 below, for details).

In the course of this thesis the scientistic views underlying the accounts of the nature of the human mind and thinking provided by contemporary cognitive science will be exposed as deeply misguided and incoherent. At the heart of this endeavour lies the application and development of the philosophical ideas of the later

[5] Hacker (2001), p.36f.

Wittgenstein, who took an explicitly anti-scientistic view in his writings[6] (e.g. BB18) (see section 5 below for details). The relevance of the present enterprise arises out of the enormous popularity and respect cognitive science enjoys nowadays among academics and non-academics alike. Today, the ideas of cognitive scientists have an enormous influence on the public's understanding of the human mind and accordingly of human nature. Yet, the 'popularisation' of their often warped scientistic ideas does have a detrimental influence on disciplines and areas of human activity that implicitly acknowledge and depend upon an understanding of human mind and nature. Among the disciplines suffering from a distorted scientistic picture of human mind and nature are, for example, modern psychiatry and neurology as well as contemporary political and economic theory and the study of human rights and criminal law. By outlining the approach adopted in this study, delineating the views of cognitive scientists and identifying the targets for philosophical investigation, the present chapter constitutes an essential first step in achieving what is the major aim of this thesis, *viz.* to expose the reductionist[7] representationalism of modern cognitive science as flawed and misconceived. As a result of exposing representationalism and its inherent conception of mind and thought as pseudo-theories, it will be demonstrated that as a model for explaining the workings of the human mind the cognitive view is inadequate.

2. Strategy in Context

In the attempt to understand and explain the nature of the human mind and the possession of psychological attributes by human beings, cognitive science, like any other academic discipline adopts its own characteristic point of view. This point of view is manifest in the set of basic tenets, premises, assumptions and beliefs, which form the explanatory framework within which mental phenomena are studied and explained by cognitive scientists[8]. The explanatory framework adopted by cognitive scientists, *the cognitive view,* will be at the heart of the following discussions. The objective is to excavate and expose the basic tenets and premises underlying the

[6] See e.g.: Vossenkuhl (1995), p.15ff.
[7] Note: Reductionism is no target of this thesis *per se*.
[8] Although cognitive science includes (strictly speaking) philosophy as an autonomous discipline, and the term cognitive scientist denotes philosophers as well as neuroscientists, psychologists, linguists etc., philosophers will be referred to independently in what follows. The term will thus denote philosophers who would consider themselves cognitive scientists as well as those working in the philosophy of mind who would not consider themselves as members of the cognitive science community.

cognitive view in order to identify the targets, which are to be subjected to philosophical investigation in the subsequent chapters. In particular, the retracement of the origins of these fundamental assumptions and the reconstruction of their use in the creation of the causal mechanistic picture of the human mind and thinking propounded by contemporary cognitive science, will benefit the examination of the *bounds of sense* (see Chapter II for details).

Fundamental to this philosophical undertaking is the method of *connective analysis,* which is the term Strawson chose to denote the kind of philosophical investigation in line with the ideas and spirit of the later Wittgenstein (see Chapter II for details). Connective analysis thus is,

> `…the description, for purposes of philosophical elucidation, of the interconnectedness of related concepts, of their implications, compatibilities, and incompatibilities, of the conditions and circumstances of use of philosophically problematic expressions. Such analysis does not terminate in logically independent elementary propositions, or in simple, unanalyzable names or concepts. It terminates in the clarity that is obtained with respect to a given question when the network of concepts has been traced through all its relevant reticulations.´[9]

Connective analysis emphasises the elucidation of conceptual connections and their ramifications (i.e. of the concepts which are constitutive of the basic assumptions underlying the explanatory framework of cognitive science) in order to delineate the bounds of sense. It thus allows for the excavation and exposition of the conceptual confusions embedded in the cognitive view. Once these confusions are exposed and their sources identified it will be possible to disentangle the knots in the thinking of cognitive scientists. In order for a connective analysis to be successful, it is essential to describe the basic premises and fundamental tenets under scrutiny as clearly and accurately as possible. Only in doing so is it possible to make somebody see the error(s) in his thinking:

> `We can bring someone's mistake home to him only when he acknowledges it as the right expression for what he feels…The point is: only when he acknowledges it as such is it the right expression *(MS 213,410)´*.

[9] Strawson (1991), Ch.II

The task one is up to is comparable to weaning an alcoholic or drug addict of his addiction. In such cases it is often the biggest problem to get someone to acknowledge their addiction in the first place. More often than not the mere suggestion of the existence of an alcohol or drug problem will be denied and fiercely resisted. Yet, it is common knowledge that the first step for the treatment of an addiction is to get the addict to acknowledge that they in fact *have* a problem. It is the same with cognitive scientists and philosophers. Because cognitive scientists and many contemporary philosophers suffer from a lack of training in attending critically to the use of concepts, combined with a general adherence to the misguided view that philosophical problems can be answered by constructing hypothesis and theories on the model of modern science, they will have to be convinced of the existence of a confusion in their thinking in the first place, before a step towards dispelling the confusion can be undertaken. Consequently, in order to convince a cognitive scientist or philosopher of the existence of a particular confusion he must be prepared to say of the description of his views: ´Yes, that is what I think. That is exactly what I think.´

Yet, doing justice to the views of cognitive scientists is not the only motive for making the exposition of the basic assumptions underlying the cognitive view the starting point of the present investigation. Another reason is the fact that it is characteristic of philosophical confusions that one is lead astray by them right from the very beginning of an enquiry. The confusions, it has been pointed out,

> 'are liable to run through an elaborate and sophisticated empirical theory precisely because these conceptual incoherencies are present in the very form of the questions the theory addresses.'[10]

I suggest that the modern representationalist stance embodied by the cognitive view is a misguided attempt to provide an answer to questions concerning the nature of human mind and thinking and more generally to problems concerning the possession of psychological attributes by human beings *per se*. The attempt is misguided because it relies on a number of misconceived philosophical doctrines and pseudo-theories while it is believed that their deficits (if they are recognized as such at all) can be overcome by the powers of modern scientific technology. However, no scientific technology, regardless of how powerful it may be, can be of any use in the

[10] Hacker PMS (1991b), p.122

attempt to answer questions or problems, which are based on conceptual confusion. Equally, our understanding of the human mind will not be advanced by answers, which rest on conceptual confusions and transgress the bounds of sense (see Chapter 2).

Finally, in order to do justice to these methodological considerations the basic assumptions underlying the cognitive view will be put into their historic context. Adopting a historical perspective is particularly called for as many of the misconceptions the current investigation will deal with manifest the lasting influence of a largely forgotten or neglected philosophical heritage. Both scientific and philosophical arguments rely on multiple layers of hidden assumptions. And both scientists and philosophers are often unaware of the subliminal ways in which these hidden assumptions guide their thinking, direct their explanations and ultimately mould their understanding. Yet, it is crucial to understand the subliminal influence of such hidden assumptions as

> `… in the active up-to-the-minute concerns of contemporary science some of the most basic issues are in a sense, not matters of deliberate, "rational" consideration, but are better understood as the residual effects of largely forgotten historical backgrounds [11]'.

In the case of cognitive science the problematic historical background is of philosophical nature and largely comprised by the representationalist ideas of 17th and 18th century philosophers like Descartes and Locke, which have been uncritically adopted by the initiators of the cognitive revolution in the mid-20th century and which have been combined with ideas of modern computer technology (see below).

3. Representationalism and the Cognitive View – A Brief Historical Sketch

Representationalism is the view that without a medium of internal representations intelligent systems could not act in the way they do:

> '…nothing binds the methodologically diverse research fields named by cognitive science as does their conviction that without internal representations over which to operate (e.g. to compute, to use, to manipulate, or to store), no intelligent system could do what it does.

[11] Horder (2001), p.122

> Consequently, cognitive scientists of nearly all stripes and theoretical persuasions tend to posit internal representations to explain how intelligent systems work.'[12]

Representationalist ideas lie at the heart of both a popular folk-psychological and cognitive scientific picture of mind and thought. They occupy a pivotal role in explaining the possession of psychological attributes by human beings. As Sterelny points out, for example,

> `the scientific and the folk picture converge on the idea that representation is central to human minds´.[13]

Cognitive scientists describe and explain the display of often-intricate behavioural strategies by invoking the notion of representations, which are derived from and constituted by our sensory input. Accordingly, we are able to competently navigate our way through the world and apparently effortlessly relate present to former situations by analysing sensory information and by using it to direct and redirect behaviour. Thus, we dash off when a lion or an overly 'enthusiastic' English bulldog is after us, for example, but we stop and wait when a friend is calling. Similarly, we are prompted to console a friend who returns from the exam schools in a flood of tears, but we would share her relief and joy (usually, by applying ample amounts of champagne, flower, whipped cream and confetti to said friend) if she emerges with a victorious smile on her face after having sat her last paper. For both the cognitive scientist and the layman it seems natural to ascribe understanding, knowledge or beliefs to any system that acts adaptively in response to complex environmental stimuli of the kind just depicted.

Historically, the modern representationalism underlying the cognitive view is deeply rooted in Descartes' vision of the mind and consciousness on the one hand, and Lockean empiricism on the other[14]. Descartes, for example, postulated that ideas are essential to consciousness and that every idea is a representation. As he regarded thought as the defining feature of consciousness *having ideas* or *operating on ideas*, *pace* Descartes, is *to think*. A similar position was held by Locke. Like

[12] Stufflebaum (1999), p.637
[13] Sterelny (1990) p.19
[14] It is noteworthy, however, that the likeness or resemblance assumption between object and its representation which is central to representationalism has been adopted by various philosophers since the times of Plato and Aristotle. (See for example: Watson (1995), Chapter 1)

Descartes, Locke believed that ideas rather than objects are the constituents of thought. In opposition to the Cartesian notion of innate ideas, however, he postulated that the human mind is furnished with ideas through the exercise of the senses. Our experience, so Locke thought, is constituted by ideas, which are representations of the objects inhabiting the *real* world. Each human being combines these ideas to construct his own individual *inner mental world* (E 2.2.2). Thus, ideas are representations of objects *outside* the mind. According to both Descartes and Locke representations, while including both general and particular features of the world, are to be thought of as events or going ons within an individual's head or mind, reflecting events outside it.

The kinds of *philosophies* that in the wake of Descartes and Locke have come to be associated with representationalism are best conceived of as a *family of related doctrines*. These doctrines share a number of key assumptions but should not be thought of as constituting a single rigidly defined metaphysical point of view[15]. While Descartes and Locke conceived of representations as ideas, contemporary cognitive science, on the other hand, thinks of them as bundles of information, i.e. symbolic representations, on which mental processes operate. And rather than being a Lockean *storehouse of ideas* modern cognitive science pictures the human mind as an information processing system similar to a computer. Analogous to the Cartesian and Lockean ideas, however, the mental representations of cognitive science are conceived of as *mental causes* and their interactions are thought to constitute *mental processes*. Influenced by modern computer technology the cognitive view takes the human mind to be a complex system, which operates on or computes information constituted by mental representations. And while rejecting the Cartesian notion of an immaterial mind, thus attempting to cast off the shadow of dualism, many cognitive scientists and contemporary philosophers believe the brain to be the material basis of mind. As a result, the terms mind and brain are often used interchangeably and synonymically[16].

---

[15] Stephen Stich points out that the "market" offers a dizzying variety of theories of mental representation and content ranging from causal covariation theories, teleological theories, functional role theories, to single factor theories, multiple factor theories, narrow theories, wide theories as well as variations of all these themes (see: Stich 1992*).*

However, rather than concerning itself with a discussion of either one of these theories the present investigation aims to examine and refute the common premises underlying all of these theories of mental representation.

[16] See for example: Chomsky 1966, 1995. Also: Stillings et al 1998

The mind/brain is envisaged as a form of biological information processing unit or biological computer, which receives, stores, retrieves, transforms and transmits information. These *operations on information* are subsumed under the term *computations* or *information processes*, indicating that in any case, whatever the mind does, some sort of information (i.e. representation) is processed and worked upon or worked with[17]. As a result, cognitive scientists are prone to explain the possession of psychological attributes by human beings by applying psychological predicates to the brain. Statements like the brain or a sub-system of the brain thinks, believes, remembers, hopes, fears or recognizes are common phrases applied in explanations of mental phenomena. However, the meaningfulness and intelligibility of such claims is questionable. The criteria for the application of psychological predicates are constituted by the various behavioural patterns which are displayed in specific contexts and which are applied against a background of widely ramifying and complex capacities manifest in human behaviour (or systems that behave in sufficiently similar ways like human beings). Yet, none of these criteria are met by the brain, and unless the ascription of psychological attributes to the brain is meant to be taken figuratively or metonymically the bounds of sense are being transgressed. The ascription of psychological attributes to the brain (or parts of the brain) constitutes a violation of the logic of whole/parthood relationships and the error underlying this violation has by various authors been described as the *Homunculus Fallacy* (e.g. Kenny 1987) or the *Mereological Fallacy* (e.g. Hacker 1993; see also: Ryle 1990). The consequences of the violation of the logic of whole/parthood relationships will be discussed *en detail* in chapter III (see also 3.1 below).

Fundamental to the consistency of the idea of a causal interaction of mental representations (mental processing) is the claim that mental representations possess *content* or *intentionality*[18]. This claim is motivated by the thought that without the

[17] Note: The cognitive view is sometimes also referred to as the *information processing view* of the mind.

[18] *Intentionality* denotes the property of mental states of *being about* something. According to Franz Brentano (1874) all and only mental states possess intentionality: *'Every mental phenomenon is characterised by what the scholastics of the middle ages called the intentional (or mental) inexistence of an object, and what we might call…reference to a content, direction toward an object (which is not to be understood here as meaning a thing)´* (Brentano 1874, p.88ff.).

Intentional in-existence, meaning existence in the mind as an essential component of a mental act, thus distinguishes mental from physical phenomena. Everything cognitive is intentional, and intentional inexistence is the key aspect of the mental. Furthermore, it implies that any non-mental

basic property of *being about something,* there can be no intentional causation and consequently no causal explanation of behaviour (see also 3.2 below). In the context of folk psychology, for example, the mental causes, or *propositional attitudes (PAs)* as they are commonly referred to, are conceived of as *psychological dispositions* (e.g. an agent's desires, beliefs, hopes, fears etc.), which are directed towards a specific intentional-state[19] (for example, mental states with representational content such as 'that I have to meet my professor today to discuss my thesis' or 'that Achilles is a lovely English bulldog'). The interaction of different propositional attitudes (i.e., the mental processing of representation) is conceived of as the basis for an explanation of human agency. An agent's behaviour is thus explained by invoking the mental causes, his propositional attitudes, for the agent's doing thus-and-so[20]. The propositional attitudes are taken to constitute an inner representation or inner map of the world, as it were, according to which an organism plots its course of action. While these ideas provide a rough sketch of the human mind and an explanation of thought and behaviour, the cognitive view manifests the attempt to provide an understanding of these features in causal mechanistic terms.

3.1 Medium and Nature of Representation

Besides the claims that information processes are representational and characterised by intentionality another central postulate of cognitive science is the idea that the central processes of the human mind constitute a *physical symbol system,* which *encodes meanings*[21]. This claim is motivated by the functionalist dilemma of not being able to account for the intentionality of mental phenomena. Contemporary computationalist, i.e. representationalist, theories are the offspring of functionalist theories of mind which originated in the 1960s. Inspired by Turing's work

---

state that seems to possess content must do so in virtue of some mental state, which possesses that content. This leads to a distinction between `underived/natural´ and `derived´ intentionality.

[19] All propositional attitudes are characterised by *intentionality*. Their intentionality or content can be described by a *that* clause.

[20] Colin McGinn, for example, elaborates this idea thus: 'First reference is made to a particular *person*; then some *attitude* is ascribed to that person; finally a *content* is specified for that attitude. Assertions of this form tell us who has what attitude towards which proposition. By making and receiving such assertions we come (it seems) to understand other people: what they do, why they want such and such, what made them hope for so and so, and so forth. Varying the three elements in the *person-attitude-content* structure gives us a seemingly powerful system for describing the minds of others (and our own), a system both antique and ubiquitous. Thinking of this system as a (tacit and unformalized) theory, we can say that folk psychology is a theory that centrally employs an explanatory ontology of persons and contentful attitudes; with these basic theoretical resources it sets about its explanatory and descriptive work.' (See: McGinn 1989, p.120)

[21] See e.g.: Newell 1980, Newell et al 1989, Stillings et al 1995, p.20.

on machine tables, Hilary Putnam (1967) put forward his *theory of functionalism*, the idea that mental states are to be identified with respect to the causal or functional role they mediate between sensations and behaviour. The functional role it was supposed, would determine the nature or identity of a particular mental state. However, though a functionalist analysis and definition of mental states has its merits with regard to such states as pain, for example, a functionalist theory falls short of explaining the decisive feature of such mental phenomena like thoughts, believes, hopes etc.. In addition to the functional role these phenomena are playing they are characterised by being about something. Functionalism fails to explain intentionality as it does not provide a clear connection between playing a particular causal role and being about something. This problem is addressed in representationalist theories of mind through the additional claim that the mental states identified and accounted for by functionalism encode meanings. Consequently, in a computationalist context cognition is viewed as a multitude of formal symbol manipulation processes which can be interpreted semantically. The possibility of giving a semantic interpretation of cognitive processes allows for giving an account of their meaningfulness/intentionality.

3.2 The Language of Thought

The idea that the human mind encodes meanings is itself not without problems, however. The claim that mental states encode meanings begs the question as to how the physical implementation of meaningful information processes is achieved. Trying to gain an understanding of the *medium* in which computational processes are implemented in is now one of the key problems concerning an overall understanding of cognition[22]. A highly influential answer to the problems surrounding the physical encoding of meanings has been provided by Jerry Fodor (1975). Writing from a representationalist perspective, Jerry Fodor regards cognition as consisting in computational operations on *internal mental representations* in a *language of thought (LOT)*. This LOT, so-called *mentalese*, is pictured as an *innate language* providing a

[22] Characteristic of the attempts to solve this problem is the renunciation of the idea of an immaterial Cartesian mind. Instead, cognitive scientists have by and large opted for a materialist position and declared the brain to be the physical medium in which cognition occurs. Adopting the view that cognition is the result of the workings of a material biological structure and not the product of the mysterious workings of an immaterial substance is reminiscent of the change in point of view which occurred among biologists in the 19th century. Then it was no longer thought that life is not the result of a special life force but rather is the result of a characteristic organisation and interplay of matter and energy.

medium of thought, while containing all the necessary conceptual resources to express any belief or hope a human being might ever harbour:

> 'At the heart of the theory (of representationalism) is the postulation of a language of thought: an infinite set of *mental representations* which function both as the immediate objects of propositional attitudes, and as the domains of mental processes. More precisely, the representationalist theory of mind is the conjunction of the following 2 claims:
> *Claim 1*(the nature of propositional attitudes): For any organism O, and any attitude A toward the proposition P, there is a (computational/functional) relation R and a mental representation MP such that MP means that P, and O has A iff R to MP…
> *Claim 2*(the nature of mental processes): Mental processes are causal consequences of tokenings of mental representations.' [23]

Closely linked to these ideas is Fodor´s aim to give a reductionist *naturalistic account of content* claiming that

> '…sooner or later the physicists will complete the catalogue they have been compiling of the ultimate and irreducible properties of things. When they do, the likes of spin, charm and charge will perhaps appear on their list. But *aboutness* surely won't; intentionality simply doesn't go that deep. It's hard to see…how one can be a realist about intentionality without also being a, to some extent or other, a reductionist. If the semantic and intentional are real properties of things, it must be in virtue of their identity with…properties that are *neither* intentional *nor* semantic. If aboutness is real, it must be really something else.' [24]

Thus conceived, the LOT is an attempt to provide a foundation for thought and meaning and consequently of rational human agency within causal reductionist terms. While few have followed Fodor in adopting this extreme hypothesis, some weaker form of a LOT view, i.e., that there is a mental language that is different from human spoken languages, has been adopted by a wide variety of cognitive scientists.

4. The Implications of the Cognitive View – Targets Revisited

After having outlined the basic tenets and believes underlying the cognitive view it is now necessary to spell out their implications for the explanations of mind and thought proposed by cognitive scientists. The contemporary form of

[23] Fodor (1987), p.16f
[24] ibid.p.97

representationalism endorsed by cognitive science (and exemplified in Fodor´s LOT hypothesis, for example) constitutes a modern attempt to provide an answer to the question as to the nature of mind and thought. In particular, it attempts to offer an account of how human agency is possible by explaining the possession of psychological attributes by human beings and by accounting for the question as to how mental states can be *about* something. How can the mental be part of nature? How, it is asked, can such apparently *puzzling* and *mysterious* properties be integrated into our existing sophisticated scientific picture of the world? It is the purpose of the following section to illuminate the consequences and ramifications of the answer given by cognitive science to these questions.

4.1 Modern Representationalism and its Cartesian Legacy

As the brief historical sketch provided above indicated modern representationalism is a descendant of both Cartesian mind body dualism and Lockean empiricism. The link between the *classical representationalism* of Descartes and Locke and the cognitive view is philosophically of the utmost significance as it shows some interesting parallels between the classical and the modern way of thinking about the mind. First, the modern representationalist position depicts thinking and other mental phenomena such as language, memory, perception etc. as cognitive processes that require the use, manipulation and storage of internal mental representations. In doing so, it retains a distinctly Cartesian/Lockean point of view. Both Descartes and Locke deemed the operation of *ideas* to be the basis of mind and cognition. Thus, it would appear that with regard to the operations of the mind the modern cognitive science point of view has not advanced substantially beyond the basic ideas articulated by Descartes and Locke. In essence, it seems that the information processing view of the mind is nothing but the classical Cartesian/Lockean view clothed in fancy dress, woven out of threads of *updated terminology*[25], as it were. Yet, despite the fancy dress the basic logical features of classical representationalism are retained in its modern pendant. For example, it is characteristic of both classical and modern representationalism that it implies an asymmetry with respect to the access and content of mental states by assuming that

[25] In terms of development and progress or thought, it appears that all that has happened is that a modernised *jargon* has been introduction into the debate. This, in turn, is indicative of the influence which modern computer technology has exercised over philosophy and cognitive science over the last decades.

the mind is asymmetrically accessible from first and third person perspectives. That is, while mental states can be ascribed to others on the basis of behavioural evidence, no such evidence is needed for the self-ascription of a mental state. Accordingly, it seems that I can only know that someone else is in pain if another person exhibits pain behaviour or if he or she is saying so. I on the other hand, do know immediately whether I am in pain. It seems that I have direct unmediated access to the contents of my mind[26].This apparent asymmetry between the first and third person access to the mind ultimately precipitates in the assumption that the mind is a private inner realm to which I have privileged access to. On the other hand, one readily believes that one has only indirect access to other people's minds. These are part of an outer public realm, which is independent of the goings on in our own minds and which is equally accessible to everyone. This *Inner/Outer picture* of the mind plays a crucial role in the characterisation of mental states as *directly accessible* and *private* and has an immediate impact on the questions regarding the explanatory priority regarding linguistic and mental content as it is intimately bound up with ideas regarding the nature of language (see 3.3 below). The Inner/Outer picture of the mind has been introduced into philosophy by Descartes and adopted by both his followers and the British empiricists. As part of the Cartesian legacy it is an integral component of the cognitive view. The examination of the intelligibility of this picture will be an integral part of the present philosophical investigation (see Chapters III and IV, also Bennett and Hacker 2003, Hacker 1997b, Kenny 1984, 1989).

Second, the link between classical representationalism and the modern representationalist stance also sheds some interesting like on the change in point of view that has occurred with regard to the mind brain relationship. It is characteristic of modern cognitive science that it renounces Cartesian dualism. Rather than conceiving of an immaterial substance as the bearer of mental/psychological properties like Descartes and his followers did, cognitive scientists view the brain as possessing psychological attributes. This change manifests itself in the fact that

[26] It is noteworthy, that this asymmetry also seems to apply to linguistic meaning. While it seems that I always know the meaning of my own words (under usual circumstances) I sometimes seem to have to rely on some kind of indirect clue or evidence to understand the meaning or words or sentences uttered by somebody else. I.e., I might have to ask that person if the meaning of certain words or a sentence are not clear to me, or I might have to infer the meaning from the linguistic or behavioural context in which the words or sentence were used. In either case it seems that there exists a similar first/third person asymmetry with the respect to linguistic meaning as exists between my mental states and other person's mental states.

cognitive scientists use the terms mind and brain more often than not interchangeably. Thus, it is the mind/brain which is viewed as a biological information processing unit which encodes amongst other things the meanings of our thoughts, believes, desires, hopes, fears etc. rather than an immaterial substance. As a result cognitive scientists often explain mental phenomena by claiming that the brain sees, hears, thinks, recognizes etc. and is engaged in a variety of other activities which are accounted for by the ascription of psychological predicates to the brain. However, such claims still have a distinctly Cartesian ring to them, as cognitive science ascribes a similar range of properties to the mind/brain in its explanations as Descartes did to the soul. The explanation of the possession of psychological attributes by human beings by reference to the brain (which Cartesians ascribed to an immaterial mind or soul), thus, seems to be an incidental corollary of the repudiation of Cartesian substance dualism by cognitive scientists. Despite this repudiation, however, cognitive science retains an important aspect of Cartesian philosophy, *viz.* its conceptual framework. Yet, the logic of this conceptual framework is deeply flawed. To ascribe psychological attributes to the mind or the brain makes no more sense than ascribing them to the soul.

As mentioned earlier, both cases constitute a violation of the logic of whole/parthood relationships and are the result of conceptual confusion. What, we have to ask ourselves, does it mean to say that the brain thinks, sees, hears, recognizes etc. What *counts* as the brain thinking, seeing, hearing, recognizing? In the *Philosophical Investigations* Wittgenstein points out that the application of psychological predicates makes only sense if they are applied to human beings or to what behaves sufficiently like a human being (PI §281). Following Hacker (1997b) and Kenny (1984,1987) Chapter III will investigate the implications of this claim for the cognitive view, and illuminate the sources of conceptual confusion, which underlie the violation of whole/parthood relationships.

4.2 Mental Representations – Issues regarding Content, Normativity, Systematicity and the Structure of Thought and Language

A vital feature of the cognitive view is the claim that mental representations encode meanings. This claim underlies the idea of a causal interaction of mental

states and the explanation of human agency, which follows in its wake. It links the idea of mental content/intentionality with an explanatory theory of speech and meaning. Because the content of mental representations can be described as attitudes or relations towards propositions it is also called *propositional content* and the content laden mental states are often referred to as propositional attitudes (PAs) (see above). According to cognitive science PAs are mental causes and provide the basis for the explanation of behaviour in our everyday lives. The possession of propositional attitudes is, in turn, explained by hypostatizing the existence of internal states, which encode those propositional attitudes. The attempt to provide an explanation of agency in terms of propositional attitudes demands that an account of the characteristic properties of content, systematicity and normativity, is provided. The cognitive view satisfies this demand by implying the claim that there exists some logical relational structure between the content-laden mental states, which account for the systematicity of content[27].

The rational logical relations in which mental states stand in respect to each other are taken to be a manifestation of the normativity of content. The normativity of content is further underlined by the characterisation of intentionality in terms of propositional content. It consists in the fact that words can be used correctly and incorrectly (according to certain rules which provide the standard for their correct application) and/or that believes can be true or false (i.e. the instantiation of certain matters of fact substantiates or refutes a certain believe). The meaning of a word or sentence is determined by the circumstances, which warrant their application, i.e. the situations in which it would be correct to use it. Propositional attitudes specify a condition for the world to satisfy if the attitude is to be fulfilled. For example, the truth of my belief that it will not rain tomorrow depends on the absence of rain and the presence of a cloudless sky and sunshine the next day. This link between normativity and content bearing phenomena is described by Wittgenstein as follows:

> ‚Der Wunsch scheint schon zu wissen, was ihn erfüllen wird, oder würde; der Satz, der Gedanke, was ihn war macht, auch wenn es gar nicht da ist! Woher dieses Bestimmen, dessen, was noch nicht da ist? Dieses despotische Fordern? („Die Härte des logischen Muß?.“)’ (PU§437)

[27] see for example: Frege G (1987), ‘Die Grundlagen der Arithmetik’

> ‘‘A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true--even when that thing is not there at all! Whence this *determining* of what is not yet there? This despotic demand? ("The hardness of the logical must.").’ (PI§437)

The notions of normativity and systematicity point at an important similarity between linguistic meaning and mental content, or rather, between language and thought. Normativity and systematicity are both distinctive features of language and thought. The language of thought hypothesis has emerged as a philosophical attempt to provide an explanation of the link between this features and account of the relation of language and thought. The systematicity and normativity of language is understood as resulting from the compositional structure of sentences. This compositional structure is thought to be paralleled by the compositional structure of inner mental representations, which constitute the linguistic vehicles of thought. Fodor’s *mentalese* is pictured as an innate language providing a medium of thought, while containing all the necessary conceptual resources to express any belief or hope etc., a human being might ever harbour. Thus, the LOT aims to provide not only the foundation of thought and meaning, but in the context of human behaviour, also of rational human agency[28].

4.3 Representationalism and Intrinsic Intentionality – The Relationship between, Natural Content and Linguistic Meaning

A further characteristic of the modern representationalist stance exemplified by the cognitive view and the LOT hypothesis is the attempt to explain linguistic meaning as resulting from mental content, and subsequently to provide a reductionist explanation of mental content by using non-intentional concepts. The basic idea is that mental content can be explained by the possession of *inner mental representations*. These representations, it is thought, possess *natural (underived)* content (also referred to as intrinsic intentionality) because of the causal relationships in which they stand. Without postulating that mental content is characterised by the possession of intrinsic intentionality a causal reductionist explanation of content would not be possible.

The mental representations by means of which mental content is explained are conceived of as *internal* states, which are ontologically independent of the

[28] see Fodor (1975), p.26ff

external world. This claim is reminiscent of Descartes' and Locke's theories of mind according to which ideas are discrete, freestanding, immaterial particulars which exist in an (inner) mental realm that is discontinuous and independent of the external world. Contrary to the Cartesian and Lockean view, however, contemporary cognitive science adopts a materialist position, claiming that mental states are states of the body and realized in the brain. As states of the body the mental states possess intrinsic (non-relational) properties, which account for their ontologically independent existence of the outside world. The independence of the outside world also accounts for the characterisation of the mental states as *inner*[29]*.* Furthermore, the internal states are hypothesized to be the *inner termini* of the causal chains (of propositional attitudes) that instigate behaviour. This claim partly reflects the functionalist ties that exist between representationalism and functionalism. Both views share the believe that behaviour is to be understood as the result of the causal interaction of mental representations.

In addition to the claim that mental representations stand in causal relations to each other, representationalism implicitly also claims that mental representations stand in a causal relation to the outside world:

> 'Thoughts are inner representations; thinking is the processing of inner, mental representations. These representations have a double aspect…their role within the mind depends on their individualist, perhaps their syntactic, properties…and they are representations in virtue of relations with the world. Mental states represent in virtue of causal relations of some kind with what they represent.'[30]

The fundamental idea expressed in this remark is that mental representations are believed to possess content as a result of their causal relations to the external world.

It is characteristic of both mental states and linguistic tokens that they share the property of intentionality. Both possess content. The representationalist idea that mental content can be explained as a result of processes, which can be described in naturalistic non-intentional terms implies that linguistic meaning is derived from mental content and can be accounted for in terms of the reductionist explanation

[29] The issue of Cartesian scepticism regarding the existence of the external world will not be addressed here.
[30] Sterelny (1990), p.39

given for mental content. The reasons why representationalism gives explanatory priority to mental content rather than linguistic meaning is partly due to the fact that that linguistic meaning is tied up with conventions for the use and interpretation of signs. Yet, causal theories of content as implied by Fodor´s representationalist stance, for example, are much more likely to be successful for mental representations than for words. This is due to the fact that whether a certain word is used in the description or reference to a certain thing, object or person to which it refers depends apart from the word meaning, above all, on the intention of the language user. Yet, entertaining certain thoughts is, more often than not, not a voluntary activity. Thus, I automatically think *Achilles* and *Velvet* on seeing my two English bulldogs before me. Similarly, on hearing the sound of water splashing against my roof top and window pain, I automatically think *it's raining*. As a result of this, representationalism assumes that a causal link between mental content and states of affairs of the outside world is easier to account for (as more consistent) as that between linguistic tokens and states of affairs of the outside world. The attempt to account for content in non-intentional reductionist terms is more likely to be successful if mental representations are ascribed natural content/intrinsic intentionality as this makes for a shorter causal chain of events that needs explaining.

The representationalist attempt to explain linguistic content as resulting from mental content and then to give a reductionist account of the latter fails to acknowledge, however, that mental content cannot be explained as a result of freestanding internal mental representations. Consequently linguistic meaning cannot be explained as the result of the "animation of otherwise dead signs by acts of understanding". The discussion in chapter VI will show that words are not injected with meaning through acts of understanding. Instead, their meaning is their use. Mental states are not internal free-standing states of the mind which have to be connected via mechanisms with the world but are intrinsically relational states. Free-standing internal representations cannot account for the normativity of content as they either presuppose what they set out to explain or fail to sustain normativity.

5. Wittgenstein and the Study of Mind

As indicated earlier, the philosophical ideas of the later Wittgenstein will provide the foundation for the present investigation. Representational theories of

mind are continuous with the Cartesian tradition in various respects, in particular, in their commitment to the (scientistic) explanatory role of mental processes and mental representation. As Wittgenstein's rejection of the Cartesian model of mind is closely linked to his philosophy of language and his ideas about the role of language in human life, his philosophy of language, which in turn is based on his revolutionary conception of the nature of philosophy and philosophical investigation, will be discussed in chapter II. Such an overview of Wittgenstein's later philosophy will provide a context for the philosophical ideas applied in the course of this thesis.

An appreciation of the revolutionary character of Wittgenstein's later views is the key to an understanding of the potential and significance of his later philosophy. In the past, the failure to highlight the radical nature of Wittgenstein's later views, has contributed significantly to the misunderstanding of his later philosophy, and consequently to the demise of its influence on current philosophical trends. Although many of Wittgenstein's insights constitute irreversible advances in the philosophy of language and the philosophy of mind[31], his ideas have yet to be absorbed by the majority of contemporary philosophers. By adopting a characteristically Wittgensteinian approach in the investigation of the cognitive view, the present study also aims to highlight the power and potential of Wittgenstein's ideas, which in the contemporary scientistic climate are all but forgotten.

[31] See e.g.: Kenny 1973

Chapter II
On the Significance and Implications of Wittgenstein's Later Philosophy - An Overview

1. A New Philosophy

In the history of philosophy Ludwig Wittgenstein stands out by having developed two entirely different and unique philosophical "worldviews". These precipitated in the publication of his (early) *Tractatus-Logico Philosophicus* and his (later) Philosophical Investigations respectively. Both works had a profound influence on the development of 20th century analytic philosophy in so far as these two works constituted the defining landmarks against which much of the philosophy, which followed in the wake of the publication of these works, was defined. Despite some continuity between the *Tractatus* of the young Wittgenstein and the Investigations of the later Wittgenstein[32] the two books stand diametrically opposed to each other. After ending his self-imposed exile and return to Cambridge in 1929, Wittgenstein undermined the major ideas that informed the *Tractatus* and gradually developed an entirely new philosophical worldview and method(s). This development was based on a novel conception of language, linguistic meaning and view of the relationship between language and reality. Together, the new ideas provided the foundation for the unprecedented and revolutionary conception of philosophy, which Wittgenstein developed and refined throughout his later years. The radical, revolutionary and, quite literally, world-shattering ideas of the later Wittgenstein found their most coherent expression in the *Philosophical Investigations*. Since its publication, Wittgenstein scholars have variously depicted the book as '…having no ancestors in the history of thought '[33], or described it as work which '…can no more be located on the received maps of philosophical possibilities than the North Star can be located on

[32] There exists continuity between the *Tractatus* and the *Investigations* with respect to both theme (e.g. the nature of linguistic representation, the relationship between thought and its linguistic expression, the intentionality of thought, metaphysics and the nature of philosophy) and philosophical judgement (e.g. the *Investigations* reaffirm the young Wittgenstein's criticism of Frege and Russell, underline his denial that philosophy is a cognitive discipline and restate his rejection of psychologism in logic).
[33] v. Wright (1984), p.14

maps of the globe.'[34] However, most people fail to come close to even a partial understanding of the radical nature of Wittgenstein's later thought, which has been highlighted by these remarks. From the days when fragments of Wittgenstein's later philosophy[35] were first circulated among Oxbridge philosophers and the posthumous publication of the *Philosophical Investigations* until today, Wittgenstein's later philosophy tends to suffer from gross misrepresentation and misunderstanding.

The controversies surrounding Wittgenstein's thought and the abundance of misrepresentation and misinterpretation have led to a relative demise in the study of Wittgenstein, and the kind of analytic philosophy inspired by his later works to the point where, as Anthony Kenny has remarked, his insights are in danger of being lost[36]. Today, this neglect is particularly problematic for the philosophical novice, who is indoctrinated early on with the theories and hypotheses of those scientistically minded philosophers, whose works tend to dominate the mainstream of contemporary philosophical thinking. In addition, it prevents the illumination of the crucial problems and understanding of the central questions, which the ever-expanding community of cognitive scientists concern's itself with. It is this community, in particular, for which Wittgenstein's later ideas could be invaluable:

> 'I…believe that there are many areas of philosophy, and many fields of scientific endeavour, in which really fruitful work is unlikely to be done by anyone who has not absorbed what Wittgenstein has to teach. Those who ignore Wittgenstein's critique of false philosophy and pseudo-science take the risk of constructing imposing edifices of thought which turn out to be nothing but houses of cards.[37],

The present chapter will provide an introduction to Wittgenstein's later philosophy and give a synopsis of its central tenets. A particular emphasis will be put on contemporary issues regarding the clarity of Wittgenstein exegesis and accuracy of the interpretation of his later philosophy. Despite the myriad of (introductory) textbooks and essays on Wittgensteinian thought, even the most eminent of Wittgenstein scholars are more often than not unsuccessful in their attempts to convey the relevance of Wittgensteinian ideas to an audience, which has already

[34] Hacker PMS (1996a), p.103
[35] Note: Of course, Wittgenstein's earlier philosophy is not immune to misrepresentation and misunderstanding either, as the dispute about the *"New Wittgenstein"* shows.
[36] Kenny A (1984), p.vii
[37] ibid., p.viii

been indoctrinated or courted by the writings of empirically minded philosophers and cognitive scientists who promise insight into the "mysteries" of mind and brain. The fact that the efforts to prevent misrepresentations of Wittgensteinian thought and to correct false judgement regarding the character of his ideas have met with little, if any, success has lead to a dramatic demise in the reputation of Wittgenstein scholarship among empirically minded philosophers and cognitive scientists. In addition, these developments have been conducive to the fostering of a virulent scientism in philosophy evident in the widespread misunderstanding of the relationship between science and philosophy. Without a profound understanding of the relationship between science and philosophy, however, any interdisciplinary study of the mind and brain is likely to get entangled in a misguided scientistic routine of theory construction and hypothesis testing as evidenced by the development of the cognitive view (see Chapter I). Consequently, the exposition of this relationship within the course of this chapter is crucial to delineating and correcting the philosophical problems surrounding the defining credo of classical cognitive science, which lies at the heart of this thesis. Finally, by sharpening the awareness and understanding of the revolutionary character of Wittgenstein's later philosophy it is hoped that the ground for preventing the undeserved and entirely ill judged dismissal of Wittgensteinian ideas[38] in the future, can be prepared.

1.1 A Break with Tradition – A New Standard of Achievement in Philosophy

Wittgenstein's later philosophy[39] (and to some extent also the kind of analytic philosophy which followed in the wake of the *Investigations*) constitutes a radical

[38] Note: It is precisely because Wittgenstein's later philosophy has no ancestors in the history of thought, that one should be doubly suspicious of philosophers and cognitive scientists who dismiss Wittgenstein alternately as a sceptic, constructivist, anti-realist or behaviourist. With regard to the misinterpretation of Wittgenstein Peter Hacker remarks, for example: *'The temptations to pigeonhole him* (Wittgenstein), *to classify his philosophy as a kind of idealism or of anti-realism, to identify his philosophy of language as a form of assertion-condition semantics standing in contrast to truth-conditional semantics or as propounding a 'use' theory of meaning, to characterize his philosophy of mind as a kind of behaviourism, to think of his philosophy of mathematics as a form of strict-finitism, and so on have evidently been difficult to resist. But resisted they must be if one is to understand Wittgenstein's later philosophy.' (Hacker 2001, p.ix)*

[39] Wittgenstein's later philosophy developed roughly from 1929 onwards after he had returned to Cambridge. As Peter Hacker points out, the reorientation necessary for the development of his new philosophy took place remarkably quickly. This is all the more astonishing, he suggests, if one conceives of the change Wittgenstein's thinking had undergone as similar to the radical transformation in style seen in the works of painters like Kandinsky or Van Gogh. Wittgenstein himself remarked: 'Meine Hauptdenkbewegung ist heute eine ganz andere als vor 15-20 Jahren. Und das ist ähnlich wie

break with over two-thousand years of Western philosophical thought, including his own earlier philosophical worldview as exposed in the *Tractatus.* In the early notes that Wittgenstein compiled after taking up philosophy again in 1929, and which have post-humously been published under the title *Philosophical Remarks*, Wittgenstein tells us, for example, that,

> 'Dieses Buch ist für solche geschrieben, die seinem Geist freundlich gegenüberstehen. Dieser Geist ist ein anderer als der des großen Stromes der europäischen und amerikanischen Zivilisation, in dem wir alle stehen. Dieser äußert sich in einem Fortschritt, in einem Bauen immer größerer und komplizierter Strukturen, jener andere in einem Streben nach Klarheit und Durchsichtigkeit welcher Strukturen immer. Dieser will die Welt durch ihre Peripherie – in ihrer Mannigfaltigkeit – erfassen, jener in ihrem Zentrum – ihrem Wesen. Daher reiht dieser ein Gebilde an das andere, steigt quasi von Stufe zu Stufe immer weiter während jener dort bleibt, wo er ist, und immer dasselbe erfassen will. ...' (PB, Vorwort)

> 'This book is written for such men as are in sympathy with its spirit. This spirit is different from the one, which informs the vast stream of European and American civilization in which all of us stand. That spirit expresses itself in an onwards movement, in building ever larger and more complicated structures; the other in striving after clarity and perspicuity in no matter what structure. The first tries to grasp the world by way of its periphery--in its variety; the second at its centre--in its essence. And so the first adds one construction to another, moving on and up, as it were, from one stage to the next, while the other remains where it is and what it tries to grasp is always the same….' (PR, Preface)

The difference in spirit to which Wittgenstein refers to was most clearly discernable in his novel conception regarding the nature of philosophy and his unprecedented view with respect to the standard of achievement in philosophy. In contrast to the great (metaphysical) tradition of Western philosophy, which varyingly conceived of the subject as an investigation into the essence of all things, an enquiry into the indubitable foundation of knowledge or the *a priori* possibility of knowledge *per se*[40]*,*

---

wenn ein Maler von einer Richtung zu einer anderen übergeht. (MS 183, 141)' (see: Baker & Hacker 2005, p.257)

[40] For Plato, philosophy was an investigation of eternal truths and a path to knowledge of the essences of all things. Aristotle, on the other hand, conceived of philosophy as an investigation into the fundamental premises and methodological principles of the sciences, to which philosophy stands in a continuum. Thus, Aristotelian metaphysics, for example, deals with the first principles of scientific knowledge and the ultimate conditions of all existence (e.g. by reflecting on existence in its most fundamental state (i.e. being *qua* being), and by pondering the essential attributes of existence). With Descartes, philosophy became an enquiry into the indubitable foundation of knowledge (e.g. clear and distinct ideas) in order to establish the unity of all knowledge. In contrast, the British empiricists led by Hume, Berkeley and Locke thought of philosophy as an investigation into the nature and limits of

philosophy, according to the late Wittgenstein, is not a contribution to human knowledge but to human understanding. Thus, Wittgenstein disposed of the hitherto uncontested view of philosophy as a *cognitive discipline,* i.e. the belief that there are philosophical propositions expressing philosophical knowledge. Instead, the later Wittgenstein held that philosophical problems are wrongly conceived of as demanding a solution. They are wrongly conceived of as questions looking for an answer. Philosophical problems, so the later Wittgenstein holds, are not solvable because their very status as problems is the result of a misunderstanding about the ways we use our language[41]. The very idea of a problem or question suggests that somewhere out there is a solution or answer to be found (if we only apply our minds hard enough). But,

> ‘there are no great essential problems, only great and compelling illusions of such problems (BT 407). (…)‘Philosophical problems are an awareness (not typically a self-conscious one) of a disorder in our concepts. (BT 421)’

Rather than striving for answers and solutions Wittgenstein suggests that the philosopher strive for an informed perspective about the way we use certain concepts. In doing so, we will arrive at conceptual clarity, which is accompanied by understanding of how a particular philosophical problems arises. The clarity and understanding the philosopher thus achieves will make philosophical problems disappear (PI§133). Consequently, for the late Wittgenstein achievement in philosophy is to be gauged by how successful a philosopher is in dissolving philosophical problems, and not by an assessment or evaluation of the complexity of

---

human understanding. With the publication of Kant’s *Critique of Pure Reason,* a paradigm shift occurred with respect to the orientation of philosophy. Kant proposed that instead of concerning itself with knowledge of objects (i.e. the objects themselves) philosophers should concern themselves with the question as to how knowledge of objects is possible *a priori*. In doing so, philosophers would gain knowledge of *synthetic a priori* propositions, which constitute the pre-conditions of all experience. And, following in this great tradition, the young Wittgenstein of the *Tractatus* aimed to disclose the *a priori* order of the world by means of logical investigation.

[41]To be exact: It is important to note that Wittgenstein was not the first to associate philosophical problems with misunderstandings of language. Michael Dummett quotes Gottlob Frege in a series of lectures published as `Ursprünge der analytischen Philosophie/ Origins of analytic philosophy´(p.14): `*So besteht denn ein großer Teil der Arbeit der Philosophie... in einem Kampf mit der Sprache´.* With Frege´s `*Grundlagen der Arithmetik´*, a book which for some made him the father of analytic philosophy, came a `linguistic turn/ die Wende zur Sprache.´ The difference between Frege and the later Wittgenstein on this matter is that Frege (like Russell in the `Principia´) saw the solution of the problem in the design of an ideal(-logical) language. This was regarded as one of the main tasks. Wittgenstein repudiated this idea in both the `Tractatus´, as well as in the `Investigations´. The peculiarity of the `Investigations´ is the turn towards an investigation of `ordinary language´. (see: PR§3, PI§108)

the theories he proposes as solutions to philosophical problems. Because philosophical problems are to be understood as conceptual problems philosophy cannot be an investigation into the most general features of the universe[42]. A philosophical investigation is a conceptual investigation (Z§458).

The difference in spirit which Wittgenstein alluded to in the *Philosophical Remarks* constituted nothing less than a radical reorientation with regard to the nature of philosophy, philosophical problems and the standard of achievement in philosophy. This reorientation stood not only in stark contrast to the views of most of Wittgenstein's philosophical contemporaries like Moore and Russell[43], for example, but also stands diametrically opposed to the conception of philosophy currently *en vogue*. This is particularly true for those areas of the philosophy of mind and psychology whose manifold philosophical propositions inform theories about mind and brain in cognitive science. With regard to its significance, Wittgenstein remarked about his "change of perspective" that,

> 'Wenn wir z.B. unsere Untersuchungen mit "Philosophie" bezeichnen, so erscheint dieser Titel einerseits angebracht, andererseits aber hat er sicherlich manch einen irregeführt. (Man könnte sagen, dass der Gegenstand, mit dem wir uns beschäftigen, einer der Erben des Gegenstandes ist, den wir „Philosophie" zu nennen pflegten.' (BB_D, p.53)

> 'If, e.g., we call our investigations „philosophy", this title, on the one hand seems appropriate, on the other hand it certainly has misled people. One might say that the subject we are dealing with is one of the heirs of the subject which used to be called philosophy.'(BB_E, p.28)[44]

Nothing but Houses of Cards

Why did Wittgenstein feel the need for such a reorientation? Writing in the tradition of Western philosophy, the young Wittgenstein had deemed logic and metaphysics to be the principal constituents of philosophy, with logic, which he thought of as the mirror of the essence of the world, providing the foundation of philosophy (NB 39, 106). Consequently, he thought of logical investigation as the method *par excellence* to explore the nature of all things (PI§89). These ideas found

[42] See e.g.: Moore GE 'What is Philosophy?', p.1f in: Moore GE (1962) Some Main Problems of Philosophy; also: PP
[43] ibid. p.1ff

their culmination in the *Tractatus.* After the completion of this book Wittgenstein abandoned philosophy thinking, with characteristic modesty, that he had solved all the problems of philosophy. Like the great philosophical systems of the past, Wittgenstein' philosophical vision of the *Tractatus* was founded on a set of metaphysical premises:

> 'Die logischen Sätze beschreiben das Gerüst der Welt, oder vielmehr, sie stellen es dar. Sie „handeln" von nichts. Sie setzen voraus, daß Namen Bedutung, und Elementarsätze Sinn haben: Und dies ist ihre Verbindung mt der Welt. (...)' (TLP 6.124)

> 'The logical propositions describe the scaffolding of the world, or rather they present it. They "treat" of nothing. They presuppose that names have meaning, and that elementary propositions have sense. And this is their connexion with the world. (…)' (TLP 6.124)

Yet, after returning to philosophy Wittgenstein realized that this metaphysical vision of the world rested on a fundamental error. Baker and Hacker have remarked that the sublime metaphysics of the *Tractatus* was *'purchased with illusion and distortion'* for as the later Wittgenstein himself realized, the idea of logic as the mirror image of the world rested on the misconceived notion that the proposition must do something remarkable (PI§93), thereby mistakenly projecting the properties of the chosen mode of representation on the linguistic entities thus represented[45] (PI§104). Metaphysics will not lead us to knowledge of necessary truths, which disclose truths about objective necessities in nature. Rather, metaphysics (i.e. metaphysical questions), express an unclarity about the grammar of our concepts (BB35); it is nothing but houses of cards that need to be destroyed (PI§118), as these have been built out of misunderstandings in the way we use our language (see above).

Investigating the Essence of the World

> 'Philosophische Untersuchungen: Begriffliche Untersuchungen. Das Wesentliche bezüglich der Metaphysik ist: sie verschleiert den Unterschied zwischen faktischen und begrifflichen Untersuchungen.' (Z_D§458)

[45] Baker & Hacker (2005), p.257

> ‘Philosophical investigations: conceptual investigations. The essential thing about metaphysics: it obliterates the distinction between factual and conceptual investigations.’ (Z_E§458)

The problem when doing metaphysics is, as Wittgenstein points out in this remark, that philosophers mistakenly think of metaphysical propositions, which are of conceptual (normative) nature, as being akin to or even identical with scientific propositions, which are factual and descriptive[46]. In contrast to the propositions of the empirical sciences, which describe contingent truths of reality, metaphysical propositions are traditionally deemed to express necessary truths not merely of this world but of any possible world. In doing so, they often take a modal form. It is said, for example, as Hacker points out that

- what is coloured *must* also be extended, that
- nothing *can* be read and green all over, or that
- one *cannot*[47] travel back in time.

As modal sentences like these are used in descriptions of causal connections or practical statements in the empirical sciences (i.e. modal sentences have an empirical use in descriptions) philosophers are inclined to view such metaphysical propositions as descriptions too. But in doing so, they fail to distinguish between the factual and the conceptual, i.e. what is (really) descriptive and what is (merely) normative. The necessity ascribed to a metaphysical proposition is merely one of convention, of grammar[48] as it were, the later Wittgenstein holds, expressing a rule for the use of a word or expression:

[46] For the following passage I am indebted to Hacker (1996a) p.118f.

[47] The metaphysical use of ‘cannot’ does not describe an impossibility or human inability, rather it is indicative of a rule for the use of a word (see e.g.: BB54)

[48] Note: Wittgenstein’s use of the term *grammar* is much broader than its customary use. *Grammar* in Wittgenstein’s writings refers not only to our *actual* uses of words, but also extends to *impossible* uses of words. Hilmy, for example, points out that in a Wittgensteinian sense *‘grammar’ … is the ledger of our actual linguistic transactions, it is as it were, the ledger of the sorts of moves that are (can be) and are not (cannot be) made in a given “language game”.’ Grammar, ‘ refers not only to the ledger of our linguistic practices but also to the ledger of meaningful actions of every kind – facial expressions people make in various circumstances, modes of dress, ways of playing games, assembling furniture, worshipping god, gardening, preparing food and so on.’* (See Hilmy: (1987, p.129))

> 'The 'necessity' which we ascribe to such (metaphysical) 'truths' is the mark not of 'a necessary fact' (since there is no such thing), but of our commitment to these concepts, of our inflexibility in employing these expressions in accord with these rules.'[49]

By disclosing necessary truths, metaphysics has traditionally been thought to provide insight into the essence of the world, thus giving philosophy a subject matter of its own[50]. Wittgenstein's critique of metaphysical investigations as obliterating factual and conceptual investigations (Z§458) shows this conception to be profoundly misconceived. '*Essence'*, Wittgenstein holds, *'is expressed by grammar'* (PI§371). The grammar of our words and expressions specifies what is to count as instances of this and that. Accordingly, our grammar is *determining* the nature of things rather than describing it (as is assumed of metaphysical propositions). Talk of essences is nothing but talk of (linguistic) conventions. It is the task of the philosopher to show that a metaphysical proposition "hides" a grammatical rule (BB55). According to Wittgenstein, it is the philosopher's task to untie these knots in one's understanding by means of conceptual clarification, which will not increase our knowledge of the world but bring about an insight into where our understanding goes wrong. After all,

> ‚Die Philosophie ist ein Kampf gegen die Verhexung unseres Verstandes durch die Mittel unserer Sprache. (PU§109)

> ‚Philosophy is a battle against the bewitchment of our intelligence by means of language. (PI§109)

1.2 Conceptual Investigations, *‚Übersicht'* and *‚übersichtliche Darstellung'*

The battle against the bewitchment of our understanding is to be fought by means of conceptual investigations. Instead of engaging in speculative metaphysics Wittgenstein holds that what philosophers need to do is to substitute wild conjectures and explanations by quiet weighing of linguistic facts (Z§448), i.e. by describing the use of language (PI§133). The use of language is determined by grammar, laying down the rules for the use of concepts and expressions. The conceptual investigations the philosophers is to undertake are thus to be understood as

[49] Hacker (1996a) p.119
[50] A subject matter distinct from the natural sciences, which study the contingent properties of objects of reality.

normative investigations in so far as they are investigations into the rules governing our use of language.

> 'When philosophers use a word and enquire after its meaning, one must always ask oneself: Is the word ever actually used like this in the language for which it was made. One will find that it isn't so, and that the word is being used contrary to its normal grammar.' (MS 109 (Vol.V) 246)

This practice not only implies the specification of the use of words in the practice of speaking a language but also involves the tracing of conceptual connections and their ramifications. Strawson later coined the term *connective analysis*[51] for this type of investigation (see also chapter I, section 2). For example, in PI§558 Wittgenstein remarks that the rule which shows us that that the word 'is' has different meanings in the sentences "the rose is red" and "two and two is four", is the one allowing us to replace the word 'is' in the second sentence by the sign '=' while at the same time forbidding this substitution in the first sentence. Similarly, the statement that while it makes sense to say about other people that they doubt whether I am in pain makes sense, but that I it does not make sense when I say it about myself, e.g. I don't know whether I am in pain (PI§246) describes a rule governing the use of the word 'to know'. Alternatively, we may simply remind us of certain aspects of use of a word or expression as Wittgenstein does in PI§183:

> ' Wir können sagen…, wir sagen auch..., oder….' (PU§183)
>
> ,We may say..., but also that…, or… .' (PI§183)

Sometimes, however, one may choose not to state a grammatical rule explicitly, but prefer to merely hint at it. The expression 'an inner process stands in need of outward criteria' (PI§580), for instance, which is part of the repudiation of conception of believing as an inner process, hints at the fact that expressions which signify 'inner processes' or 'inner states' are grammatically bound up with behavioural criteria for

[51] `*What this amounted to was the description, for purposes of philosophical elucidation, of the interconnectedness of related concepts, of their implications, compatibilities, and incompatibilities, of the conditions and circumstances of use of philosophically problematic expressions. Such analysis does not terminate in logically independent elementary propositions, or in simple, unanalyzable names or concepts. It terminates in the clarity that is obtained with respect to a given question when the network of concepts has been traced through all its relevant reticulations.´* Strawson (1991), p.10

their application in the third person case[52]. These few examples by no means exhaust the myriad of ways in which we may state grammatical rules to remind us of the way certain words are used. But they suffice to demonstrate how one might go about conducting a philosophical investigation *à la* Wittgenstein. Furthermore, they show that grammatical rules in a late Wittgensteinian sense comprise

> `…not just school-grammatical or syntactical rules, but also explanations of meaning (PG§68,143; PR78). (...) Such rules include
> (a) definitions whether in formal (`"Bachelor" means "unmarried man"´) or material mode (´Bachelors are unmarried men´);
> (b) analytical propositions (´All bachelors are unmarried men´);
> (c) colour-charts and conversion-tables (BB4, LFM 118);
> (d) ostensive definition (BB12, 90);
> (e) explanations by exemplification (PI §§69-79);
> (f) expressions of the 'geometry' of colour like `Nothing can be red and green all over´(PR51-2;LWL 8);
> (g) propositions of arithmetic and geometry(PR 143,170,216,249;LWL 8,55; PG 319,347).´[53]

All of these various instantiations of grammatical rules set standards for the correct use of a concept or expression, and thus determine their meaning. To give the meaning of a word is to specify its grammar[54]. These examples also illustrate that there is not one single unified approach one can follow in investigating the grammar of words and in striving to make philosophical problems disappear. As Wittgenstein remarked,

> '…Es gibt nicht *eine* Methode der Philosophie, wohl aber gibt es Methoden gleichsam verschiedenen Therapien.' (PU§133)

> '…There is not *a* philosophical method, though there are indeed methods, like different therapies.' (PI§133)

Although a picture of Wittgenstein's later ideas is emerging, the suggestion that the description of rules for the uses of language will ultimately lead to the dissolution of philosophical problems, may still seem a bit of a leap. To gain a firmer

[52] See e.g.: Baker & Hacker (2005), p.292
[53] See: Glock (1997), `A Wittgenstein Dictionary´, p.152
[54] ibid, p.150

grasp on this idea it is helpful to consider the following remark from the *Philosophical Investigations*:

> ‘Es ist eine Hauptquelle unseres Unverständnisses, dass wir den Gebrauch unserer Wörter nicht *übersehen.* – Unserer Grammatik fehlt es an Übersichtlichkeit. – Die übersichtliche Darstellung vermittelt das Verständnis, welches eben darin besteht, dass wir ‚Zusammenhänge sehen’. Daher die Wichtigkeit des Findens und des Erfindens von *Zwischengliedern.*
>
> Der Begriff der übersichtlichen Darstellung ist von uns von Grundlegender Bedeutung. Er bezeichnet unsere Darstellungsform, die Art, wie wir die Dinge sehen. (Ist dies eine ‚Weltanschauung?’).’ (PU§122)

> ‘A main source of our failure to understand is that we do not command a clear view of the use of our words.--Our grammar is lacking in this sort of perspicuity. A perspicuous representation produces just that understanding which consists in 'seeing connexions'. Hence the importance of finding and inventing *intermediate cases*.
>
> The concept of a perspicuous representation is of fundamental significance for us. It earmarks the form of account we give, the way we look at things. (Is this a 'Weltanschauung'?) (PI§122)’

Here Wittgenstein identifies the lack of an overview[55] (*Übersicht*) of the various uses of our words, as one of the main sources of conceptual confusion. By tabulating the various grammatical uses of words in the form of a surveyable representation (*übersichtliche Darstellung*), however, one achieves the kind of clarity in the use of our language in the light of which philosophical problems will disappear. The philosopher will not only gain an insight into the various ways a word is used (and thus demarcate the bounds of sense) but also obtain an understanding of its conceptual connections and their ramifications. Consequently, one may think of an overview as helping us to put our ideas as to what can be said about the world in order.

> ‘What one can do when one has an overview, is *survey the linguistic environs of a problematic expression.* One can ‘see’ (take in) the use of the problematic concept, its manifold

[55] Following Peter Hacker, *Übersicht* will be translated as ‘overview’ and *übersichtliche Darstellung* as ‘surveyable representation’. See Baker and Hacker (2005), p.308

> grammatical connections with related concepts, as well as the differences between it and the concepts with which it is liable to be confounded.'[56]

By gaining an overview of the grammar of a particularly problematic concept, the philosopher will gain an understanding of a specific segment of our language, which ultimately culminates in an insight where our understanding went wrong and how we came to tie the knots into our understanding in the first place. It is in this way that the new way of dealing with philosophical problems resolves or dissolves problems rather than providing answers to them (AWL 27f.)[57].

From the Quest for Truth to the Quest for Sense

Wittgenstein's call for an overview of a segment of our language to dissolve philosophical problems is linked to his critique of metaphysical investigations as confusing factual and conceptual investigations (Z§458) (see above). When investigating the essence of the world the metaphysician is prone to take words like "knowledge", "being", "object", "I" etc. and use them in ways not specified by their grammar. That is, metaphysicians are prone to violate the bounds of sense. Doing so results in the knots in our understanding, which Wittgenstein aims to resolve by means of conceptual investigations. Instead, he advises us to ask whether the words found in metaphysical questions and propositions are 'ever actually used in this way in the language, which is its original home' (PI§116). Gaining an overview of a segment of our language helps the philosopher to answer this question and untie the knots in his understanding by bringing

> '…Wörter von ihrer metaphysischen, wieder auf ihre alltägliche Verwendung zurück.'
> (PU §116)

> '… words back from their metaphysical to their everyday use'. (PI§116)

In the *Tractatus* the young Wittgenstein wrote:

[56] Baker & Hacker (2005), p.309

[57] However, conceptual clarity does not consist in the accumulation of fresh knowledge or new facts, i.e. the measures by which scientists gauge their success or failure. Rather conceptual clarity consists in the 'Übersicht' (the surview or perspicuous representation) the philosopher gains over a (specific) segment of our language.

‘Die meisten Sätze und Fragen, welche über philosophische Dinge geschrieben worden sind, sind nicht falsch, sondern unsinning. Wir können daher Fragen dieser Art überhaupt nicht beantworten, sondern nur ihre Unsinnigkeit feststellen. Die meisten Fragen und Sätze der Philosophen beruhen darauf, daß wir unsere Sprachlogik nicht verstehen.
(Sie sind von der Art der Frage, ob das gute mehr oder weniger identisch sei als das Gute.)
Und es ist nicht verwunderlich, daß die tiefsten Probleme eigenlich *keine* Probleme sind.’ *(TLP4.003)*

*‘*Most propositions and questions, that have been written about philosophical matters, are not false, but senseless. We cannot, therefore, answer questions of this kind at all, but only state their senselessness. Most questions and propositions of the philosophers result from the fact that we do not understand the logic of our language.
(They are of the same kind as the question whether the Good is more or less identical than the Beautiful.)
And so it is not to be wondered at that the deepest problems are really no problems.‘ (TLP4.003) [58]

Although the later Wittgenstein’s views on logic, language and the logic *of* our language differed dramatically from his earlier ones his view, that most philosophical questions and propositions constitute transgressions of the bounds of sense, remained constant[59]. In a series of lectures given in the early 1930s Wittgenstein explains that to say of a combination of words that they constitute nonsense is tantamount to saying that they are meaningless combination of marks or sounds. As such they do not belong to language, (although their constituents do) and lack any use (see: LWL 48). That the bounds of sense have been transgressed can be shown by explaining the language game[60] to which these marks or sounds belong and

[58] Note: The suggestion, that *‘Most propositions and questions, that have been written about philosophical matters, are not false, but senseless’* undermines the conception of philosophy as a cognitive discipline. Traditionally, philosophical questions have been conceived of as deep and significant. Flawed attempts to answer them were simply false but not nonsensical. To deny that philosophical questions constitute questions in search for an answer is tantamount to denying the possibility of philosophical knowledge (expressed in philosophical propositions).

[59] *‘Nonsense according to the Tractatus, results from failing to observe the rules of the logical syntax of language. Within philosophy, this typically occurs when philosophers endeavour to say what can only be shown, ask questions about the essential nature of things and try to answer them, rather than pointing out their illegitimacy. For such questions characteristically employ formal concepts as if they were genuine, material concepts. And attempts to answer them involve the use of pseudo-propositions which try to assert that certain internal properties and relations obtain (TLP 4.12-4.1272). The result is combinations of signs that do not satisfy the logical requirements of representation, that violate the formation rules for propositions by employing what are in effect variables as if they were names.’* See: Hacker PMS (2000a), p.94f.

[60] Wittgenstein uses the idea of language games to illuminate the conception of language as a rule-guided activity. The understanding of a language involves the mastery of techniques concerning the applications of rules, for our *language games*. In this context, it is also important to note that such

demonstrating that they have in fact no use within that game. To say that a combination of words, marks or sounds is nonsense is to exclude it from language. Importantly, however, this does not imply that they could no be given a sense. Rather, it requires that if one wishes to include a certain sign or combination of signs into language, one needs to specify rules for their use (i.e. one needs to specify what counts as doing thus and so, for example).

Metaphysical questions and theories provide the paradigm example of such transgression of the bounds of sense. The theories or hypotheses, which scientistically minded philosophers are prone to propose in answer to a metaphysical problem or question are nothing but symptoms of their conceptual entanglement and not, as some believe, answers to *general questions about the universe,* which increase our knowledge of the world*.* According to the later Wittgenstein, it is the philosopher's task to untie these knots in one's understanding by means of conceptual clarification. Wittgenstein scholars have generally characterised this part of Wittgenstein's approach as the negative or therapeutic aim of philosophy, i.e the destruction of houses of cards resulting and the disentangling of conceptual confusion. The purpose of this is to convey to the confused philosopher an understanding of the absurdity and illusory nature of the basic claims and premises informing his views. This is complemented by what has commonly been regarded as the positive aim of Wittgenstein's later philosophy, i.e. to gain an overview of the grammar of a problematic concept or expression, which helps the philosopher to gain insight into where one's understanding goes wrong (where imagination takes over the task of the intellect) by delineating the workings of language[61]. Or to put it in Wittgenstein's own words, what is necessary is to put

> ' …in order our ideas as to what can be said about the world...[62]´

---

rules are not applied according to a theory. Rather, the rules governing the use of words in our language are determined within the public realm of our *language community.* These rules thus constitute an integral part of the *form of life* of a given language community*.* In a Wittgensteinian context this term is meant to stress the intertwining of culture, worldview and language:
'Das hinzunehmende, Gegebene – könnte man sagen – seien Lebensformen. What has to be accepted, the given, is - so one could say – forms of life.´ Wittgenstein, PI (II) p.226/226e; See also:
`...Und eine Sprache vorstellen heißt, sich eine Lebensform vorstellen. / ....And to imagine a language means to imagine a form of life.´(PI§19)

[61] Naturally, the negative and positive aims of philosophy overlap.
[62] Moore (1970b), p.323

2. Sources of Conceptual Confusion

It has been remarked earlier, that Wittgenstein identified the lack of an overview of the grammar of certain concepts and segments of our language as one of the main sources of philosophical confusion (see above). This "deficiency" makes us prone to fall victim to a variety of features of our natural language, which in the past have frequently entangled the understanding philosophers of as well as scientists. A variety of these sources of conceptual confusion have been identified[63]:

a) Analogies in the Surface Grammar of Language

Expressions like 'to have a car' or 'to have a house' look like the expression 'to have a mind'. Yet, while the former indicates some kind of ownership, the latter does not. And while 'to have a house', for example, denotes a concrete entity, i.e. 'a house', 'to have a mind' does not (except in the confused views of (neo-)Cartesian philosophers, and cognitive scientists (see chapter III for details). Similarly, activity verbs like 'to row', 'to sail', 'to play' and process verbs like 'to evolve', 'to grow', look like the psychological verbs 'to think', or 'to believe'. Yet, these superficial analogies gloss over the fundamental differences in the uses, and thus the meanings, of these verbs. The circumstances, under which thinking, for example, can correctly be said to be an activity or a process, are very narrowly defined by the rules for the use of the concept. Importantly, this does not include uses which allow for conclusions like 'thinking is a mental process', 'thinking is an activity of the mind' etc. as encountered in contemporary cognitive science. An investigation of the use of the concept of thinking shows, for example, that thinking is neither an activity of the mind nor a process going on in the mind. Neither the "mind", nor the brain meet the criteria that would warrant the ascription of thought in such cases (see chapter IV).

b) Pictures and Metaphors Embedded in Language

Numerous pictures and metaphors are tied to the wide-range of expressions, which are part of a natural language, and which are a natural part of everyday discourse as turn of phrase or *facon de parler*. However, in philosophical or scientific contexts they

[63] E.g., Hacker, `Insight and Illusion´, p.168ff.

are frequently misconstrued and understood figuratively, rather than being taken as the metaphors or synonyms they are. Thus, philosophers and cognitive scientists have been prone to interpret expression like 'I am out of my mind', 'or what is going on in his mind', as talk about an entity with an *inside* and an *outside*, for example. The prevalence of this *'Inner/Outer Picture of Mind',* which is prevalent in cognitive science, philosophy as well as popular culture is but one consequence of this misunderstanding (see chapter III for details). Similarly, Rodin's famous depiction of *Le Penseur* is partly the precipitate and partly invited the perpetuation of the picture of thinking as talking to oneself privately *in foro interno,* as it were (see chapter V). Expression of this kind, tempt us to believe that they are a genuine if picturesque description of reality, and that somehow the *facts* conform to them. Yet, as Bennett and Hacker point out, it is no conceptual confusion in calling the lower reaches of a mountain its foot*,* but there would be if one went looking for its shoe.

c) Projecting Grammar onto Reality

The tendency to project grammar onto reality is rooted in the conception that all names and all sentences are descriptions. For example, it seems natural to think of the statement 'nothing can be red and green all over' as a description of reality, like 'English bulldogs have a unique character an personality'. Likewise, one tends to conceive of a statement like '5 >1' as a *description* of these numbers. We take these statements as descriptions of reality, as certainties, and we are convinced of their truth. If distinct at all, these certainties are different merely in degree, with mathematical certainties enjoying the highest degree of certainty. Yet, this is mistaken, as there are different *kinds* of certainties, with different kinds of grounds (or none at all), and different kinds of consequences. In other cases, substantives make us look for a thing that corresponds to it (BB1). Traces of this particular sort of mistake can be found in the non-technical German term for substantive, *Namenwort*, which suggests the association of an entity or object with a word. This *picture of language*[64], thus has led philosophers and (today) also cognitive scientists to believe

[64] This 'Augustinian picture of language', which Wittgenstein describes and discusses in the opening section of the *Philosophical Investigations*, is generally regarded as an `Urbild´ or `proto-theory´ of the nature (Wesen) of language. Among the implicit assumptions of this picture are, e.g., the ideas that `describing´ and `ostensive´definition belong to the essence of language, or that to understand the `meaning´of a word is tantamount to knowing which object the word stands for. (see: PI§1,2), also Baker & Hacker (2005), Chapter I

that the word ‘mind’ denotes and entity of some kind, and that memories are stored as *neural engrams* in the brain mediated through neurophysiological mechanisms like long-term potentiation (LTP). It is not surprising then, that if one tends to think of substantives as names of objects, one will also tend to conceive of psychological expression as names of psychological objects, processes, states or events.

d) Projecting the Features of one Language Game onto Another

Another source of confusion is linked to the tendency of employing an expression belonging to one language game on analogy with the use that belongs to another, but which has no place or quite a different one. In such cases, to say that an expression is nonsense is tantamount to saying that this expression is not part of the language game to which it seems to belong (PG§81). Thus, one may hold for instance, that horses are larger than donkeys, that blue is darker than white, that it is true that either p or not –p, etc. Yet, we do not reflect on the differences in what it means to say, that it is true that. The truth of a proposition in mathematics is no more akin to the truth of an empirical proposition than a chess queen is akin to a queen[65] (see also: preceding paragraph). Because of the particularity or specificity of each language-game, each confusion requires independent treatment. Each misguided analogy must be investigated in its own right, and in each case a separate demonstration of non-conformity with the rules of the specific language-game in question is required.

Despite the crucial role of our grammar in the genesis of conceptual confusions, not all confusions are the result of misunderstandings of language *per se*. Human culture and human nature provide additional sources of conceptual confusion:

e) Scientific Theory and Explanation

> ***‘****Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does. This tendency is the real source of metaphysics, and leads the philosopher into complete darkness.’ (*BB_E,p.18)

---

[65] Hacker (1997a), p.170

The tremendous success which the natural sciences have enjoyed particularly in the twentieth century, have led many philosophers to think that in order for philosophical investigation to be successful, it must be modelled on the natural sciences[66] (see chapter I, also section 3 below). Like scientists, philosophers thus should make use of hypotheses and theories for the explanation of philosophical puzzles. However, this conception rests on the misunderstanding of the nature of science and philosophy, and neglects the important distinction between empirical and conceptual problems. While empirical problems are amenable to the tools and means of scientific investigation, conceptual problems cannot be solved by applying the hypothetico-deductive approach encountered in the natural sciences (see section 3 below). The consequence of this development is the obfuscation of the conceptual nature of philosophical problems, and the promotion of a speculative metaphysics which follows in the wake of the failure to distinguish factual and conceptual questions.

f) The Craving for Generality and Unity

Human beings display a craving for generality, i.e. a tendency to look for something in common to all the entities, which we commonly subsume under a general term. This also leads them to metaphysical speculation and consequently to transgression of the bounds of sense.

> 'We are inclined to think that there must be something in common to all games, say, and that this common property is the justification for applying the general term "game" to the various games; whereas games form a family the members of which have family likenesses. Some of them have the same nose, others the same eyebrows and others again the same way of walking; and these likenesses overlap. The idea of a general concept being a common property of its particular instances connects up with other primitive, too simple, ideas of the structure of language. It is comparable to the idea that properties are ingredients of the things which have the properties; e.g. that beauty is an ingredient of all beautiful things as alcohol is of beer and wine, and that we therefore could have pure beauty, unadulterated by anything that is beautiful.' ($BB_{E, p.18}$)

A historical example helps to illuminate this point. Plato, for instance, yielding to the craving for generality, thought that whenever we encounter one or more examples of

[66] Wittgenstein's contemporary and former mentor Bertrand Russell was a great champion of this idea.

a phenomenon X there must be something, which is implicit in all these examples and that entails the *Xness* of X. According to this principle, one might postulate beauty as underlying all beautiful things. In the same way, justice may be postulated as underlying all just acts. Plato considered the ideas of justice and beauty as the common nature or essence of all just acts and all beautiful things. According to Plato these universals cannot themselves exist in the world of sense as their instantiations, the particulars do. They are considered to be immutable and indestructible. This theory of ideas leads Plato to the assumption of his (metaphysical) realm of ideas in which these universals supposedly exist. The universals are supposed to enjoy existence independent of their respective particulars, rendering the distinction between universals and particulars a distinction between two things.

> `He (the metaphysician) moves from the supposition that there must be *something* in common to two examples of Xness, namely X itself, to the supposition that there must be some *thing* in common between them and, therefore, that X itself or Xness is a sort of *thing* other than those things which are examples of Xness.[67]´

In a similar manner, human nature's craving for unity, informs the search for ever more general theories. However, this may incline us to subsume a variety of distinctly different phenomena under one special case, as in the case of thinking for example (see chapter V).

3. Science and Philosophy

Throughout his career scientism in philosophy was Wittgenstein's *bête noire*[68]. Philosophers, Wittgenstein pointed out, constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does (BB18). Today, with few exceptions, the presence of scientistic currents within the mainstream of contemporary philosophy is stronger than ever, and has contributed considerably to the diminishing influence of Wittgensteinian ideas. This "triumph of scientism", however, is the result of misunderstanding the relationship of science and philosophy, and the power and limits of the respective disciplines by many philosophers and cognitive scientists. As a result, cognitive scientists tend to

[67] White (1989), p.12
[68] See e.g., Vossenkuhl, (1995) p.17f.

believe that scientific investigation can bring new insight to problems and questions where philosophical thinking faltered in the past. The champions of this view fail to acknowledge the crucial difference between factual and conceptual questions, and consequently are likely to end up under the spell of some illusion, which is the result of crude speculative metaphysics. The following discussion will substantiate this claim by juxtaposing the nature of empirical scientific investigations with Wittgenstein's latter conception of philosophy as a conceptual investigation.

Empirical Investigations and Conceptual Investigations

Today it is common to answer the question *'What is science?*', by immediately invoking the description of various experimental methods and their application like, PET, fMRI, PCR, single cell recordings, titration etc.. This tendency to identify science with scientific methodology and experimental technique implies the belief that it is sufficient to point at a few examples of scientific practice, in order to convey the nature of scientific investigation. The understanding of science, which emerges as a result, is one of a continuous series of experimental activities, which go on and on apparently oblivious to outside influences. Within the grip of this picture, scientific disciplines seem to distinguish themselves from each other only by the kind of experimental method applied (e.g. cognitive neuroscience may thus be conceived of distinguishing itself from physics only by the type of experiment conducted and the type of experimental method applied). However, such views neglect the fact that scientific practice cannot be reduced to experimental procedures. It fails to acknowledge the manifold explicit and implicit assumptions informing the design of experiments and the interpretation of experimental outcomes. Above all, it fails to recognize that science embodies a rich and intricate cultural tradition with a complex history. To be sure, science is a hypothetico-deductive endeavour, which proceeds by constructing and testing hypothesis through experiment. By means of hypothesis testing scientists proceed to the construction of theories, which describe, predict and explain the behaviour of the scientific phenomenon under investigation. Yet, as a result of the intricate interplay of culture and history of ideas, the interpretation of scientific results is also susceptible to influences that cannot be controlled by the most sophisticated and up to date experimental methods.

Human practices like scientific investigation are based on certain taken for granted principles and premises, which reflect the subject matter they occupy themselves with and the social contexts in which they take place. Such practices presuppose the ability to adopt and make us of a system (or systems) of concepts. The ability to use systems of concepts is integral to the practice of science, in so far as it allows for the definition and classification of the types and kinds of phenomena of interest to scientists. Thus, a taxonomy is created and an ordered catalogue of phenomena for investigation is established. This catalogue of phenomena, i.e. a catalogue of what is taken to be really real in the domain of scientific enquiry, constitutes the ontology of science[69]. An ontology will be among the presuppositions of a science at each moment in its development, and involves the postulation of the existence of observable entities like proteins, cells, cloud clusters, planets or galaxies, as well as of the existence of unobservable entities like rays, tectonic plates or inner mental representations, for example. By identifying and classifying phenomena, scientists form *conceptual presuppositions* about the meanings and interrelations of concepts. In addition, scientists also employ *factual presuppositions* to describe the behaviour, constitution, connection and interrelations of or between the various scientific phenomena. Importantly, while factual presuppositions can be tested by observation and experiment, conceptual presuppositions can only be evaluated by means of philosophical investigation, that is, by examining their plausibility, utility and coherence. For example, is it just a matter of fact, that I cannot feel your pain, or is it a matter o how the word pain is to be used in everyday language? If it is a matter of fact, it could have been otherwise. But, if it is a matter of the uses of words, one should not even be able to make sense of the alternative. In contrast to such conceptual matters, matters of fact can be adjudicated by observation and experiment. Thus, matters of fact are amenable to the kind of hypothetico-deductive form of investigation characteristic of science. Science, it can be said is the province of factual questions. Conceptual matters, on the other hand, do not relate to aspects of truth and falsehood. Rather, they relate to the rules for the

[69] Note: The ontology of a science is also revealed in the ways in which scientific theories are created and tested. Theory building and hypothesis testing are based on a system of models and metaphors, patterns of analogy through which concepts are modified and extended into new domains. The choice of models and metaphors involves presuppositions about what sorts of beings exist in the domain of a science. (E.g. *Inner mental representations*, which are ascribed a propositional structure, and which are encoded in the physical symbols system (of the brain) in a language of thought. *Volitions* as the source of willed action etc.)

correct use of concepts, and are adjudicated by an analysis of meaning. An analysis of meaning may reveal unnoticed confusions, contradictions or other faults in what may appear a coherent conceptual system. Such flaws can only be revealed by investigating the interrelations and ramifications among the meaning of concepts, and thus the rules for their use. Such investigations, as the later Wittgenstein pointed out are the province of philosophy. Philosophy is concerned with the analysis of the meaning of conceptual presuppositions. As the adjudicator of the bounds of sense, its concern is not with matters of truth and falsehood, but with matters of sense and nonsense.

The distinction between factual and conceptual presuppositions, and consequently between empirical and conceptual questions has important implications as to how the relationship between science and philosophy is to be understood. In contrast to the claims of philosophers past and present, philosophy is neither the *queen* nor the *under labourer* of science. Philosophy is not an investigation into the most general features of the universe, nor does it hand over its subject matter to scientists as soon as scientific technology has sufficiently advanced. There is no room for the application of scientific methods in philosophy. In contrast to the grandiose claims of some scientists, science will not be able to solve conceptual problems, as their very nature as questions of meaning does not render them amenable to empirical investigation. Yet during their training, scientists tend to acquire a picture of science, which leads them to believe that the majority of the problems encountered during the planning, implementing and publication of research is taken care of by sophisticated technology and continuing empirical investigation. For example, meaning is generally presumed to have been taken care of by the positivistic principle of giving an 'operational definition'. However, such an approach often leads scientists to focus on one use (i.e. meaning) of the concept of e.g., thinking or memory, which is only part of a field of (closely) related uses. This neglect results in a distortion of the polymorphous character of the phenomena under investigation (see e.g., Chapter V). Hence, it is important for scientists to realize the interplay of factual and conceptual issues influencing their work. Scientific understanding is articulated in the form of cognitive claims and hypotheses, which are expressed by means of propositions. As all such claims and hypotheses involve the use of concepts, they may therefore either be defective in point of truth or in point

of conceptual coherence, with the matter being an important precondition of the possibility of the former[70]. Philosophy and science should thus be understood as independent yet complementary forms of investigation, each of which can contribute in its own unique way to the genesis of insight and understanding.

4. The Difficulty of Philosophy

The difficulty of unravelling and dispelling conceptual confusions and to dissolve philosophical problems, is directly correlated to the degree of entanglement of our understanding:

> 'Die Philosophie löst die Knoten in unserem Denken auf die wir unsinniger Weise hinein gemacht haben; dazu muss sie aber ebenso komplizierte Bewegungen machen wie diese Knoten sind.' (PB§52)

> 'Philosophy unties the knots in our thinking, which we have tangled up in an absurd way; but to do that, it must make movements which are just as complicated as the knots. The complexity of philosophy is not in its matter, but in our tangled understanding.' (PR§52)

It is the bumps in his understanding, which the philosopher must suffer through whilst striving for that overview of the grammar of a philosophically problematic concept, and which will ultimately allow him to uncover a piece of nonsense, that will make him see the value of his discovery (PI§119). What makes the philosopher's task even more difficult is the fact that disentangling one's understanding is not only a difficult intellectual exercise, but also tends to take a heavy emotional toll on him. On the one hand, this is because philosophical problems irritate us, like a persistent itch or personal conflict with a parent or friend. On the other, this is because disentangling one's understanding means parting with ideas, which tend to be deeply engrained in our thinking, and which philosophers tend to cling to like the friendship of an old comrade. The philosopher's treatment of a question, Wittgenstein says, is like the treatment of an illness (PI§255). The notion of philosophy as a form of therapeutic endeavour is illuminating hear (see above), as the problem one faces, is comparable to the dilemma of the alcohol or drug addict. In both cases, the first stop towards salvation is the acknowledgment of the existence of a problem, but in order to arrive

[70] See Benett & Hacker (2001), p.3ff.

at this point usually a great deal of resistance and self-denial has to be overcome. In addition, the philosopher steeped in confusion and accomplished in the creation of metaphysical systems and theories, will find it very difficult to appreciate the power and significance of Wittgenstein's later philosophy, as this new way of thinking tends to be exceedingly hard to establish (CV§48). The idea to study the use of a concept and trying to gain an overview of its grammar, for example, tends to sound rather repulsive to the ears of both the mainstream philosopher and cognitive scientist. It seems deeply counterintuitive, and one may wonder why the understanding of the use of a word should be so important to us (RPPI§548). This problem gets even worse when the emphasis is shifted from the understanding and clarification of a philosophical problem, to such seemingly evident and apparently instinctively intelligible psychological phenomena as introspection and thinking. The fact that we can seemingly watch ourselves think or introspect creates the illusion of an intuitive understanding of these phenomena (PI§314), which stand in the way of acknowledging the existence of misconception and conceptual confusion.

Besides these "emotional" difficulties, which stand in the way of fruitful philosophical investigation, there are also general cultural challenges to overcome that obstruct the path to clarity of understanding. Our understanding of mind and substance like accounts of the soul for example, are closely linked to practices and beliefs rooted in the Christian religion. While soul and mind have always been related, they were welded into one by Descartes in the seventeenth century. To swim against the current of thought established by this great tradition is a formidable challenge. Wittgenstein was not unaware of the fact that his revolutionary ideas called for a radical reorientation with regard to the idea and practice of philosophy, and as such posed a substantial intellectual and emotional challenge. He once remarked that if his name will be remembered, then only as the *terminus ad quem* of Western philosophy, like the name of the one who burnt down the great library of Alexandria (MS183,63). In the context of this thesis, this remark foreshadows the intellectual and emotional challenges lying ahead, but also highlights the significance of this endeavour.

Chapter III
Crypto-Cartesianism, Brain-Body Dualism and Related Confusions – A Prelude

1. Crypto-Cartesianism and Brain-Body Dualism

At its most basic, the endeavour which cognitive science embarks on can be described as the attempt to explain and understand the *possession of psychological attributes by human beings*. In the present chapter the meaningfulness of explanatory statements depicting the role of the brain in the mediation of mental phenomena will be investigated. In doing so, it draws on earlier discussion of this issue by Bennett and Hacker (2003), and Kenny (1984, 1989). The form these statements take is the result of endorsing the cognitive view of mind and brain as a framework for the explanation of the mental (see Chapter I). A characteristic feature of these explanatory statements is the ascription of psychological predicates to the brain. As described in the previous chapter, cognitive scientists picture the brain as a computational device which receives, stores, retrieves, transforms and transmits information. Accordingly, cognitive scientists claim that the brain sees, hears, smells, believes, imagines, and even thinks, when explaining cognition:

> 'A hunk of biological tissue (the brain) is known to think, remember, attend, solve problems, want sex, play games, write novels, exhibit prejudice and do a zillion other things'.[71]

Among the *zillion other things* the brain or rather, a special sub-system within the brain, is supposed to be able to do is *recognizing faces,* for example[72]. Similarly, Francis Crick asserts that

> '…what you see is not really there; it is what your brain *believes* is there….'[73]

However, such claims are mistaken. As Wittgenstein points out,

> 'Es kommt darauf hinaus: man könne nur vom lebenden Menschen, und was ihm ähnlich ist (sich ähnlich benimmt) sagen, es habe Empfindungen; es sähe; sei blind; höre; sei taub; sei bei Bewußtsein; oder bewußtlos.' (PU §281)

[71] Gazzaniga, MS, Ivry RB, Mangun GR (1998), p.2f
[72] ibid. p.2
[73] Crick F, (1995), p.30

> 'It comes to this: Only of a human being and what resembles (behaves like) a living human being can one say: it has sensations, it sees; is blind; hears; is deaf; is conscious or unconscious.' (PI §281)[74]

The present chapter will develop the philosophical insight, which underlies this remark, and illuminate its consequences for explanatory statements in cognitive science. Wittgenstein's remark pinpoints the cardinal error, which cognitive scientists are frequently guilty of in their attempts to explain the possession of psychological attributes by human beings. They ascribe psychological predicates, which are only applicable to human beings or organisms that behave in sufficiently similar ways to human beings, to a part of that being: to the brain or to parts of the brain. In doing so, cognitive scientists violate the logic of whole/parthood relations and commit a *mereological error* [75]. The danger of mereological errors in cognitive science lies in the fact that they often conceal what is left to be explained. They lead us to thinking that we have arrived at a satisfactory explanation, when in fact all that has been achieved was to thwart the chance of obtaining a proper understanding of the mental phenomenon in question. In the following discussion, the tendency of cognitive scientists to ascribe psychological attributes to the brain and its parts will be referred to as the *mereological fallacy*.

---

[74] Note: In *De Anima* Aristotle also urges us *'not to say that the soul pities or learns or thinks but that the man does in virtue of his soul* (DA 408b13-15).*'*

[75] Mereology (from the Greek μερος, 'part') describes the relations of part to whole and the relations of part to part within a whole. Its roots can be traced back to the early days of philosophy, beginning with the pre-socratic atomists and continuing throughout the writings of Plato (especially the *Parmenides* and the *Thaetetus*), Aristotle (especially the *Metaphysics*, but also the *Physics*, the *Topics*, and *De partibus animalium*), and Boethius (especially *In Ciceronis Topica*). Mereology has also occupied a prominent role in the writings of medieval ontologists and scholastic philosophers such as Garland the Computist, Peter Abelard, Thomas Aquinas, Raymond Lull, and Albert of Saxony, as well as in Jungius's *Logica Hamburgensis* (1638), Leibniz's *Dissertatio de arte combinatoria* (1666) and *Monadology* (1714), and Kant's early writings (the *Gedanken* of 1747 and the *Monadologia physica* of 1756). As a formal theory of parthood relations, however, mereology made its way into modern philosophy mainly through the work of Franz Brentano and of his pupils, especially Husserl's third *Logical Investigation* (1901). The latter may rightly be considered the first attempt at a rigorous formulation of the theory, though in a format that makes it difficult to disentagle the analysis of mereological concepts from that of other ontologically relevant notions (such as the relation of ontological dependence). It is not until Leśniewski's *Foundations of a General Theory of Manifolds* (1916, in Polish) that the pure theory of part-relations as we know it today was given an exact formulation. Because Leśniewski's work was largely inaccessible to non-speakers of Polish, it is only with the publication of Leonard and Goodman's *The Calculus of Individuals* (1940) that this theory has become a chapter of central interest for modern ontologists and metaphysicians. Contemporary formulations of mereology grew out of the recent theories of -- Leśniewski's and Leonard and Goodman's. Indeed, although such theories came in different logical guises, they are sufficiently similar to be recognized as a common basis for most subsequent developments.

A historical example helps to illustrate how the mereological fallacy can distort our understanding. In his book *The Legacy of Wittgenstein* Anthony Kenny discusses the consequences of mereological errors in the context of Descartes´ writings:

> 'If we see some animal approach us, the light reflected from its body depicts two images of it, one in each of our eyes, and these two images form two others, by means of the optic nerves, in the interior surface of the brain which faces its cavities; then from there, by means of the animals spirits with which its cavities are filled, these images so radiate towards the little gland which is surrounded by these spirits, that the movement which forms each point of one of the images tends towards the same point of the other image which represents the same part of this animal. By this means the two images which are in the brain form but one upon the gland, which, acting immediately upon the soul, causes it to see the form of this animal.'[76]

Kenny, using the term *homunculus fallacy* to describe the conceptual issue surrounding whole/parthood relationships, points out that Descartes commits the homunculus fallacy (i.e. a mereological error) when he claims that the soul encounters images in the pineal gland. This is due to the fact that according to Descartes a soul is no more a human being than a brain is. While it is not philosophically flawed to speak of images in the brain *per se* (e.g. it would not be philosophically objectionable to say that a neurologist might observe these images in the brain whilst investigating the pineal gland), it is logically incorrect and misleading to suggest that there are images in the brain which are visible to the soul, and that the perception of those images, appearing on the pineal gland, constitutes seeing (i.e. is a satisfactory explanation of human vision). Seeing is *not* constituted by the souls´s perception of images on the pineal gland, but by the various forms of behaviour we exercise or are engaged in when we read, walk around, drive a car, watch television or numerous other things that involve having a look around, stare at a screen, or gaze at the daffodils whilst we "*float on high o´er vales and hills…"*. By claiming that seeing constitutes the perception of images on the pineal gland by the soul, our understanding of the nature of seeing, i.e. what seeing really consists in, gets distorted. Henceforth, one labours under the illusion of understanding the nature of vision, as a conceptual confusion has been cast in the form of an empirical discovery. Descartes' claim constitutes a conceptual confusion, because if it was an empirical discovery that there are images in the brain, which are visible to the soul, then one should in principle be able to conduct experiments to further investigate this

[76] Descartes R, '*The Passions of the Soul*' , in Cottingham, Stoothoff, Murdoch (eds), 'The Philosophical Writings of Descartes', Vol.I, p.341f

observation. But exactly this is not possible, because we do not know *what it would be like* for a brain to see (or to be blind for that matter). Unless the meaning of expressions like the soul (or brain) perceives, beliefs, thinks etc. has been explained, i.e. unless one knows what *counts* as the soul (or brain) doing thus and so, and unless one has established what could be cited as *evidence* for the soul (or brain) doing thus and one is unable to investigate this "phenomenon" experimentally. From a modern perspective, one might offer as a possible correction of the Cartesian claim a description of human vision of the following kind: 'A person is *able* to see, because the light falling on the retina causes the photo pigments in the membranes of the retinal cones and rods to change their conformation. This leads to the activation of various ion channels. The nerve impulses generated as a consequence of this activity are transmitted to the primary visual cortex etc.'

The fundamental flaw in Descartes´ explanation is that the very same problems arise about Descartes´ *explanans,* as exist with regard to his *explanandum*, and which prompt an explanation in the first place. Because the meaning of expressions like the soul perceives, beliefs, thinks etc. has not been stipulated (i.e. there exist no rules for the use of such expressions), for example, it has not been specified what counts as a soul doing thus and so, an explanation of the explanandum is still to be given. Thus far, however, nothing has been explained (see also 1.2. below). The following discussion will further clarify that the brain is not a logically appropriate subject for psychological predicates. Wittgenstein's remark in PI§281 is meant to direct our attention to the rules and criteria which underlie the application of psychological predicates. It is a grammatical (normative) statement pointing out a potential conceptual issue, by making a claim about the correct logical application of certain predicates. In order to appreciate the weight and significance of Wittgenstein's insight, however, it is necessary to gain an understanding of the causes which underlie the modern mereological error. Furthermore, in order to expose the inconsistency underlying the mereological fallacy convincingly it is indispensable to examine the grounds which warrant the application of psychological predicates to human beings.

1.1 The Causes Underlying the Persistence of Mereological Confusion in Modern Cognitive Science

In Descartes´ case described above, it is easy to discover and understand the gaps in his account. This is mainly due to the fact that his physiological descriptions have been shown to be inadequate. Yet, as the examples of modern cognitive science illustrate, this particular philosophical error can happily coexist with the most detailed of physiological descriptions, and the most sophisticated items of scientific information. What is the reason for this? First, it seems that this is partly due to the fact that cognitive scientists are unaware of the implications of the Cartesian ideas which infuse their chosen field of study. It is not a matter of rational consideration that the application of psychological predicates has been shifted from an immaterial mind or soul to a material brain, but the residual effect which the Cartesian legacy exerts on cognitive science (see Chapter I & II). Despite the fact, however, that this legacy is more often than not openly acknowledged, the impact and lasting influence of the flawed Cartesian conceptual framework which cognitive science endorses remains neglected and misunderstood. Cognitive scientists appear to underestimate the subliminal ways in which the Cartesian conceptual framework they work with influences their scientific arguments and moulds their explanations and understanding. Although the vast majority of cognitive scientists vehemently reject the dualist conception of body and mind championed by Descartes, the *partial* repudiation of the Cartesian legacy does not mean that other aspects, which still exercise their lasting influence, have been equally recognized and repudiated. One of the reasons for such carelessness may lie in a false belief in the powers of modern science. The repudiation of Cartesian metaphysics often follows in the wake of an emphatic reference to scientific progress and the ever-increasing power and sophistication of scientific technology. The conviction that with the advancement of scientific technology and resulting scientific progress philosophical problems are solved quasi automatically, is spread widely among cognitive scientists and those philosophers, which champion the application of the scientific method in philosophy[77]. However, this view rests upon a profound misunderstanding of both the nature of science and the nature of philosophy (see Chapter II). It reflects the ideologization which science has undergone, particularly in the 20th century. One of the symptoms of this affliction is the fact that the claims of (cognitive) scientists are more often than

[77] See, for example, Dennett (1991)

not unjustifiably scientistic, and vastly exaggerate the powers of science. Eric Kandel, winner of the Nobel prize for medicine in 2000, and his colleagues are just one prominent example who exhibit this misguided belief in the powers of science whilst defining the goal of cognitive neuroscience as the examination of

> '… classical philosophical and psychological questions about mental functions in the light of cell and molecular biology.'[78]

However, views like this neglect that no flawed philosophical theory or question can be corrected or answered by applying the methods of science. They betoken an ignorance of the distinction between factual and conceptual questions, and a general disregard for conceptual clarity. A more outspoken example of this kind of contemporary scientism is provided by the famous Oxford chemist and science writer Peter Atkins, who seems rather keen to underline the "omnipotence" of science:

> 'Although poets may aspire to understanding, their talents are more akin to entertaining self-deception. They may be able to emphasize delights in the world, but they are deluded if they and their admirers believe that their identification of the delights and their use of poignant language are enough for comprehension. Philosophers too, I am afraid, have contributed to the understanding of the universe little more than poets…They have not contributed much that is novel until after novelty has been discovered by scientist…while poetry (and philosophy) titillates and theology obfuscates, science liberates.'[79]

It is the scientistic attitude underlying those statements, which is rife within the community of cognitive scientists, and which is partly responsible for the fact that so many cognitive scientists unthinkingly adhere to a mutant form of Cartesian metaphysics, and unwittingly trust in technological progress whilst neglecting the need for conceptual clarity and hygiene.

The second reason why cognitive scientists are so tempted to ascribe psychological predicates to the brain is, in philosophical terms, much more interesting and of much greater significance. It gives us an insight into the ways in which "*the philosopher in us"* works and how we fall victim to our "*metaphysical disposition"* (see Chapter II), whilst struggling for an understanding of the human mind. Philosophically, it seems that the tendency to commit a mereological error is linked to

[78] Kandel ER, Schwart JH, Jessel TM, (2000), p.313
[79] Atkins (1995), p.123

a misunderstanding of the differences, which exist between the ascription of psychological attributes to oneself and to others. While the ascription of psychological attributes to others is justified by reference to behaviour, the ascription of a psychological predicate to oneself is not. For example, someone's saying that one is in pain does not depend on the observation of one's own pain behaviour. If that is the case, how is it that we can say what we see, hear, feel and think or belief? It is a widespread belief, not only harboured by scientists, that the mental is a special private domain accessible only to its subject. We are inclined to think of this private domain as an *inner mental realm* and we speak accordingly of our *inner mental life*, or simply *the inner*. In contrast, behaviour, observable by anyone appropriately situated, belongs to the public domain. The public domain is conceived of as *the outer*. According to this *Inner/Outer picture* or *metaphor*[80] we have direct privileged access to our own mind, to the mental events and processes going on *in our heads,* by ways of introspection (and thus without having to observe what we do or say). Other people, on the other hand, depend on behavioural evidence in order to find out what someone thinks or how someone feels. It seems that they can find out how things are with me only indirectly. Due to a misunderstanding of the faculty of introspection we assume, as Bennett and Hacker point out that:

> '…the *access* which introspection gives each person to his own mental states and processes is a partial and limited access to processes going on in his brain[81]'.

These assumptions do not represent the latest thinking currently *en vogue* in contemporary cognitive science. Rather, they have already been shared by both Cartesian and empiricist philosophers. The difference between the *classical* and modern position(s) lies merely in the fact that both Cartesians and empiricists implied a substance dualism, and consequently assumed that introspection provides us with direct and immediate access to our own immaterial mind, rather than the going ons in our brains. These ideas are crucially linked to the representationalism endorsed by these schools of thought, and which continue to be applied by contemporary cognitive science in the context of cognitive view (see Chapter II). Cognitive science thus replaces the Cartesian dualism of mind and body, with a dualism of brain and

[80] As Gilbert Ryle points out, 'this antithesis of outer and inner is of course meant to be construed as a metaphor, since minds, not being in space, could not be described as being spatially inside anything else, or as having things going on spatially inside themselves.', *in,* 'The Concept of Mind'*, London, Penguin Books, 1994, p.14*

[81] Bennett & Hacker (2003), *p.85*

body. As a result, the mereological error persists in modern cognitive science, despite the application of highly advanced and sophisticated technology. The immaterial mind of Cartesian dualism is merely substituted with the material brain, while the logical character of dualism is essentially retained in the conceptual framework endorsed by cognitive scientists.

Of course, this claim may initially be somewhat difficult to get to grips with. After all, both body and brain are material substances and the suggestion that dualism is retained is not immediately intelligible and may seem rather absurd. One is inclined to point out that body and brain are not two distinct kinds of substance in the same way that mind and body are within the context of Cartesian metaphysics. Yet, the range of psychological properties cognitive scientists ascribe to the brain are the same characteristics that Cartesians ascribe to the mind. But psychological predicates are no more applicable to an immaterial mind, than they are applicable to the brain or parts of the brain. Cognitive science thus applies a conceptual framework to the brain that is roughly identical to that of Cartesian metaphysics. Whilst the immaterial Cartesian mind is replaced by the material brain, the logical structure of dualist psychology is retained. Consequently, cognitive scientists are prone to *explain* the possession of psychological attributes by a human being by ascribing those attributes to the human brain. However, the fact that the brain rather than Descartes' immaterial soul is now the subject to which these predicates are ascribed does not change the fact that the *explanandum* is still looking for its *explanans*. As long as this crypto-Cartesianism persists, the endeavour of cognitive science will suffer from conceptual confusion, a fundamental flaw, which obscures the fact that no real explanation of the mental phenomenon in question has been achieved. The persistence of crypto-Cartesianism in contemporary cognitive science thus stands in the way of a correct understanding of the mental phenomena cognitive scientists strive to illuminate. Under the influence of the Cartesian legacy, the modern explanations of mental phenomena provided by cognitive scientist are, although much more sophisticated, just as flawed as the much older explanations offered by Descartes. Like the explanations of their Cartesian ancestors, the modern versions also constitute transgressions of the bounds of sense (see below).

1.2 Substantiating the Mereological Principle: The Grounds for the Ascription of Psychological Predicates

The point that Wittgenstein makes in PI§281 is not a factual one, but a conceptual one. Whether psychological attributes can intelligibly be ascribed to the brain is a philosophical and therefore a conceptual question. It is not a question, which can be answered by the means of scientific investigation. It is not factually wrong that brains don't see, hear, believe, desire, remember or think. Rather, it constitutes a *transgression of the bounds of sense* to say such at thing, save in a metaphorical or metonymical sense. We do not know what it is for a brain to see, hear, believe, desire, remember, or think. Neither do we know what it would be for a brain to be blind or to be deaf, or to be content, or to be thoughtless. The issue addressed here calls for conceptual clarification not for empirical investigation as the claim that the brain thinks, hears, sees remembers etc. is not based on the *observation of (newly discovered) facts*. Rather such claims are the product of a particular line of thought, which cognitive scientists adopt in the *interpretation* of their experimental results. And until one knows what it would be for a brain to do or not to do so, i.e. until we know what sort of evidence would support the ascriptions of such attributes to the brain, we are not clear about the meaning of such statements. Any interpretation or explanation, which involves the ascription of psychological attributes to the brain, is flawed, and will not contribute to advancing our understanding. Consequently, the conception of the brain as a form of biological information processing unit or biological computer which receives, stores, retrieves, transforms and transmits information, leading to such statements as the brain beliefs, thinks, sees etc, (see Chapter II, Sec.2) constitutes as transgression of the bounds of sense.

The *logical grounds* for the ascription of psychological attributes to a subject are the behaviour of the subject, of the human being, but not of his brain or any other part of his body. Nothing a brain does could possibly constitute a ground for ascribing vision, hearing etc. to the brain. The brain is not a logically appropriate subject for the ascription of psychological attributes. This follows from the mereological principle (the principle that psychological predicates which only apply to human beings as a whole cannot be applied to parts of a human being) on which these claims are based. The mereological principle is best understood as a convention, rule or norm, which determines what does and does not make sense. As a convention the mereological principle could in theory be changed. Yet, in order to implement such a change a,

large number of other things would have to change too. The mereological principle is secured by a widely ramified network of conceptual connections manifest in a network of rules, which forbid the application of psychological concepts to parts of human beings. To change the application of the mereological principle would therefore necessitate the reformulation and rearrangement of the rules which underlie these conceptual connections, and which form an integral part of the intellectual landscape of our socio-cultural environment. In other words, such rearrangement and reformulation would lead to the changing of the meanings of numerous familiar concepts, which constitute a fundamental part of our form of life (see Chapter 2). In fact, we could not resolve to abandon this normative relationship without instigating changes in our form of life.

The significance of the mereological principle thus arises out of the conceptual commitments or rules relating to the application of expressions, which describe and explain the psychology of human beings. An examination of these rules by studying language games using psychological expression reveals that the applicability of psychological predicates to others is regulated and warranted by their behaviour. For example, if I hear my mother, who is just bustling about in the kitchen, suddenly scream ‘Ouch!’, and see her assuaging her head or hand I know immediately that she is in pain. Her pain behaviour, i.e. her scream, warrants my immediate pain-ascription to her. Importantly, I ascribe pain ‘to my *mother*’, not to her head or her hand. I don’t think (or say) that her head or her hand is in pain. Similarly, if by rushing out of the door I accidentally trap Giacomo’s (i.e. my bulldog’s) paw in the door, I think that I accidentally hurt him and that he is in pain because of his wails. I do not think (or say) that his paw is in pain. It is my bulldog Giacomo who is in pain. In the same way, I would not think that my teacup is in pain, after I accidentally dropped it and its handle broke off. Neither would I think (or say) that my armchair beliefs that he is going to be up next for bulk rubbish collection, because I just dispensed of my sofa in this way. Neither the ascription of pain to my tea cup, nor the attribution of a certain belief to my arm chair makes sense, as the mereological principal forbids the application of such psychological predicates to something other than human beings or anything that behaves like a human being. Both expressions constitute transgressions of the bounds of sense. At this point, however, one may be tempted to argue that in certain stories and fairytales, for example, psychological predicates are frequently meaningfully applied to things like stones, plants and other usually

inanimate objects, and that in these cases these ascriptions occur well within the bounds of sense. Thus, the *Ents*[82] encountered in *The Lord of the Rings* rose against the wizard Saruman because they believed that he was in league with the evil Sauron. Likewise, the little teacup "Tassilo" encountered in Walt Disney's *The Beauty and the Beast* rubs his nose, i.e. his handle, in pain after the candelabra "Lumiere" accidentally dropped some wax on it[83]. While it is true that in these cases the bounds of sense have not been transgressed, it is important to note that this is because Tassilo, by rubbing his nose/handle behaves just like a human being would. Likewise, the Ents display a characteristic human behaviour when they learn about Saruman's alliance with Sauron. They voice their indignation and call for action, also a characteristically human behaviour. In such cases the bounds of sense are not transgressed because the ascription of belief and pain is modelled on the pain behaviour of human beings. Both Tassilo and the Ents display the behavioural criteria, which warrant the ascription of pain and belief. The mereological principle is (implicitly) acknowledged in these cases. It is the different forms of behaviour displayed by human beings, which provide the primary ground for the ascription of psychological attributes to others. Thus, when cognitive scientists say that the brain sees, hears or thinks etc. the bounds of sense are transgressed because the brain, like my mum's head or hand, and like Giacomo's paw, (and unlike Tassilo or the Ents) does not exhibit the behavioural criteria which would justify the application of these predicates. To say that the brain sees, hears or thinks makes as much sense as to say that the brain is blind, is deaf or behaves thoughtlessly. Thus, when Antonio

[82] In J.R.R. Tolkien's story the *Ents* are described as tree-herders, i.e. guardians of the trees of the forests of Middle Earth and essentially trees themselves.

[83] To the left, the candelabra *Lumiere;* to the right, *Tassilo* with his mother, the teapot called *Mrs Potts*.

Damasio tells us that core consciousness[84] is the result of first order neural representations telling, i.e. "*narrating",* a story which is perceived/heard by (higher) second order neural representations about, '*how the organism's own state is affected by the organism's processing of an object*[85]' , one may feel inclined to question the intelligibility of such an account and press for clarification. Similarly, if someone argues that schizophrenics experience hallucinations because '*the right hemisphere is talking to the left hemisphere*[86]' one should be worried about the degree of insight prevailing in modern neuropsychiatry.

It is also important to note, that contrary to the beliefs of some philosophers the ascription of psychological predicates to others is neither based on inference nor on inductive evidence[87]. The ascription of pain described above, for example, is not inferred. Inductive evidence is gathered by establishing the co-occurrence of two sets of independently identifiable events. Thus, it pre-supposes a non-inductive identification of the phenomena that are observed to stand in correlation to one another. Consequently, to understand the notion of some particular mental state, one has to be familiar with the kind of behaviour that counts as evidence for its occurrence. Pain behaviour is logically good evidence for being in pain. As Wittgenstein points out:

> 'Ein 'innerer Vorgang' bedarf äußerer Kriterien.' (PU§580)
>
> 'An 'inner' process stands in need of outward criteria.' (PI §580)

But, one may want to ask, what kind of behaviour does the brain display when it beliefs something for example? What are the behavioural criteria, which the brain exhibits when it perceives, believes, thinks or does the myriad of other things that human beings are usually engaged in? What are the outward criteria, which indicate

[84] Damasio distinguishes between proto-self (a coherent collection of neural patterns which map, moment by moment, the state of physical structure of the organism in its many dimensions (~state of the organism in homeostasis)), core consciousness, (a sense of self only in the here and now), and extended consciousness (a sense of self spread out over time, which adds a historical dimension to our (core) sense of self). (See: Damasio 1999, Chapters 5 & 6)

[85] Damasio (1999), p.170

[86] Frith (1999), p.27

[87] Bertrand Russell, for example, holds that 'When I see the sun, I am often aware of my seeing the sun; thus 'my seeing the sun' is an object with which I have acquaintance. When I desire food, I may be aware of my desire for food; thus 'my desiring food' is an object with which I am acquainted. Similarly we may be aware of our feeling pleasure or pain, and generally of the events, which happen, in our minds. This kind of acquaintance, which may be called self-consciousness, is the source of all our knowledge of mental things. It is obvious that it is *only* what goes on in our minds that can be thus known immediately.', *in, 'The Problems of Philosophy', Oxford, Oxford University Press, 1989, p.26f.*

the brains' believing or not believing something? Like the mereological principle described in PI§281, this remark indicates a *rule of grammar, which* regulates the use of psychological predicates. (Pain) behaviour is best thought of as a *criterion*[88] for the ascription of pain to a subject. Put in more general terms we can also say that the physical expression of a mental process is a criterion for that process (i.e. the meaning of a psychological predicate is partly constituted by a characteristic manifestation of a typical form of behaviour).

> 'Criteria, unlike symptoms (inductive correlations), determine the meanings of expressions for which they are criteria.'[89]

However, to say that the various forms of behaviour are criteria for the ascription of psychological predicates is not to say that the meaning of a psychological predicate is equivalent to the behaviour that warrants its ascription. People can be in pain and not show it. They can also think or believe something and not exhibit it or not admit that they are. And it is possible for people to pretend to be in a certain psychological state. It is possible to pretend to be in pain by exhibiting pain behaviour yet not to be in pain at all (a form of behaviour often displayed by centre forwards in the goal zone). That is, any criterial evidence for the ascription of psychological attributes to others is capable of being invalidated by evidence to the contrary (i.e. a centre forward may pretend to be in pain after having been in contact with an opponent in order to secure a penalty. After the referee has blown the whistle and decided to award or not award a penalty the centre forward who had been pretending to be in pain more often than not "miraculously" recovers form his injury in an instant).

---

[88] Anthony Kenny explains that Wittgenstein argued that it is necessary to distinguish between two kinds of *evidence*, which one may have for the occurrence of states or affairs. Wittgenstein thus comes to distinguish between *criteria* and *symptoms*. A piece of evidence may be called a symptom of a particular state of affairs if the connection between a certain kind of evidence and the conclusion drawn from it is a matter of empirical discovery (through the application of theory and induction). Yet, where the relation between evidence and conclusion is not something discovered by empirical evidence, but is something that must be grasped by anyone that is in possession of the concept in question, then the evidence is not a mere symptom, but is a criterion of the state of affairs in question. Kenny gives the following example: 'A red sky at night may be a symptom of good weather the following morning; but the absence of clouds, the shining sun, etc., tomorrow are not just symptoms but criteria for good weather.' (see Kenny 1989, p.5)

[89] Hacker, ‚Wittgenstein – Meaning and Mind' (Part II), p.250

2. Aspects of the *Inner* and the *Outer* Investigated: Introspection, Direct and Indirect Access, Privacy and Private Ostensive Definition

One of the reasons which has been identified above as underlying the tendency of both philosophers and cognitive scientists to ascribe psychological attributes to the brain was their inclination to conceive of expressions of thought and experience as descriptions of one's own inner mental states. Like their Cartesian ancestors[90] they hold that psychology studies the human mind from the outside, i.e. by investigating human behaviour. The main difference to the thinking of Cartesians of previous generations is, however, that contemporary cognitive scientists and philosophers belief that with the advancement of scientific technology in the late 20th century, brain imaging in particular, one is finally able to find out what lies behind the outside (i.e. the behaviour) – the workings of the human brain. The well known British neuroscientist and director of the Royal Institution, Baroness Susan Greenfield, famously explained to a TV audience that as a result of modern brain imaging techniques, scientists can '…see the brain think'. A corollary of the conception of expression of thoughts and experiences as descriptions of observations made *in foro interno*, as it were, is the claim that we have *direct and privileged access* to our mental states by means of *introspection*. Conversely, it is claimed that we have only *indirect access* to other people's states of mind, thoughts and experiences. The 'inner' appears to be hidden behind the 'outer'; what goes on inside is to be *inferred* by observing someone's behaviour. In other words, we are prone to think that when we ascribe psychological predicates to others we can only do so on the basis of indirect behavioural evidence. However, this conception manifests a deep confusion regarding the nature of first- and third-person psychological propositions.

The aim of this section is to closely scrutinize this picture of a private *Inner* realm of mind and an *Outer* public realm of behaviour. By showing that this picture is the result of a misconstural of various figures of speech such as natural metaphors and similes, which crop up naturally in every day discourse, I hope to provide a *prophylactic* against its seductive simplicity. However, an investigation into the muddles which provide the foundation for the *Inner/Outer picture* of the mind is also crucial to obtaining a deeper understanding of the confusions underlying the

[90] In 20th century neuroscience the Oxford nobel laureates Charles Sherrington and his pupil John Eccles were prominent champions of Cartesian substance dualism.

mereological fallacy. Both the Inner/Outer picture and the mereological fallacy are a manifestation of the crypto-Cartesianism persisting in cognitive science. Together they provide an example of the intricate links that exist between the analogies, pictures and metaphors that often mislead us when we contemplate the nature of the human mind. As Wittgenstein put it:

> 'Der besonderen Täuschung, die hier gemeint ist, schließen sich, von verschiedenen Seiten, andere an. ...' (PU§96)

> 'Other illusions come from various quarters to attach themselves to the special one spoken of here. …' (PI §96)

The contemporary picture of a private inner realm to which subjects have privileged access to, and of a public outer realm which is immediately accessible for everyone, is of partly Cartesian and of partly empiricist origin[91] (see also Chapter II). This view is problematic as it invites a misunderstanding and misconstural of the logical character of experience and its ascription. The difficulty in freeing ourselves from this picture lies in the fact that it is so deeply embedded in our language and thus appeals to our most basic intuitions regarding our views of mind and the world (PI §115). It is equally appealing to the philosopher, as it is to the laymen contemplating the intricacies of human psychology. This aspect will become clear as the philosophical investigation into the Inner/Outer picture progresses and we try to disentangle the knots in our understanding. Some of the claims we are going to encounter in the following discussion may strike us as deeply counterintuitive and as having no obvious rationale. But as will also become clear, this is because we tend to be stuck in viewing philosophical problems and questions from a theoretical , speculative perspective rather than the more appropriate grammatical one. If one adopts a grammatical perspective, however, we come to realize that, despite appearances to the contrary, some of the claims and ideas proposed by philosophers and cognitive scientists are not genuinely intelligible. But as Wittgenstein has

[91] It is important to note, however, that the empiricist adoption of this picture is a result of the influence of Descartes´ ideas outside the Cartesian tradition proper. For Descartes and his followers ideas, impressions and sense-data are all (private) mental entities. For the British empiricists these mental entities are (similar to the Cartesian tradition) epistemologically prior to the physical substances which are part of the 'external' or 'outside' world. Like Descartes, Locke, Berkeley and Hume postulated that the mind is better known than the body (i.e., the internal is more certain than the external; the private is prior to and better known than the public). (See for example: Kenny 1973, p.114)

pointed out, one learns to recognize when the bounds of sense have been transgressed not least by the bumps one's understanding gets by running its head up against the limits of language (PI §119).

2.1 The Metaphor Exposed and Examined

The Inner/Outer picture is based on a number of highly questionable assumptions, which take us hostage while leading to misconceived assumptions about the character of experience. Some of them have already been encountered *en passent* in our discussion of the mereological fallacy. For clarity's sake it is worth, however to restate them at this point of the investigation. The importance of the investigation into the Inner/Outer picture is not least due to the fact that some of the misconceptions associated with this picture, i.e. misconceptions about the meaning of psychological predicates, would undermine our claim that it makes no sense to ascribe psychological predicates to brains if they stand uncorrected (see below). The following assumptions lie at its core:

a) The mind is a *private inner realm* that stands in contrast to an outer public realm of material (physical) phenomena and behaviour. This view is exemplified by claims like that of Antonio Damasio who in the tradition of William James[92] views consciousness as an

> 'entirely private, first-person phenomenon which occurs as part of the private, first-person process we call mind.'[93]

b) Introspection is a means to get access to the private domain, i.e to what goes on in the 'theatre of mind', as it were. It is a means for *perceiving* our mental going ons.

c) The inner is *directly accessible* only by oneself, whereas others have only *indirect access* to it. Emphatically propounding a Cartesian position, Francis Crick remarks

[92] James in emphasizing the Inner/Outer dichotomy implicitly conceives of the mental as a private inner realm: `*…everyone admits the entire incommensurability of feeling as such with material motion as such. "A motion became a feeling!" – no phrase that our lips can frame is so devoid of apprehensible meaning. Accordingly, even the vaguest of evolutionary enthusiasts, when deliberately comparing material with mental facts, have been as forward as anyone else to emphasize the `chasm´ between the inner and the outer worlds.´*[92] *(see:* James W 1890, `The Principles of Psychology´, p.149)

[93] Damasio (1999), 'The Feeling of What Happens' , London, Heinemann, p.12

> '(that) strictly speaking, each individual is certain only that he himself is conscious.'[94]

d) The meaning of psychological predicates can be grasped independently of any conceptual connection to behavioural criteria. They are names of inner entities, mental experiences like states, events, processes etc. We know what psychological predicates mean through *private ostensive definitions*, that is, by associating them with the experiences we have and to which we have privileged access. This assumption has not only been tacitly presupposed in the writings of a variety of philosophers since Descartes and the British empiricists, but is also presupposed in the modern representationalist ideas underlying the cognitive view. Fodor's ideas on a language of thought, as well as Chomsky's notion of a "language acquisition device", for example, exemplify how the 'privacy assumption' influences contemporary thinking in cognitive science.

The outline presented above identifies three closely connected misconceived dualities, i.e. an inner-outer duality, a direct-indirect duality and a public-private duality -, as providing the core our conception of the mental. They are deeply embedded in the ways philosophers, cognitive scientists as well as laymen talk about the mind and psychological phenomena. In addition, they are closely linked to misconceptions regarding the nature of introspection and the meaning of psychological predicates which have attached themselves to these dualities, and which necessitate philosophical investigation. The following will investigate each one of these misconceptions in turn.

The Inner-Outer Duality

Antonio Damasio provides an illustrative example of the persistent influence which the inner-outer duality exerts on contemporary cognitive science. In *The Feeling of What Happens* , for example, he postulates that

> 'the term feeling should be reserved for the private, mental experience of an emotion…this means that you cannot observe a feeling in someone else, although you can observe a feeling in yourself when, as a conscious being you perceive your own emotional states.'[95]

[94] Crick (1995), 'The Astonishing Hypothesis', London, Touchstone, p. 107
[95] Damasio (1999), p.42

It is immediately obvious that Damasio's claim is an echo of the voices of his Cartesian ancestors. The assumption that one can only observe a feeling in onself, but not in others, is not only a clear allusion to both the inside-outside duality but also to the direct–indirect duality[96]. It carries the clear implication that when we talk of something like feeling we speak of somebody else's feelings we talk about something which lies *behind* the behaviour. Although the sense of the attributions of psychological predicates to third parties appears, at first glance, unproblematic, philosophical problems seem to arise with respect to their truth (in particular, when compared with first person ascriptions of that kind). But this is where Damasio and other cognitive scientists go wrong, as they 'look for an inside behind that which in our metaphor is the inside (NFL, p.280) .' In order to clarify the insight underlying this statement it is of paramount importance to gain an understanding of how the metaphor of *inner* and *outer* is applied (NFL, p.280).

A first step in getting an understanding of the application of this picture lies in answering the question under which *circumstances* it is employed? Both Anthony Kenny and Peter Hacker point out that it is highly conspicuous that with the exception of neuroscientific and philosophical discourse nobody would claim that a feeling like pain, for example, is something mental[97]. In fact, quite the opposite is the case. We routinely invoke the notion of physical pain, whilst referring to headaches, broken legs or arms, or stomach or toothaches. Talk of mental pain on the other hand, if invoked at all, refers to states of mind which are characterised by feelings of grief, sadness or depression. The verbal expression of such states of mind often involves figures of speech like, he is a broken man, or he suffers from a broken heart. But things start to become muddled, when we project the relation between entertaining a feeling or an experience, i.e. when experiencing some form of pain like having a toothache, for example, and behaviour (e.g. assuaging one's jaw), onto the picture of the inner and the outer (See e.g.: LSD, p.118). In itself this comparison is natural and even harmless. After all,

> 'We say of another that he has toothache on the basis of observing his behaviour (something external), but he says I have a toothache without observing his own behaviour.' (NFL, p.278)

[96] Note: The confusion of the inner-outer runs a misuse of the direct and indirect. (See e.g. Hacker 1997c, p.131)
[97] Hacker (1997a) p.278, Kenny (1984) p.125

It is true that I would never know whether my mother is in pain (after she accidentally spilled water whilst preparing tea, for example) unless she reveals it to me by screaming ‘Ouch!’, or by rubbing her reddened hand. On the other hand, she would know perfectly well whether she is in pain, without having to look for a behavioural expression. While I can be in doubt whether my mum is in pain or feels happy or sad or whether she is not, she cannot be in doubt about these things. However, it is this pattern of thought that is highly misleading, and leads to the hypostatisation of a metaphysical *inner mental realm* whilst engaged in philosophical reflection or scientific discourse. It is in such moments, that our understanding is bewitched by the means of language (*MS 219,11; PI §109)* and we tend to fall victim to our metaphysical dispositions, while we project the grammar of the *outer* upon the grammar of the *inner*.

Let’s consider the following question: In which cases can it meaningfully be said that the mental is inner? Well, I may be angry with somebody but not reveal it. Equally, I may think that current neuroscientific thinking about the mind is a lot of nonsense, but chose not to disclose it to the audience of Oxford neuroscientists to which I present the findings of my memory research,so as not to affront them. And, while I may be deeply in love with Cécile, I may decide that now is not the time to tell her about it. In contexts like these, that is, in so far as the *mental* refers to *mental states* that can under certain circumstances be concealed and its behavioural manifestations suppressed, it can meaningfully be said that it refers to something inner. Yet, if I tell Cécile that I love her, that her smile makes my entire universe resonate, or that I feel as if I had been taken hostage by my thoughts of her, unable to escape, then I have revealed (rather emphatically so, I believe) what in the metaphor of the inner and the outer *is* the inner. And it would make no sense to press me to reveal an inside (i.e. reveal a love) behind the inside that I have just revealed (i.e. “behind” my love for Cécile, which I have just declared). It would simply be wrong to assume after my declaration of love, that some love is still hidden from view. Similarly, it would be wrong to assume that if I scream with pain because Emmy, my dance partner, steps on my (already highly bruised) foot, that my pain is still concealed, and that my scream is merely a behavioural manifestation. Furthermore, if Cécile shows me photographs of her trip to Cambodia, I too can see some of the things she saw there, although at that moment I do not look into anything (i.e. I do not

look into her mind or brain for example). It is in this sense that the inner does not stand behind the *outer*. The inner *has no explanatory priority* over the outer. If we watch the famous 'Glockenspiel' on Munich's *Marienplatz*, for example, we may well wonder what it is that makes the owlet figures dance? That is, we may wonder what goes on inside the clock, that makes the figure dance? To find out we could climb the town hall tower and have a look inside the Glockenspiel. In such a case, it is the climbing of the tower and the looking inside the Glockenspiel, which *counts* as revealing the inside. To ask ourselves 'what, in the case of the human mind, *counts* as revealing the inner?', is the final push in the attempt to free ourselves from the deceptive force of the Inner/Outer picture.

If I am unsure whether Cécile is sad or happy or whether she is in pain, if I am in doubt about what she thinks about Emmy's new Latin dress, or about what she sees outside the window, she can tell me. Yet, if she does tell me what she sees, she does not *let me see inside her* nor does he *turn her insides out* (NFL, p.279). Cécile can tell me what she thinks, or point out to me what she sees. This is what counts as letting me know what she thinks or showing me what he sees. But to see what she sees is not to peer into her soul. That is, the meaning of the sentence 'I see what she sees' is not identical with the meaning of the sentence 'I can see into her soul', for example. While the former is a statement of a simple fact, *viz.* that I see the same house, car, dog, tree, or Latin dress, the latter is a statement about the insight and understanding I have into Cécile's character or personality (i.e. I know what moves her, I can empathize with her reasons for doing this-and-that and thinking thus-and-so). Similarly, on a stroll through the London Zoo Cécile can point at an elephant and say 'the elephant I rode in Cambodia looked like this'. Yet, this does not constitute a case of indirectly communicating what colour she saw. Rather, this is what is called showing me what colour she saw. The purpose of these various examples is to illustrate that it is the various behavioural manifestations, which provide the criteria for a person's revealing what goes on in his or her mind. Rather than depending an analogical extrapolation from personal experience, the ascription of psychological predicates is based on the *behavioural expressions* of what goes on inside a person. And rather than being mere symptoms, that is, inductive evidence, of how things are with a person, they are criteria of how things are with a person. Thus, contrary to the claims of Damasio and other cognitive scientists and philosophers we do not just

have to make do with indirect evidence, i.e. behaviour, when we ascribe psychological predicates to other people. It follows then that it is wrong to think that behaviour is nothing but externalities, and likewise, to think as Damasio suggests that one cannot only observe a feeling in oneself, but not in others[98]. If these claims strike us as counterintuitive or even absurd, it is helpful to remind us that they are not to be thought of as epistemological but as grammatical statements. They are not to be thought of as statements about what is true or false, but as statements about what makes and what does not make sense.

The Inner/Outer picture has a particular appeal when we ponder the nature of thinking. It is often the case that unless somebody voices his thoughts we cannot tell what that person is thinking. But to assume that because of a person may refrain from disclosing his thoughts that they are hidden is profoundly confused. To say that Cécile's thoughts are hidden from me, because they are in her mind amounts to no more than saying that *she thinks her thoughts* and *I cannot guess what she is thinking*:

> ' "Was ich im Stillen bei mir denke ist ihm verborgen" kann nur heißen, er könne es nicht erraten, aus dem und dem Grunde nicht erraten; nicht aber, er könne es nicht wahrnemen, weil es in meiner Seele ist.' (BPP_{II} §977)

> ' "What I think silently to myself is hidden from him" can only mean that he cannot guess it, for this or that reason; but it does not mean that he cannot perceive it because it is in my soul. ' (LW §977)

Likewise, if my dance teacher Lorraine, tells me that there is still a lot of work to do on the Jive, before Emmy and I can even dream of challenging Bryan and Carmen, it would be nonsensical to claim that what she told me are only words, and that she has still kept her thoughts to herself[99]. Both Cécile and Lorraine can tell me what they think. But when they tell me that they were thinking this and that, the criteria for the truthfulness of them saying so are not the same as those for a true description of a process. The criteria for a description are not the same as those for a confession. The truth of a description of a process is checked by observation of the process,

[98] Damasio (1999), p.42
[99] She rarely does, as far as my dancing is concerned.

whereas the criteria for the truth of a confession that he thought this and that are his sincerity. Neither Cécile nor Lorraine say that they thought this and that on the grounds of *observing* their thoughts. Rather, they gives expression to their thoughts by telling me about them.

2.2 Issues Surrounding the Faculty of Introspection

It has been noted earlier that the inner – outer duality is closely linked to a conception of introspection as a form of perception. Many philosophers and cognitive scientists believe that it is in introspection that we get access to the contents of our own consciousness, which is metaphysically speaking private property, known to each one of us in a unique way (LPD p.277) [100]. Philosophers and psychologists alike have confused the ability to say how things are with us, i.e. what we are feeling or what we are perceiving, with the ability to see (in the case of introspection, for example, we readily assume that we see with *the mind's eye*, as it were). Thus, we think we can say what is *within* us. David Hume, for example, claimed:

> 'Since all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Everything that enters the mind, being in reality a perception, 'tis impossible anything should to feeling appear different. This were to suppose that even where we are most intimately conscious, we might be mistaken.' (THN, Bk. I, Pt. iv, Sect. 2)

Similarly, Immanuel Kant viewed what he called the *inner-sense* (CrPR A22/B37) as the source of our knowledge of the objects of our private mental realm. It is this *inner-sense*, which provides us with an immediate, non-inferential, acquaintance with them. Furthermore, according to Descartes our knowledge of the inner world is not only immediate; it is also indubitable**.** For many, including the laymen not working in cognitive science or its related fields, these assumptions are intuitively appealing. In fact, they appear unquestionable. But once we have investigated the assumptions, which lend these ideas their seductive power, once we have taken off *'the pair of glasses which sit on our nose and which shape the form of whatever we look at'* (PI§103), however, the conception of introspection as a form of perception starts to loose its intuitive appeal. Instead we recognize that, in fact, this view is distorted by

[100] As William James points out, 'The word introspection need hardly be defined – it means, of course, the looking into one's own mind and reporting what we there discover. Everyone agrees that we there discover states of consciousness.', *in* 'The Principles of Psychology', p.185 (Vol..1)

conceiving of introspection as some form of inner vision. This misconception is the result of drawing a false analogy between our ability to say what we perceive and our ability to say what is on our mind (i.e. what we think, feel, intend, desire etc.). In order to gain an understanding what is wrong with this analogy, and why it is in fact incorrect to think of introspection as a form of perception, we have to remind ourselves of the defining features of our perceptual faculties and how the various forms of perception (seeing, hearing, smelling, etc.) are exercised. Above all, we have to get an understanding of the set of criteria, which determine what counts as the exercise of our perceptual faculties.

A range of factors influence the exercise of our perceptual faculties, the most prominent of which are

- the state of the relevant perceptual organ,
- the conditions of observation, and
- the observational skills of the subject.

These factors provide us with a rough idea of the range of behavioural criteria which are connected to the exercise of our perceptual faculties, and which justify or exclude the ascription of predicates of perception to a subject. As in the case of the ascription of psychological predicates, it is the various behavioural criteria which circumscribe the bounds of sense around the predicates of perception, and within which such predicates can meaningfully be ascribed.

In the case of introspection, however, we are hard pressed to identify any such criteria. First, introspection does not involve the use of any perceptual organ. We do not use our eyes, ears, nose or any other organ in order to find out what is on going on *in foro interno*, nor do they have any use in the expression of what is on our mind. And because introspection does not involve the use of any perceptual organ, the behavioural criteria, which would warrant the ascription of predicates of perception in are absent in the context of introspection. Consequently, predicates of perception cannot be meaningfully applied in the context of introspection. Second, the use of these sense organs is tied up with the concept of observation. But like the concepts of perception, the concept of observation does not get any grip in the context of introspection. People do not observe anything in order to be able to say what they

think, hope, expect, want or intend. The use of the word observation is bound up with the use and consequently the behavioural criteria linked to the exercise of our perceptual organs[101]. Again, it is the absence of these criteria, which tells us that no observation of any kind underlies our ability to say what is on our mind. For example, I might suffer an excruciating headache the morning after *Summer Eights*. Could I be said, other than metaphorically, that I have a headache because the voices of my better angels where shouted down by my unnatural craving for 'just one more pint of Pimm's'? Would better hearing have reduced my Pimm's Consumption? Surely not. Because of the absence of the relevant behavioural criteria there is also no analogue of good and poor hearing, or of good or poor eyesight. Neither *more light* nor *having a second look* will improve my faculty of introspection (or anybody else's for that matter). Finally, in relation to introspection there are also no observational skills that might be greater or lesser, and which might be honed with practice etc.. There is no more a *mind's eye* in anything other than a wholly metaphorical sense than there is a mind's ear, or nose. And where we do invoke the metaphor of the mind's eye, we speak of seeing *in* not *with* the mind's eye. So what enables us to what is going on in the minds eye? Surely nothing that is in any way connected to our visual abilities. Misled by such metaphors as the 'mind's eye', and by homonyms such as 'feeling pain' the prevailing tradition in philosophy and cognitive science has confused the *grammatical* exclusion of ignorance with the presence of knowledge. But the exclusion of ignorance, for example, the senselessness of the statement 'I do not know whether I am in pain' also *excludes knowledge*. I know that I am in pain is not an epistemic claim, and I feel that a pain in my right ankle is not a perceptual claim. Similarly, that statement that I am aware of a sudden burning sensation in my stomach, after I have enjoyed one of those fatal midnight Kebab's at Hassan's, is not a cognitive judgement.

If introspection is not a form of perceiving the inner vision of our mind's eye what then is it? What is the *nature* of introspection? In order to answer this question we have to look at the *grammar* of the concept. Once again we have to investigate how and under what circumstances a concept - introspection - is applied. To do so, let's start by asking what are the circumstances in which we describe other people as introspective? Two conditions stand out. First, people are described as being

[101] Hacker (1991), Chapter II

introspective when they are prone to reflect intensively upon their character, their motives for action, their beliefs, their attitudes, their emotions or their moods, for example. In this context, introspection is best understood as a form of reflexive thought. It is the term under which we subsume kinds of behaviour, which qualify as introspective. Thus, understood in this sense introspection is a means by which we gain an understanding of ourselves. It is a way by which people acquire self-knowledge. However, it is not a form of perception, which enables us to arrive at such knowledge or understanding of ourselves. Secondly, there may be situations in which one may note the waxing or waning of one's pain or one's affections for one's girlfriend. I may, for example, find that the pain in my bruised right foot decreases considerably, if I put it on a chair in order to rest it in an elevated position. Similarly, I may find that my passion for Beethoven has waned away in favour of an increasing obsession with Chopin. Understood in this second sense introspection is a matter of paying attention to one's moods and emotions, sensations and feelings.

In the light of this investigation, one is led to doubt explanations of mental phenomena involving the 'orthodox' conception of introspection as perception, and which has been revealed as deep misconceived in the above. Lawrence Weiskrantz's, for example, conceives of introspection as a neural monitor, which observes incoming 'visual sensations', and enables the perception of one's inner goings ons. In the context of the phenomenon of blindsight, he takes the disconnection of this neural monitor, i.e. a disconnection of what he calls visual sensation and visual perception, from the visual inputs as the underlying cause of he blindsight[102]. Given the results of the investigation into the concept of introspection, his ideas seem profoundly confused. Equally ill conceived is Philip Johnson-Laird view of introspection as a parallel processing device, which enables the neural system to construct a model of its own use of such models in a series of continuously ascending neural representations[103]. This suggestion is part of his attempt to construct a computational account of introspection. Johnson-Laird's view is a prime example of an attempt to provide a scientific foundation for a nonsensical idea (introspection as a form of inner perception). Together, these two examples[104] provide some insight into the confusion prevailing in cognitive science regarding the

[102] Weiskrantz (1999), p.319
[103] Johnson-Laird, The Computer and the Mind, Fontana London, p.363

notion of introspection, and the way this misconception leads to deeply flawed accounts of mental phenomena. Weiskrantz's and Johnson-Laird's ideas are representative of a host of neuroscientific explanations and theories about mental phenomena which constitute not so much insight but illusion.

2.3 The Direct-Indirect Duality

A further corollary of the conception of introspection as the means by which we perceive the contents of our private inner world is the idea that this inner realm is a kind of *private property* (see e.g.: LPD p.277). The objects of this inner world, one is tempted to think, are essentially owned. As *Gottlob Frege* pointed out, sensation is impossible without a sentient being and the existence of an inner world presupposes somebody whose inner world it is. Thus, it seems natural to suppose that every individual should have privileged access to his very own mental realm. This is enhanced by the fact that other people can apparently only come to know what I think or how I feel (i.e. whether I am in pain) if I tell them or show them (i.e. nurse my head or cheek, wear a gloomy expression on my face), whereas I do not have to rely on such behavioural criteria in my own case. Whilst it is true that it is the behavioural criteria which justify the application of psychological predicates to people, it does not follow that one has direct access to oneself, because one can say how one feels without having too refer to such criteria. This becomes evident once one gets a firmer grasp on the concept of getting access to someone or something.

A case in which it makes sense to speak of having access to something or somebody might be the following: I do have access to Jesus college library, for, as a member of Jesus College, I am granted the permission to use the college's facilities. Similarly, I have access to the world-wide-web via the computers in the college library, which I am allowed to use for such purpose. Furthermore, I have access to the Principal of Jesus College. Even more so, my access to the Principal might be described as *privileged*, for as president of the graduate common room (GCR) of Jesus College I am, unlike other members of college, able to see him on short notice or at inconvenient hours. To think that there is anything to such cases of 'having access to something or somebody' which is comparable to having access to the going ons in my mind, which would justify the claim that I have 'direct or privileged access' to my thoughts or feelings, is misconceived. When I feel my head ache, for

example, I can say so without reference to any evidential grounds (such as my affiliation to Jesus College). Any person who has mastered the use of a language can make assertions[105]. The utterance of such assertions like, I believe that Oxford will win the dancesport varsity match, or I think that Bryan Watson is the best jiver in the world, does not depend on me having privileged access to my beliefs or thoughts etc.. Rather, I have *reasons* or *grounds* for believing such-and-so or thinking this-and-that.

Similarly misconceived is the idea that others only have *indirect access* to my inner world and must make do with indirect evidence for describing me as happy or sad, or thinking that I am sincere or that I am a liar. It makes sense to speak of indirect access or evidence only in cases where it would make, in principle, also make sense to speak of direct access or evidence. It is important to note that the characterizing of evidence or knowledge as indirect is tantamount to drawing a contrast between two types of knowledge or evidence (i.e. direct and indirect evidence or knowledge). Having a pain, as mentioned previously is not to observe anything. The lesson to be learnt here is that there is no more a direct a way to finding out what a person thinks, believes, intends, experiences, feels etc. than the person sincerely telling me what he thinks, believes, experiences. Similarly, there is no more direct a way of finding out what another person sees than by his showing me what he sees. Although, there are circumstances in which it can be said that I have come to know something about somebody 'indirectly', such circumstances do not imply that I was unable to inspect the nature of the other person's consciousness. Rather, Cécile might find out that I am in love with her because Louisa, who I confided in her, told her so. In this case it could reasonably be said that Cécile found out indirectly about my feelings for her. In other cases, I might, for example, find out indirectly that my new acquaintance is a fan of rock music by having a glance at her CD collection. But if I chose to tell Cécile that I love her, or if I am being run away with by passions and just walk up to her kiss her, than for sure, Cécile would not think, nor could it reasonably be said of her, that she found out about my affections for her indirectly. That is, my love would no longer remain hidden. Thus, it only makes sense to talk of indirect knowledge, if it also does make sense of direct knowledge, as this distinction implies a contrast. In context of the mind and the application of

[105] See for example: Bennett & Hacker 2003, p.93

psychological predicates, however, there is no contrast which would warrant the meaningful application of expression of direct and indirect access. The direct – indirect duality is nothing but a metaphysical myth, resulting from a misleading analogy (LSD, p.13):

> '…we compare the case of a person saying 'I am in pain' as opposed to someone saying *of* him on he basis of his behaviour 'he is in pain' with the case of, e.g., knowing that there is a penny in my pocket by feeling it ('directly') as opposed to knowing that there is a penny in your pocket by hearsay ('indirectly').'[106]

2.4 The Public - Private Duality

> 'Nobody else has my pain. Someone may have sympathy with me, but still my pain belongs to me and his sympathy to him. He has not got my pain, and I have not got his feeling of sympathy.'[107]

This remark by Gottlob Frege nicely illustrates another source of misunderstanding regarding the picture under examination. It lies in the idea that another person *cannot have my experiences*. My pain as well as your sympathy are inner objects, mental representations in the jargon of modern cognitive science, which belong to our respective subjective inner worlds. As described in the introduction to the preceding sub-section, a corollary of the idea of introspection as the means by which we perceive the contents of our private inner world is the notion that this world is a kind of private property. For example, we readily assume that if two people do have a headache, there are two headaches not one. Whilst we may concede that two people have the same pain, i.e. the same stinging sensation in their head, we are at the same time inclined to think that all this means, however, is that the head aches are exactly alike. Yet, it does not imply that they are identical. Thus, while the pains of two people may be qualitatively identical, we still conceive of them as numerically distinct[108]. The inclination to argue this way seems to arise from the assumption that the location of a sensation is a criterion for who has it. The burning sensation which Michelle feels in her stomach is a burning sensation in her stomach.

---

[106] Hacker (1997a), p.287
[107] Frege G, (1984) 'Thoughts', *in 'Collected Papers on Mathematics, Logic and Philosophy'* , Oxford, Blackwell, p.361
[108] See for example, Hacker (1997b) p.20

Alice might feel a similar burning sensation in her stomach, but since Alice's stomach cannot be in the same place as Michelle's stomach, her burning sensation cannot be in the same place either.

The mistake we fall victim to in such cases is that we project the grammar of expression containing physical objects onto expressions pertaining to experience. That is, we wrongly take the concept of location of sensation to be analogous to the concept of location of a physical object. But, the location of a sensation is not a criterion for who has it.

> ”Der Andre kann nicht meine Schmerzen haben.” Welches sind meine Schmerzen? Was gilt hier als Kriterium der Identität? Überlege, was es möglich macht, im Falle physikalischer Gegenstände von ”zwei genau gleichen” zu sprechen. Z.B. zu sagen: ”Dieser Sessel ist nicht derselbe, den du gestern hier gesehen hast, aber er ist ein genau gleicher.”
>
> Soweit es Sinn hat, zu sagen, mein Schmerz sei der gleiche, wie seiner, soweit können wir auch beide den gleichen Schmerz haben. (Ja es wäre auch denkbar, daß zwei Menschen an der gleichen nicht nur homologen Stelle Schmerz empfänden. Bei siamesischen Zwillingen, z.B., könnte das der Fall sein.)
>
> Ich habe gesehen, wie jemand in einer Diskussion über diesen Gegenstand sich an die Brust schlug und sagte: ”Aber der Andre kann doch nicht diesen Schmerz haben!” Die Antwort darauf ist, daß man durch das emphatische Betonen des Wortes ”diesen” kein Kriterium der Identität definiert. Die Emphase spiegelt uns vielmehr nur den Fall vor, daß ein solches Kriterium uns geläufig ist, wir aber daran erinnert werden müssen.’ (PU§253)

> ‘ ”Another person can´t have my pains.” – Which are my pains? What counts as a criterion of identity here? Consider what makes it possible in the case of physical objects to speak of “two exactly the same”, for example, to say “This chair is not the one you saw here yesterday, but is exactly the same as it”.
>
> In so far as it makes sense to say that my pain is the same as his, it is also possible for us to have the same pain. (And it would also be imaginable for two people to feel pain in the same – not just the corresponding – place. That might be the case with Siamese twins, for instance.)
>
> I have seen a person in a discussion on this subject strike himself on the breast and say: “Bur surely another person can’t have THIS pain!” – The answer to this is that one does not define a criterion of identity by emphatic stressing of the word “this”! Rather, what the emphasis does is to suggest the case in which we are conversant with such a criterion of identity, but have to be reminded of it.’ (PI §253)

In order to obtain an *Übersicht* over the grammar of expressions containing physical objects Wittgenstein asks us to ponder what enables us in the case of physical objects to distinguish between being identical *(dasselbe)* and being exactly the same *(das gleiche)* yet not identical. For example, I have two IKEA shelves of the type 'Billy'. They are exactly alike, yet not identical. One of them is located in my study the other one in my living room. My friend Toby who comes to visit me in my new flat may observe that the shelve 'Billy' in my new living room is the same shelve 'Billy' he saw in the living room of my previous flat. Furthermore, as I show him round my new home, he may observe that I now have another shelve 'Billy' (newly acquired) standing in my study. In this case it would make sense for him to say that this is not the 'old' shelve 'Billy' he knows from inhabiting the living room of my old flat, but that this is a 'new' shelve 'Billy', which is exactly like it. Now, does the same distinction, between being identical and exactly the same apply to sensations, like pain for example? In order to answer this we have to clarify what counts as a criterion of the identity of a sensation. What is called 'having a burning sensation in the same part of our body' *is* having the same sensation in the same part of the body (i.e. Michelle in her stomach and Alice in hers). Doctors, for example, may frequently ascribe the same pain to various patients if the respective pains are sufficiently alike in intensity, phenomenal characteristics and location. Even if one went on to object that a 'corresponding location' is not 'the same location', and that hence two people cannot have *the same* pain one would still be mistaken. For, if we imagine as Wittgenstein did the case of two Siamese twins who are conjoined (at identical places) at the head or at the back, we must concede that in such a case it is logically possible for two people to have the same pain, and not just in corresponding places (PI §253). Thus, location is not a criterion for the possession of an identical sensation. Possessive phrases like, being mine or being yours, or being his or being hers characterize *who* has a headache. They do not single out or identify the headache in question (PI §302), nor do they single out or identify any experience as implied by the public-private duality .

2.5 The Final Straw that Breaks the Camel's Back or the Inner/Outer Picture of the Mind and the Meaning of Psychological Predicates

Thus far, the present discussion has identified various different sources of confusion underlying the Inner/Outer dichotomy (e.g. flawed ideas about the faculty of introspection and the concepts of direct and indirect access and privacy). The resulting distorted picture of mind, which itself is part of the confusion underlying the mereological fallacy, is deeply rooted in these conceptual confusions. In order to free ourselves from this misguided picture of the mental it is vital that all of its roots are identified and pulled up. Only in doing so can we make sure that the weed will be thoroughly eradicated, as it were.

At its most fundamental the metaphor of the Inner and the Outer is linked to an erroneous conception of language, which leads to a misunderstanding of the meaning of our psychological vocabulary. This flawed conception of language is rooted in what with the publication of Wittgenstein's *Philosophical Investigations* came to be known as the Augustinian picture of language:

> 'Nannten die Erwachsenen irgend einen Gegenstand und wandten sie sich dabei ihm zu, so nahm ich das wahr und ich begriff, dass der Gegenstand durch die Laute, die sie aussprachen, bezeichnet wurde, da sie auf ihn hinweisen wollten. Dies aber entnahm ich aus ihren Gebärden, der natürlichen Sprache aller Völker, der Sprache, die durch Mienen und Augenspiel, durch die Bewegungen der Glieder und den Klang der Stimme die Empfindungen der Seele anzeigt, wenn diese irgend etwas begehrt, oder festhält, oder zurückweist, oder flieht. So lernte ich nach und nach verstehen, welche Dinge die Wörter bezeichneten, die ich wieder und wieder, an ihren bestimmten Stellen in verschiedenen Sätzen, aussprechen hörte. Und ich brachte, als nun mein Mund sich an diese Zeichen gewöhnt hatte, durch sie meine Wünsche zum Ausdruck. *(Augustinus, in den Confessionen I/8)*
>
> In diesen Worten erhalten wir, so scheint es mir, ein bestimmtes Bild von dem Wesen der menschlichen Sprache. Nämlich dieses: Die Wörter der Sprache benennen Gegenstände Sätze sind Verbindungen von solchen Benennungen. In diesem Bild von der Sprache finden wir die Wurzeln der Idee: Jedes Wort hat eine Bedeutung. Diese Bedeutung ist dem Wort zugeordnet. Sie ist der Gegenstand, für welchen das Wort steht. (PU §1)
>
> ' "When they (my elders) named some object, and accordingly moved towards something, I saw this and I grasped that the thing was called by the sound they uttered when they meant to point it out. Their intention was shewn by their bodily movements, as it were the natural language of all people: the expression of the face, the play of the eyes, the movement of other parts of the body, and the tone of voice which expresses our state of mind in seeking, having, rejecting, or avoiding something. Thus, I heard words repeatedly used in their proper places in various sentences, I gradually learnt to understand what objects they signified; and after I had

> trained my mouth to form these signs, I used them to express my own desires." (Augustine, Confessions, I.8)
>
> These words, it seems to me, give us a particular picture of the essence of human language. It is this: the individual words in a language name objects- sentences are combinations of such names. – In this picture we find the roots of the following idea: Every word has a meaning. This meaning is correlated with the word. It is the object for which the word stands. (…)' (PI §1)

The most significant misconception inherent in this picture is the idea that first-person present tense psychological statements are descriptions of one's own mental states. The Inner/Outer picture of the mind, endorsed by the rationalist and empiricist schools of philosophical thought as well as contemporary cognitive science, is deeply rooted in the Augustinian picture of language. Within the empiricist tradition, for example, the foundation of knowledge is comprised by the *ideas* which are obtained from the use of our inner and outer senses[109]. Accordingly, the fundamental building blocks of language were thought to be the names given to simple ideas. As a result, the private mental objects inhabiting our inner world were deemed to provide the foundation of language. The most modern upshot of this conception is the idea that the fundamental rules of language are *innate* and given *antecedent* to any experience[110]. This view is endorsed throughout a large part of the cognitive science community. Thus, in the 'classical' as well as the modern case the roots of language, and consequently, the meaning of any psychological expression, are conceived as essentially mental. It is in this way that the predominant picture of the mind leads to the idea of a *private language* (i.e. a language that can be understood by nobody but it's owner (PI§243)) A consequence of the received conception of language outlined above is that the subject-matter of language is viewed as consisting exclusively in the subjective experience of the speaker. It implies that the words of the language, referring to the names of sensations, feelings, moods, or sense-data obtain their

[109] The *locus classicus* of this conception is to be found in John Locke's *Essay Concerning Human Understanding*. Locke's philosophy of language is based on the idea that the meaning of words is given by their association with the *'ideas in the mind of him that uses them.' (Essay 3.2.2)*

Contrary to Locke, Descartes, did not reflect deeply on the nature of language. In formulating his (in)famous Cogito-argument, he took the meaning of a number of basic terms such as thought and existence, for instance, for granted: *'And when I said that the proposition I am thinking, therefore I exist is the first and most certain of all to occur to anyone who philosophizes in an orderly way, I did not in saying that deny that one must first know what thought, existence and certainty are… .' (Principles of Philosophy, P.X; in, Cottingham, Stoothoff, Murdoch, 'The Philosophical Writings of Descartes, Vol.I, p.196)*

[110] See e.g., Chomsky (1966), ‚Cartesian Linguistics'

meaning by way of *private* ostensive definition[111]. In this context, concentrating on an experience and associating the experience with a name, for example, is often understood as a mental analogue of physically pointing at an object. However, a private language thus conceived is impossible and the entire conception of language associated with the Inner/Outer picture of the mind rests on conceptual confusion.

The question we have to examine in order to disentangle us from the confusion underlying the idea of a private language is whether psychological expressions can be assigned a meaning by means of private ostensive definition? We may ask, for example, whether we learn the meaning of pain in this way? Did we master the use of 'I am in pain' by naming a private experience? Wittgenstein considers the following analogy:

> '...angenommen, es hätte Jeder eine Schachtel darin wäre etwas, was wir " Käfer" nennen. Niemand kann je in die Schachtel des Andern schauen; und Jeder sagt, er wisse nur vom Anblick seines Käfers, was ein Käfer ist. Da könnte es ja sein, dass Jeder ein anderes Ding in seiner Schachtel hätte. Ja, man könnte sich vorstellen, dass sich ein solches Ding fortwährend veränderte. Aber wenn nun das Wort " Käfer dieser Leute doch einen Gebrauch hätte? so wäre er nicht der der Bezeichnung eines Dings. Das Ding in der Schachtel gehört überhaupt nicht zum Sprachspiel, auch nicht einmal, als ein Etwas: denn die Schachtel könnte auch leer sein. Nein; durch dieses Ding in der Schachtel kann 'gekürzt' werden; es hebt sich weg, was immer es ist.
>
> Das heißt: Wenn man die Grammatik des Ausdrucks der Empfindung nach dem Muster von 'Gegenstand und Bezeichnung' konstruiert, dann fällt der Gegenstand als irrelevant aus der Betrachtung heraus. (PU §293)

> '…Suppose everyone had a box with something in it: we call it a 'Beetle'. No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle. – Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. – But suppose the word 'beetle' had a use in these people's language? – If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as a *something*: for the box

[111] To define a word by means of ostensive definition is to define a word by pointing to its literal counterpart. For example, I can give an ostensive definition of a chair by pointing to a chair saying, "this is a chair." This kind of definition stands in contrast to dictionary-style definition which only uses words to explain the meaning of other words. In ostensively defining a term we obtain a sense that we are making a real connection between language and the world. As the Augustinian picture of language illustrates we are tempted to regard ostensive definition as the paradigmatic form of definition and the kinds of objects that can be so named as the paradigmatic form of words

might even be empty.- No, one can `divide through´ by the thing in the box; it cancels out, whatever it is.

That is to say: if we construe the grammar of the expression of sensation on the model of 'object and designation' the object drops out of consideration as irrelevant.' (PI §293)

If 'beetle' were to have a use in the common language of the group of people or 'box owners' referred to in this paragraph of the *Philosophical Investigations*, the term would not be related to whatever is in each persons' private box. In fact, the content of the box *cancels out* as far as communication in the public language between the box owners is concerned. Consequently, with regard to the meaning of those expressions containing the word 'beetle', the contents of the boxes are utterly irrelevant. And the same is true in the case of talk about pain or any other *private* experience. If we assume, for example, that the *private* experience of pain varies significantly from person to person whilst applying the 'Inner/Outer' picture of the mental, the differences between the various pain experiences would neither be verifiable nor falsifiable. As a result, the public use of the word 'pain' could not be affected by these private differences either. An entirely private pain sensation, or any other private experience in fact, 'drops out of consideration as irrelevant'. After all, 'a wheel that can be turned though nothing else moves with it, is not part of the mechanism' (PI§271). Two conclusions can be drawn from these considerations. First, our public talk about our experiences or sensations like pain, for example, is not dependent on the character of a *'private'* experience. Second, a sensation like pain, is not a private experience. The meaning of the word pain or any other sensation or experience is not determined by naming an experience in one's private inner realm.

3. Methodological Queries: Objections and Replies

Anthony Kenny points out that cognitive scientists often shrug off and defend their practice of ascribing psychological attributes to the brain as *a harmless pedagogical device*[112]. Indeed, in response to the claim that ascribing a psychological attribute to the brain makes no sense, cognitive scientists often respond by pointing out that actually *they didn't mean it like that* or that such statements are to be understood in a *metaphorical sens*e. That is to say that the words or expressions applied were used in a special sense as *analogical extensions* of their usual

[112] Kenny A (1984), p.125

application. They must not be taken to have their *ordinary meaning*. Rather, the words or expressions in question, i.e. that the brain believes that…, has been given a *new meaning* as the predicate is not used in the ordinary sense. Again, Antonio Damasio provides an excellent example of the kind of illusion and excuse alluded to here. In elaborating his neurobiological theory of consciousness he states that:

> `looking back, with the *license of metaphor*, one might say that the swift, second order nonverbal account (=core consciousness) narrates a story: "that of the organism caught in the act of representing its own changing state as it goes about representing something else." …This plot is incessantly repeated for every object the brain represents….´[113]

Such claims are problematic, however. What could it possibly mean to say that one adopts *metaphoric licence* when proposing a neurobiological theory of consciousness. Is this tantamount to saying that certain words adopt a *new meaning*? If it does then there is a problem. For there is no evidence that Damasio, or cognitive scientists in general, have *really* given *new meanings* to the words they use which would legitimise the ascription of psychological attributes to the brain or parts of the brain. Let alone render talk of second order neural representations, which narrate stories intelligible. When Descartes postulated that the *soul sees* images projected on the pineal gland, for example, he applied the verb *to see* in its ordinary sense (i.e. as applied in statements like, *I see an English Bulldog chewing on my guitar case*). Similarly, Damasio and other cognitive scientists, contrary to any claims brought forward, also continue to apply the psychological terms and expressions in question in their ordinary sense. Just like Descartes did. If that was not the case, cognitive scientists would not draw the kind of inferences from their research which they do draw. This is reflected in statements like

> '…the brain recognizes faces' …It appears as though a special subsystem in the brain sees faces; it is triggered to produce the percept for our conscious lives by the configuration of elements. The special face processor does not know or care about what elements it is composed of; as long as they are in proper arrangement, a face is perceived. What could be more fascinating than to study how the brain does such things?'[114]

[113] See: Damasio (1999), p. 179. According to Damasio the *story narrating* core consciousness is what in his terms is a second order neural representation: *core consciousness occurs when the brain's representation devices generate an imaged, nonverbal account of how the organism's own state is affected by the organism's processing of an object, and when this process enhances the image of the causative object, thus placing it saliently in a spatial and temporal context.´* (see ibid: p.169)

[114] Gazzaniga et al 1998 p.2

It is important to note that recognizing means 'recognizing' here, i.e. to identify something or someone as having been encountered before, and being able to make a discriminatory judgement as a result of the recognition. Furthermore, it is important that the special subsystem *sees* faces in the ordinary sense of *to see*. Only thus can it pick out or recognize the proper elements of a visual scene and contribute to the perception of a face (needless to point out that what is meant here is not that a human being perceives a face here, but his or her brain).

Yet, as the previous discussion showed, it makes no sense to say of a brain that it thinks, hears, sees, imagines, remembers, recognizes, perceives or believes as we do not know what would count as a brain thinking, hearing, seeing, imagining, remembering or believing. Furthermore, if the cognitive scientists really had given a new meaning to those kinds of expressions, they would have had to adopt a new conceptual framework with new rules for the application of certain familiar terms. Again, this is because the bounds of sense are drawn by the rules of language. If one wanted to change the rules for the use of certain psychological concepts, like believing or recognizing, for example, one would have to change a lot else as the rules for the application of certain concepts are part of an intricate and widely ramified network of conceptual connections which form an integral part of our form of life (see Chapter I, see also above).

Although perception is indeed a fascinating field of study, our understanding of this psychological phenomenon will not be advanced by explanations of the kind cited above. Rather, in cases like this cognitive scientists engage in a form of speculative metaphysics (or psychology) propounding seriously flawed explanations and theories. Thus, philosophy can not only add perspective to cognitive science but also make a real contribution to moving neuroscience forward, and help to put in on sure footing. And philosophy can do this contrary to the perceptions of some cognitive scientists:

> '...to understand how a biological system works, a laboratory is needed and experiments have to be performed. Ideas derived from introspection can be eloquent and fascinating, but are they true? Philosophy can add perspective, but is it right? Only scientific method can move a topic along on sure footing.'[115]

[115] ibid. p.2

4. Summary and Conclusion

A variety of misconceptions have been discussed and clarifications proposed. Contrary to the common practice in cognitive science it is wrong to ascribe psychological predicates to the brain in order to explain the possession of psychological attributes by human beings. The behavioural grounds, which provide the criteria for the ascription of a psychological predicate, are partly constitutive of the meaning of the predicate. They are conceptually connected with the attribute in question. Yet, the brain does not exhibit pain-behaviour. There is no such thing as a brain engaging in such activities like assuaging it's limb, or moaning of its aching head. Consequently, explanatory claims like those propounded by Gazzaniga, Crick, Damasio and others are mistaken. They wrongly assume that they provide explanations of psychological phenomena but instead further misconceptions about the human mind by distorting our understanding of the phenomena investigated. What cognitive science can do, on the other hand, is to observe the neural concomitants that might be involved when a person is suffering from pain. However, such neural activity is not a form of pain behaviour. Neural phenomena may be inductively correlated with being in pain. Yet, the correlation is an *empirical discovery*, which *presupposes* the concept of pain and its connection with non-inductive behavioural evidence that warrants the application of the concept of pain.

The philosophical investigation into the mereological fallacy in cognitive neuroscience showed that the misconceptions and conceptual confusions which cognitive scientists fall victim to are manifold and profoundly interconnected. The mereological fallacy, it has been shown, is closely linked to the Inner/Outer picture of the mind which has dominated philosophy since Descartes and the British empiricists. This picture, in turn, is the result of profound conceptual confusions regarding the nature of introspection, privacy, and the nature and foundation of language. Thus, the claim that *I know that I am in pain*, for example, is not an epistemic claim; different people can have the same experience, share the same feeling or can harbour the same thought; inner states stand in need for outward criteria; a private language is impossible. It is important to understand that these statements are not theses constituting improved alternatives, which would override other philosophical or scientific theses. Rather, these statements are to be

understood as grammatical remarks, arrived at by investigating the uses of certain words and expressions. Grammatical remarks pertaining to the use of psychological predicates highlight logical connections between words and the psychological phenomena to which they relate. They are, as Norman Malcolm remarks, “truisms“ since they merely spell out some features of our familiar uses of an expression[116].

Part of their function is to describe the rules for the use of concepts, and thus, to delineate the bounds of sense. They are reminders, which the philosopher assembles to provide a prophylactic against conceptual confusions (PI§127). As such they are not philosophical claims, which have been proven to be correct (or incorrect), but rather they are examples of what it means to provide an overview (PI §122). The purpose of these grammatical elucidations is to demonstrate that some established ways of providing explanations of mental phenomena are incorrect and to demonstrate that what appeals to many as an intuitively comprehensive picture of the human mind is, in fact, incoherent and mistaken. In doing so, the achievement of this chapter lies primarily in the exposition and destruction of what is a customary but profoundly misconceived way of thinking about the human mind and brain. Due to the crypto-Cartesianism persisting in the cognitive view, it invites us to conceive of mind and brain in the ways examined in this chapter. Besides the critique of the widespread ascription of psychological predicates to the brain, the present chapter has dealt a first decisive blow to the mentalist tendency to explain cognition in terms of *inner* mental states and processes. To expose these defects is central to providing alternative ways of understanding psychological phenomena. This will be carried out in subsequent chapters.

[116] Malcolm (1995), p.85

Chapter IV
The Cognitive View and the Mind

1. The Mind: Preliminary Observations

What is the human mind? What is its nature? As described in chapter II, cognitive scientists tend to answer these two questions in a broadly physicalist manner. First, classical cognitive science views the mind as a *biological information processor, a representational system, which receives, stores, retrieves, transforms and transmits information*. Second, the nature of the mind is thought to be material and realized in the neural circuitry of the brain. Consequently, the terms mind and brain are used as synonyms, and the conception of the mind/brain is widely spread among cognitive scientists and philosophers. As a result of the nowadays prevalent believe that an understanding of humanity can only be arrived at by applying the causal hypothetico-deductive framework of explanation encountered in the natural sciences[117], and the equally wide spread tendency to view the mind as some form of biological computer, these reductionist ideas have gained a deceptive appeal. From a Wittgensteinian point of view, there is good reason to be deeply suspicious about the intelligibility of these ideas. Despite the appealing mix of modern scientistic and classic Cartesian and empiricist metaphysics, these ideas are nothing more than wild conjectures, a prime example of the confusion that ensues when the need for conceptual clarity is neglected. They are nothing but bad philosophical questions and theories, which are not amenable to empirical investigation, but demand the quiet weighing linguistic facts (Z§447). Following the later Wittgenstein's principles of philosophical investigation, the present chapter will expose this reductionist notion of the human mind as a biological information processor as the manifestation of

[117] The immense success that the sciences enjoyed, particularly in the 20th century, combined with their tremendous economic potential and the promise of financial gain, resulted in a triumphant increase in their popularity. This made the rifts between the natural sciences and the humanistic studies (which had developed and grown for centuries) finally seem unbridgeable (see e.g. Snow CP, 1993). Most importantly, this development was accompanied by a general devaluation of humanistic understanding, which gave rise to *scientism* - the illicit extension of the methods and forms of explanation of the natural sciences. The wide spread belief in the omnipotence of the causal hypothetico-deductive framework of explanation encountered in the natural sciences (methodological scientism), and the ontological reductionism informing the cognitive view are but two contemporary examples of this development.

misconceived philosophical theorizing. The investigation of the concept of mind and its conceptual connections will provide the starting point of this discussion.

The reductionist notion of the mind as a biological information processor ultimately rests on the misinterpretation of various pictures of the mind, which are deeply embedded in our language. These pictures surface, for example, in our everyday parlance about our intellectual faculties, our intellectual abilities and inabilities or in talk about our intellectual virtues and vices (for details see section 2 below). Unfortunately, figures of speech like having a broad or narrow mind, like having a sharp, cunning, agile or vicious mind are often erroneously taken at face value, and believed to refer to a single entity. Furthermore, the adjectives used in these expressions are taken to describe distinct attributes of that entity[118]. However, *substantives are substance hungry*[119], and as a result, what is a mere *facons de parler* philosophers and cognitive scientists are prone to "take it as read". This, in turn, results in questions with regard to the nature of the entity – the mind - (e.g. what is this entity, the mind, made out of?). However, such questions are pernicious as they invite the construction of flawed philosophical and/or scientific theories in answer to misunderstood philosophical questions (misunderstood in the sense that they rest on conceptual confusions which are the result of a lack of an overview over the grammar of our concepts). Following the late Wittgenstein, however, we have to turn our attention to the way in which the term *mind* is used, investigate how the concept is applied in its natural habitat, in order to arrive at an understanding of the nature of mind (PI§370). The challenge we face in order to arrive at an understanding of *mind* is not one of devising sophisticated theories, but to bring back the word mind from its metaphysical to its everyday use (PI§116)

2. Investigating the Mind in its Natural Habitat

The notion of mind as an entity is a characteristic misconception featuring in discussions among philosophers and cognitive scientists. As a result, it is not surprising to find the adherents of the cognitive view give in to the temptation of providing a theory and definition of the mind by suggesting that the mind is a

[118] For Cartesian dualists the entity was an immaterial substance, whilst for the majority of modern cognitive scientists and philosophers the entity is a material substance, i.e. the brain.
[119] See: Hacker (1997b) p.64

biological information processing unit, which receives, stores, retrieves, transforms and transmits information. The following discussion will provide an overview of the grammar of the concept of mind and trace its links to the numerous psychological predicates and dispositions of human beings, which it is associated with. In doing so, it draws on earlier discussion of the subject by Hacker (1997b) and Kenny (1988),

In ordinary discourse the *concept of mind* is most readily associated with the intellect, that is, in particular with *intellectual powers* and their deficiencies or the lack of such powers. Consequently, the concept frequently features in talk about the character and quality of somebody's intellect or somebody's intellectual faculties and thus, is connected to descriptions of somebody's personality or traits. For example, we readily describe clever, intelligent people as having a potent, powerful or sharp mind. Similarly, we speak of someone as having a quick, agile or subtle mind if a person easily understands the various nuances of a problem and/or is particularly quick at solving it. In contrast, we are prone to describe people who we perceive as generally intellectually challenged as having a slow mind if they are prone to err or fail in such situations. Furthermore, we would speak of someone as possessing a versatile mind if a person is accomplished not only in the sciences, for example, but also possesses knowledge and expertise in the humanities. When talking about another person's character or personality, we might remark that the person at the centre of our discourse has a dirty or filthy mind, if this individual uses a lot of swear words, or if his discourse is littered with sexual innuendos, for example. Alternatively we might describe somebody has having an idle mind, if the person we are aiming to portray is particularly lazy and frequently displays a lack of concentration. Finally, somebody we think of being in possession of a persuasive mind distinguishes himself by his ability to make good arguments and easily convince people of his point of view. These few examples provide us with a first vague idea of the extensive ground[120], which the concept of mind covers. Including aspects of one's ability (or inability) to solve problems to one's powers of persuasion (or lack thereof), the concept of mind seems to be applied to more contexts than the simple conception of the mind as a biological information processor allows room for, (i.e. there seems to be a single minded emphasis/over emphasis of the problem solving abilities of human beings. The concept o mind applied in cognitive science seems to put undue

[120] As far as conceptual geography is concerned.

emphasis on problem solving abilities, things hat can easily be picked up or measured by IQ tests, but neglect those aspects of mind, like someone's powers o persuasion, who involve taking into account its more discursive aspects).

Of equal prominence to the association of mind and intellect, are the connections of the concept of mind with the concept of *will*. A strong-minded individual, is not easily influenced or swayed by others, and may go about doing something with perseverance. Thus, I may keep on pursuing my philosophical investigations, despite the fact that the neuroscientific community chooses to disregard their relevance. Similarly, Nick may continue to work on his thesis, despite his housemates tempting him with red wine, a curry and the latest episode of *The West Wing.* Strong-minded individuals may also often serve as examples for others, and be referred to in the context of the kind of motherly advice delivered in such phrases like *"anything goes if only you set your mind to it…"*. In contrast, a weak-minded individual may be easily swayed and influenced. For example, if I was tempted by Jumin, Sugi, Robin and Nick with red wine, curry and the latest episode of *The West Wing,* I would desperately fail in trying to keep my mind on my thesis and off the fact that I could spend an enjoyable evening with my friends[121]. Apart from being strong-minded or weak minded, somebody may also be indecisive or be in two minds about something. That is, someone may be unable to make up his mind or change his mind constantly. For example, I may be said to be in two mind's about whether this is the right time to tell Cécile that I love her, or not. Or I may be in two mind's about continuing the work on my thesis tonight, or on spending some time with my friends.

Besides connections of the concept of mind with intellectual power and will, it has strong links with the concepts of *thought, memory and opinion*. *a) Thought:* As my brother was just released from hospital, I no longer need to be concerned about his health and I can now turn my mind back to my philosophy thesis without having to feel guilty about it. And although I am struggling with this chapter I am sure that if I put my mind to it I will be able to finish it today. *b) Opinion:* To make up one's mind is to form an opinion, while to be in different minds about something is for two people to have different opinions on a certain matter. For example, I have finally managed to make up my mind about Cécile. I am going to tell her. Also, Dave and I are in

[121] Of course, this is not to say that an evening writing one's thesis cannot be enjoyable.

different minds about who is he best jiver in the world. He thinks it´s Slavik, whereas I believe that it is Bryan Watson. *c) Memory:* To bear or keep something in mind is to remember something. For something to slip out of my mind is for me to forget something. And, somebody who is prone to daydreaming may be described by his peers as being frequently absent-minded. On the other hand, the fact that I can't get Cécile off my mind means, that I am unable to forget about her, and that I think about her all the time (except for those times when I am writing my thesis, of course).

This latter case points at an important aspect of the use of the concept of mind. In various instances a particular use of the concept can have more than one meaning or be applied in more than one context. Its use runs criss cross in all directions (PI§67). Thus, the fact that I can't get Cécile off my mind can mean either that I constantly think about her, or that I am unable to forget about her. Similarly, the ascription of a great mind to somebody can either be taken as a depiction of a brilliant thinker, or alternatively, it can be taken to describe a visionary who propounds ideas of great and unprecedented originality[122]. Furthermore, if someone is ascribed a beautiful mind, such a person may either be thought of as harbouring a particularly strong interest in the arts and the humanities or, alternatively, such a person may also be believed to possesses a particularly gentle and caring nature.

Finally, the use of the concept of mind is bound up with various pictures of room, space, agency, and of parts. For example, for someone to have a narrow or small mind is for such a person to be petty and/or prejudiced. Likewise, parents who lost their child in a car accident, for example, may be out of their mind with grief and anger, and various thoughts may have crossed their minds with regard to how to punish the drunken driver who killed their daughter. Alternatively, thoughts about Cécile may be at the forefront of my mind, whereas I may keep putting thoughts of my thesis deadline to the back of my mind. Expressions like these constitute a metaphorical portrayal of our degree of attentiveness to things of importance. Or to put it differently, such portrayals indicate the relative importance in which various matters stand in relation to our lives. In the context of agency, somebody may be said

[122] Of course, these characteristics are not mutually exclusive and can be true of one and the same person. For example, Wittgenstein was both a brilliant thinker and a great visionary in as much as in his later phase he propounded a radical and revolutionary approach to philosophy. Similar things can also be said about Galileo, for instance In both cases, the ascription of a brilliant mind can be used to refer to different aspects of the personality or abilities of the person to whom a great mind is ascribed to.

to have a quick or slow mind as a result of solving problems swiftly and adeptly, or because of somebody's inability of doing thus. And, I may be in two minds about whether to participate in the foxtrot competition today, and if I am being honest with myself come to the conclusion that it's not worth doing because I have only half a mind for doing it. The similarity between the uses of these expressions and the pictures of mind discussed earlier on in this section is noteworthy. They underline the subtle modulations and shifts in the meaning of the concept of mind, as a result of its manifold uses, which run criss-cross in all directions.

The examples described here are by no means meant to provide an exhaustive overview of the various uses of the concept mind. They suffice, however, to illustrate how one may be mislead by the various uses of the concept and come to picture the mind as a single discernable entity who is referred to in these various expressions. From there it is only a small and natural step to start asking questions of the kind *What is the mind?, What is it's nature?.* The lack of an overview of the various uses of the concept of mind leads to confusion among philosophers and cognitive scientists, who go about their daily business whilst being utterly unaware of the intricate ramifications of the concept of mind, and the pictures which are associated with it. By conceiving of the meaning of the concept of mind as uniform, they distort the character of the phenomenon in question. The cognitive view of the mind constitutes the precipitate of this distortion. The misunderstanding of various figures of speech containing the concept of mind leads to a mistaken quest for definitions and theories, while the true nature of these expressions as convenient ways of talking about characteristics human traits, faculties, powers and their exercise, remains unrecognised.

The misconceived definitions, theories and resulting hypothesis delude philosophers and cognitive scientists into dreaming up solutions of questionable intelligibility to misconceived difficulties (or as Wittgenstein put it, false solutions to false difficulties) (RPPI§1063). The difficulties that arise in the wake of these misconceived solutions are then attributed to the particular difficulty of the problem under investigation, whose impalpable nature makes it seem mysterious (Z§125,126). Often, the failure to catch hold of the problem is then attributed to the fact that the investigation itself is still in its infancy. But of course, this is only the

result of misunderstanding the nature of science and the nature of philosophy (see Chapter II), which induces cognitive scientists to embark on experimentation without having gained a proper understanding of what it is, that makes the human mind seems so mysterious and unfathomable (RPPI§1093).

The present investigation into the natural habitat of the concept of mind underlines, that talk of the mind is a metaphorical or figurative means of giving expression to the characteristic psychological and personal attributes of human beings. And *only* of human beings (or beings that behave in sufficiently similar ways to human beings). The concept of mind could easily be dropped from these figures of speech, without rendering the expressions meaningless and without reducing their informational content. Applying the concept of mind when referring to one's own or another person's *personality, views, opinions, ideas, mood or intellectual abilities* is a convenient figure of speech, which makes our talk much more vivid and animated. In these contexts it is a harmless rhetorical device, which, however, causes havoc when it is torn out of context by philosophers and cognitive scientists, and considered in isolation from its natural habitat. Contrary to the prevailing conception among philosophers and cognitive scientists, the examples given above demonstrate that in speaking of the mind one is not referring to a singular entity. Rather,

> ' when we say that someone changed his mind, that he has a dirty mind, and that he has turned his mind to such-and-such a question, we do not imply that there is one thing, a mind, which has changed, is dirty and has been turned. Indeed the *only* thing we are speaking of is the person, and from case to case (and phrase to phrase) we are saying different things *of* the person.' [123]

The claim that the concept of mind does not denote an entity of some kind may seem counterintuitive and puzzling, and thus, may be easily taken amiss. It is important to note, that this claim does not imply that minds do not exist. Minds do exist, but they are not entities (neither material nor immaterial). These observations are of a grammatical kind, and what is argued for here, is a shift in perspective. Minds are not the kinds of objects philosophers and cognitive scientists are aiming to explain, when they ponder questions like *What is the mind?*, and *What is its nature?*. These questions are the result of a profound conceptual confusion, which surrounds the

[123] Bennett & Hacker (2003), p.105

concept of mind, and which makes philosophers and cognitive scientist embark on the investigation of a single entity, instead of turning their attention to the contexts in which the word mind is used in (Z§447). The nature of mind can only be grasped if the grammar of the concept is investigated[124].

The previous investigation has highlighted that minds are not entities. To think otherwise marks, as Bede Rundle once put it, a shift from the metaphorical to the metaphysical.[125] In thinking of the mind as a kind of entity, philosophers and cognitive scientists fall victim to the bewitchment of their intelligence by means of language (PI§109). As pointed out in the above, substantives are substance hungry. Consequently, (philosophical) questions of the form *What is…?,* should be treated with care and put our sense for nonsense on high alert.

3. Aristotelian Psychology as the Foundation of an Alternative Approach to Understanding and Investigating the Mind

Modern cognitive science and philosophy have inherited the conception of the mind as a kind of entity from Descartes. Descartes' view of the mind replaced a tradition of Aristotelian thought, which provided the first conceptual framework for the investigation of human psychology. Following Kenny (1984, 1989), the following discussion will provide an outline of Aristotelian psychology and ideas about the mind. By comparing the Aristotelian and Cartesian frameworks of thought directly, it will be possible to illustrate the misleading features of the Cartesian framework concisely, which lead to the entity view of mind. Furthermore, it will provide a powerful argument for the (re-) endorsement of the Aristotelian conceptual framework in philosophy and cognitive science, which construes the soul or mind not as an entity but as a characteristic *attribute* of (all) living beings. As Jonathan Barnes points out, Aristotle's concept of mind is

[124] Note: Language is the means by which we represent and understand the world. It originates in the socio-cultural microcosms of human experience, *viz.* their forms of life. As such language mirrors the intellectual landscape of a culture. It is for this reason that tracing the conceptual connections of the concepts of a language can reveal an entire history of thought.
[125] see: Rundle (1997), p. XXX

'at least as good a buy as anything else currently on the philosophical market. Philosophy of mind has for centuries been whirled between a Cartesian Charybdis and a scientific Scylla: Aristotle has the look of an Odysseus.'[126]

3.1 The Aristotelian Conception of the psuchē

In his treatise *De Anima,* Aristotle lays out his psychology at the centre of which lies the idea of the ***psuchē*** *(gr.)* (*anima* (lat.)). For Aristotle the ***psuchē*** or soul[127] was not a separate entity existing besides the body, but rather, was to be understood as the *characteristic powers of a given living being* (DA414^{a}20-21). This conception stands in stark contrast to the immaterial substance which Descartes had envisioned, and the glutinous substance contemporary cognitive science envisages. In an Aristotelian context, the soul, is to be understood as a *biological concept,* and not as a metaphysical one as in the Cartesian tradition. Consequently, the Aristotelian and Cartesian views represent two fundamentally different conceptions of the nature of man. For Aristotle, his conception of the soul is the principal of both animal and vegetable life (DA, 402^{a}7-8).

Aristotle believed in a hierarchy of three different kinds of soul existing in nature:

a. The *nutritive soul*, which he conceived to be the most fundamental principle of life. It is the *power* in virtue of which all biological organisms are said to have life, *viz. growth, nutrition and reproduction* (DA,415^{a}23-26).

b. The *sensitive soul*, which comprises *the powers of perception, desire and locomotion*. The sensitive soul is only possessed by higher organisms such as non-human and human animals. Plants, on the other hand, only possess a nutritive soul.

c. The *rational soul*, which is unique to human beings or creatures similar or even superior to human beings (DA,414^{b}27-28) and comprises the *powers of intellect (thought/reasoning) and volition*. Characteristically, Aristotle's conception of the (rationale) soul is also devoid of any ethical or religious connotations.

---

[126] Barnes J (2003) 'Aristotle's Concept of Mind', in: Articles on Aristotle – Psychology & Aesthetics, Barnes J, Schofield M, Sorabji R (eds.), London, Duckworth

[127] Henceforth the term soul, which is the traditional translation of the Aristotelian term ***psuchē*** , will be used to refer the Aristotelian conception of the mental.

Aristotle's conception of the soul sprang from his ideas on *matter* and *form*. The soul, so Aristotle, is the *form* of a natural body, which has life (DA, 412^{a}20). Every natural body, which has life is a *substance*, in the sense of being a composite (DA, 412^{a}15-16). The concept of substance is related to the concept of matter, as substance is an umbrella term denoting,

> '…matter, which in itself is not a this; secondly, shape or form, which is that precisely in virtue of which something is called a this, and thirdly that which is compounded of both.' (DA 412^{a}6-9)

All natural bodies including *non-living* and *living* bodies possess both form and matter. In order to describe the form of living bodies one has to describe their souls, as they are the forms of those natural bodies, which are (potentially) alive (DA 412^{a}19-21)[128]. Thus, to specify the form of a body is to say what kind of thing it is and to define the form is to state what it is to be that kind of thing.

The distinction between form and matter provides the basis for Aristotle's understanding and description of persistence through change. As Bennett and Hacker point out,

> 'Aristotle introduced the notion of *matter* as a technical term to pick out that which has a capacity for substantial change, and *form* to pick out that which makes a certain matter into the kind of substance it is.' [129]

It is in this way that Aristotle came to conceive of the soul as the form of the living body. At the same time, Aristotle also thought of the soul as the first actuality of the living body, *viz.* the unexercised dispositional power of an animal (DA412^{b}6-7). In contrast, Aristotle refers to the exercise of a dispositional power as a second actuality. In general, the actualities of a substance are both the things a substance is and the things a substance can be engaged in (e.g. the things the substance is doing) at a certain time. Aristotle counts the powers of a substance, which includes both the active and the passive potentialities the substance possesses, among the

[128] It has been pointed out that as far as Aristotle was concerned, the Greek language already had a term, ***psuchē***, to denote the forms possessed by living bodies (a subclass of natural bodies). (see: Everson S 'Psychology', in: *The Cambridge Companion to Aristotle, Barnes J ed., Cambridge University Press, Cambridge, 1995*).

[129] Bennett & Hacker (2002) ,The Motor System in Neuroscience: A History and Analysis of Conceptual Developments', *Progress In Neurobiology 67:1-52*

substances´ actualities[130]. It is important to note that Aristotle does not want the soul to be understood as being a part of the body or one thing with the body. This, he remarked, would be like thinking that the wax is identical to the imprint on it (DA412^{b}8-9). Rather, soul is in general

> 'substance in accordance with the account of the thing. And this is the `being what it was´ for a body of this kind. In the same way, if some tool, say an axe, were a natural body, its substance would be being an axe, and this then would be its soul. And if this were separated from it, it would not continue to be an axe, except homonymously, whereas as it is, it is an axe.' (DA412^{b}13-17)

That is, the axe derives its power to chop from its being constituted by the matter of iron and wood formed into matter and handle. The *form* of the axe is its power to chop. The power to chop wood is the first actuality of the axe.

This short exegesis illustrates how Aristotle came to conceive of the soul as an attribute of a living thing, *viz.* its characteristic powers, rather than an immaterial substance as envisioned by Descartes.

> '…substance is, as we said, spoken of in three ways, as form, as matter, and as the composite, and of these matter is potentiality, form actuality, and since the composite is in this case the ensouled thing, it is not that the body is the actuality of the soul but that the soul is the actuality of some body. And for this reasons they have supposed well who believed that the soul is neither without body nor a kind of body.'

In conceiving of the soul as the powers of a living being, rather than as an entity existing independent of the body, Aristotle's ideas stand not only in opposition to the conception of mind which was offered (much later) by Descartes but also to the teachings of his contemporary Plato. Both Plato and Descartes thought of the soul as an entity enjoying an existence separate from the body. Consequently, in contrast to the dualist Platonic and Cartesian psychological frameworks, Aristotle never runs into the difficulty of having to answer questions with regard to the relationship between mind and body. Within the Aristotelian conceptual framework, there is no room for a mind-body problem. Moreover, as Aristotle does not conceive of the soul as and entity, his framework of thought does not invite violations of the logic of

[130] see: Bennet &Hacker (2002) ibid., p.3

whole/parthood relationships. Conceiving of the soul as the powers of a living being does not allow for attributing to the soul the distinctive powers of the living being whose soul it is. Aristotle's urges us

> 'not to say that the soul pities or learns or thinks but that the man does *in virtue of his soul* ' (DA 408b13-15)

This aspect highlights one of the distinguishing features between classic Cartesian and contemporary ideas about mind encountered in cognitive science. Within these traditions of thought, psychological attributes are ascribed either to the mind (classic Cartesianism) or the brain (modern cognitive science) (see chapter III). The Aristotelian conception of the mind on the other hand, is immune to these kind of mereological errors, and can be a powerful antidote to transgressions of the bounds of sense.

3.2 Aristotle – From a Contemporary Point of View

Anthony Kenny has adapted and developed the Aristotelian conception of the soul in a modern context. In accord with the Aristotelian conception of the soul as the characteristic powers of a living being Kenny suggests that if one is to give a definition of mind at all, it is best to think of the mind as a second order ability, *viz* the *capacity to acquire intellectual skills*[131]. Understood in this way, the mind is the capacity[132] that enables human beings to acquire those skills and abilities, which form

> 'the basis to engage in behaviour of the complicated and symbolic kinds which constitute the linguistic, social, moral, economic scientific, cultural, and other characteristic activities of human beings in society.'[133]

[131] See: Kenny (1989), p.3ff
[132] Note: Although, the following will emphasize the conception of mind as a cognitive capacity, following Aristotle's conception of the rational soul, the mind is not only a cognitive capacity (i.e. comprised of intellect), but also a volitional „capacity" (i.e. comprised of will). In this context, the human will is to be understood as the ability to act for reasons. Its freedom derives from the special characteristics of practical reasoning.
[133] Kenny (1989), p.7

The most important skill, which human beings possess, is the knowledge and command of language, as it is through our language ability that we acquire knowledge of other abilities and learn to master and exercise those abilities. To have knowledge of something, simply is to be able to do something. An important distinction that is to be made in this context is that between ability, the possessor of the ability, and the vehicle or mediator of an ability. I am the *possessor* of my ability to dance, my ability to play guitar, and my abilities to think, imagine, memorize and remember. I am the one who has the ability to do all these things and more. Of course, it is not only human beings who possess abilities of a certain kind. My laptop has the ability to perform calculations with hyper-complex numbers twice as fast as my old machine. Similarly, *Neuhauser Zoigl*[134] can intoxicate you about twice as fast as normal beers can, because of its extraordinary potency. The *vehicle* of my laptops ability to perform very complex calculations very quickly is the processor inside the machine. The vehicle of the *Neuhauser Zoigl`s* ability to induce intoxication is the amount of alcohol contained in the beer. Finally, the vehicle of my ability, to dance, to play guitar[135], to think, to imagine, to memorize and to remember is the brain. As these examples illustrate, an ability is essentially something intangible. It is, as it were, *an abstraction from behaviour*[136]. In contrast, the vehicle of these abilities depicted above is always something solid and concrete. The vehicle of my new laptops' ability to perform calculations with hyper complex numbers twice as fast as my old machine is the processor inside my new laptop. Similarly, in the case of human beings, the vehicle of their intellectual capacity is the brain. One advantage of this conception of mind is that it provides an unambiguous guideline for empirical investigations:

> 'If one is to say what each of these is, for instance what the capacity for thought is, or for perception or nutrition, one should first say what is thinking and perceiving.' (DA 415a16-17)

That is, in order to understand what the ability to φ is one needs to understand what it is to φ. In other words, in order to gain an understanding of the ability to φ one has to

[134] Zoigl is an unfiltered, bottom fermented type of beer, and has the status of a Bavarian Delicacy. Its production is restricted to certain areas of the upper palatinate (Oberpfalz), where it is brewed by private individuals in possession of special legal permission to brew.
[135] Note: It is assumed that both learning how to dance and learning how to play guitar involves a variety of intellectual skills proper such as , thinking, memorizing etc.
[136] Kenny (1984), p.28

study the exercise of that ability. And in order to study the exercise of the ability to φ, one needs to have an understanding of the symptoms and criteria adjudicating the use of the concept of φing, i.e one needs to know what counts as φing .

The conception of the mind as a second order ability (i.e. capacity), and of the brain as the vehicle of this second order ability, has important implications for cognitive science. First, it should put a stop to cognitive scientists and philosophers to adopt the entity view of the mind, which in the past has lead to the concoction of imaginative but logically-flawed (i.e. nonsensical) theories of mind-body interaction (e.g. dualism, behaviourism, mind-brain identity theory, functionalism etc.). In the wake of the entity view of the mind a host of questions arose, which have plagued philosophers and cognitive scientists for years. The infamous mind-body problem or the problem of the explanatory gap[137] are but two prominent examples. However, the preceding discussion of the concept of mind gives rise to some considerable doubt about their status as (philosophical) problems. It has shown that talk about the human mind does not refer to a kind of *thing*, but rather to constitutes an idiomatic way of talking about the characteristic powers of human beings, and their exercise. The question 'what is the mind?', mistakenly led philosophers and cognitive scientists to view questions about the relationship between mind and body as problem in its own right; a problem, to which there is a solution if one only probes hard enough. With the benefit of an overview of the concept of mind at hand, however, we can see how false grammatical analogies and simplistic questions have led to the illusion of a problem. Consequently, there can be no solution to the mind-body problem. Rather, our informed perspective regarding the concept of mind, should prevent us from granting this age old philosophical problem e very status of a problem. The mind body-problem, simply disappears. Second, it prevents cognitive scientists and philosophers from falling victim to the "substance hunger" of questions like "What is the mind?", which in turn leads to misconceived definitions and explanations of the workings of the mind as constituted by the cognitive view. Wittgenstein points out, for example, that questions like,

[137] It is commonly believed that human experience cannot be fully explained by the mechanical processes implied by reductionist conceptions of the mind such as the cognitive view, i.e. there seems to be a gap between mechanics of the brain and the experiential quality of our emotions, sensations etc. Consequently, it is thought that something extra, maybe a substance/entity of different metaphysical kind, must be added to "fill the gap", as it were.

‘”Die Fragen, Was ist Länge?”, „Was ist Bedeutung?“, „Was ist die Zahl Eins?“ etc., verursachen in uns einen geistigen Krampf. Wir spüren, daß wir auf nichts zeigen können, um sie zu beantworten, und daß wir gleihwohl auf etwas zeigen sollten. (Wir haben es hier mit einer der großen Quellen philosophischer Verwirrung zu tun: ein Substantiv läßt uns nach einem Ding suchen, das ihm entspricht.) (BB_D, p.1)

‘"Questions like What is length?", "What is meaning?", "What is the number one?" etc., produce in us a mental cramp. We feel that we can't point to anything in reply to them and yet ought to point to something. (We are up against one of the great sources of philosophical bewilderment: a substantive makes us look for a thing that corresponds to it.’ (BB_E,p.1)

On the other hand, adopting an Aristotelian position by viewing the mind as the ability to acquire intellectual abilities illustrates, what kind of insight empirical investigations into the human mind can yield: Thus, neural events and processes can be correlated with the exercise of a human beings ability to think, perceive, imagine etc. Moreover, cognitive scientists can try and discover what neural states, events and processes are empirical necessary conditions for human beings to be able to exercise their intellectual abilities. Third, if minds are capacities, they not only are not entities, but most importantly, they can’t be *physical* entities. If it is wrong to think of the mind as a physical entity, it is wrong to conceive of the mind as biological information processing unit. By now, it should be abundantly clear that this view does not constitute so much a particularly insightful conception of the mind, but rather a piece of sophisticated scientistic metaphysics; the result of a gross lack of conceptual clarity. While one would hardly think of a car’s ability to accelerate from 0-100km/h in 4,1 seconds as a material entity, mind-brain dualists do so consistently in their investigations. Finally, in the Aristotelian tradition, there is no room for talk about an inner and an outer. After all, it would make little sense to conceive of my ability to remember the colours of the rainbow, of something inner. What would the outer be?

4. Representations and Information Processing

The previous section has exposed the serious misconceptions and confusions, which underlie the entity view of the mind. The plausibility of accounts of human and animal behaviour in terms of the information processing view of the mind, i.e. the view of the mind as a biological information processor, seems highly questionable in

the light of the preceding investigations and the Aristotelian alternative. However, the entity view of the mind is not the only questionable idea that has found its way into the body of basic tenets and fundamental premises, which make up the cognitive view. Even if one was prepared to dispose of the entity view of the mind, and accepts that the exercise of our intellectual abilities is correlated with certain neural events and processes, cognitive scientists and philosophers may still insist that these neural going ons constitute (mental) representations (of the external world), by means of which the brain constructs a map of the world in order to plot its course of action[138]. In order to thwart such ideas, the present section will investigate the idea of *mental representations*. The meaningfulness of the notion of internal mental representations, conceived of as bundles of information derived from the exercise of our perceptual input, will be subjected to a conceptual analysis.

The contemporary conception of mental representations as physical symbols is a rather recent development. While the classical representationalists like Descartes and Locke, for example, assumed that the end product of the exercise of our perceptual capacities is a *pictorial representation*[139], contemporary representationalists championing the cognitive view conceive of the result of the exercise of our perceptual capacities as a *symbolic representation*. In their seminal article *Cognitive Psychology and Information Processing*, Lachman, Lachman and Butterfield (1979) emphasised the importance of the computer analogy to the study of the mind and mental phenomena, defining the subject matter of cognitive psychology as being, about how people take in information, how they recode and remember it, how they make decisions, how they transform their internal knowledge estates, and how they translate these states into behavioural outputs. Accordingly a person driving a car through the country side must

[138] Bechtel and Graham (1999), p.48ff

[139] Famously, Descartes wrote, *'If we see some animal approach us, the light reflected from its body depicts two images of it, one in each of our eyes, and these two images form two others, by means of the optic nerves, in the interior surface of the brain which faces its cavities; then from there, by means of the animals spirits with which its cavities are filled, these images so radiate towards the little gland which is surrounded by these spirits, that the movement which forms each point of one of the images tends towards the same point of the other image which represents the same part of this animal. By this means the two images which are in the brain form but one upon the gland, which, acting immediately upon the soul, causes it to see the form of this animal.'* (See: Descartes 1984)

> ‘have represented the landmark’s appearance in *(his)* memory, *(and permanently)* match up *(his)* current perception of the landmark to its stored representation.’[140]

Similarly, in the context of the study of vision it has been argued that if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information[141]. This dictum implies that

> ‘there must be symbols inside our heads for the things we see symbols which themselves are unlike the things they represent. Upon opening up a patient’s head for a brain operation, the surgeon does not find there a miniature stage-set of the world! All he finds there is a pink blancmange-like mass of brain cells. So it is an inescapable conclusion that there must be a symbolic description in the brain of the outside world, a description cast in symbols which stand for the various aspects of the world of which sight *(and the other senses)* makes us aware. In fact, when we began by asking ‘What goes on inside our heads when we see?’ We could as well have put this question as ‘When we see, what are the symbols inside our heads that stand for things in the outside world?’…The idea of visual experience as a symbolic process may seem a strange one. The likely reasons for this is that the world we see…is so very clearly ‘out there’ that it can come as something of a shock to realize that somehow the whole of this world is tucked away in our skulls as an inner representation which stands for the real outside world.’ [142]

It would be wrong to regard these remarks as isolated instantiations of the representationalist ideas of a few individuals, working in small restricted areas of cognitive science. Rather, they manifest some of the most influential articulations of the modern representationalist stance, which have contributed to the fact that representationalist ideas very much constitute the orthodox view among cognitive scientists:

> ‘The principle function of the central nervous system is to represent
> and transform information and thereby mediate appropriate decisions and behaviours.
> The cerebral cortex is one of the primary seats of the internal representations maintained
> and used in perception, memory, decision making, motor control, and subjective
> experience…’[143]

---

[140] Lachman R, Lachman J, Butterfield E (1979) Cognitive Psychology and Information Processing, *Hillsdale New Jersey, Erlbaum,* p.7
[141] Marr (1982), p.3
[142] Frisby JP (1979), Seeing, Illusion, Brain and Mind, p.8f
[143] de Charms RC, Zador A (2000) Neural Representation and the Cortical Code. *Annual Review of Neuroscience 23:613-647, p.613*

The question, which will be investigated in the following, will be whether it really makes sense to conceive of mental representations as symbolic descriptions of the outside world in the brain. To settle this question we have to examine the ways in which the concepts of *description* and more specifically, *symbolic descriptions* are used.

4.1 Descriptions and Symbolic Representation

Descriptions, in the ordinary sense of that term, are strings of words or symbols and take the form of an expression or proposition, which may depict the characteristic features of a certain event (e.g. my account of the last Guns'n'Roses show at Docklands Arena) or a certain state of affairs (e.g. that the relationship between the Bush administration and 'old' Europe is highly strained). Furthermore, the proposition or expression can also be a description of the features of an object (e.g. Degas´ painting *The Blue Dancers* is the perfect depiction of female grace and beauty without being chauvinist, and it reminds me of Cécile). The descriptions of the various features can be objective or subjective, they can be accurate or inaccurate, be very detailed or only sketchy and they can be true or false. As such, descriptions are on a par with other forms of expression and speech used in ordinary discourse such as excuses, apologies, recommendations, justifications, declarations, pleas, oaths, questions and answers, for example. The concept of description is sometimes also applied in a wider context, however. For example, in certain cases a pencil sketch of a part of Oxford can serve as a map, i.e. as a description of the way to my house or college. Similarly, a blueprint may be a description of the layout and dimensions of a house, a plan may be a description of a process to take action or how to find a hidden treasure on an island, and a diagram retrieved from a *Kinder* egg may be a description of how to construct the surprise toy contained inside the chocolate egg. If one goes beyond these contexts, however, the concept of description starts to loose grip. Sometimes, as in the study of visual processes pictorial representations like images, e.g. pictures or photographs, are thought of as descriptions. However, my picture of Cécile and my photograph of her dancing the Waltz cannot properly be said to be descriptions, although descriptions of the picture or the photograph may well include a description of what it is a photograph or picture of. But, my picture of Cécile doesn't describe her, it depicts her, and my photograph

of Cécile waltzing doesn't describe her dancing, it shows her dancing. But although pictures or photographs may serve the same *purpose* as descriptions, they could not be said to be symbolic descriptions themselves. A picture of a bullet framed with guns and roses may be a symbol of *Guns'n'Roses*, but it is neither a description of the band nor their music. Thus, in the ordinary sense of description, or symbolic description, their expression may take a written symbolic form or a spoken verbal form. These kinds of descriptions can be found in articles, newspapers, books, on the world-wide-web, or one may hear them presented on television, radio or at the cinema. But in this normal, ordinary sense of the term description, they are not found inside a person's brain. Thus, contrary to the claims of John Frisby, it is not an "inescapable" conclusion that there must be a symbolic description in the brain of the outside world (see above).

At this point (as in the case of the mereological fallacy discussed in chapter III), it might well be objected that cognitive scientists were applying the concept of a symbolic description in an extended or figurative sense. After all, scientists must not be refused the right to introduce new terminology or extend the use of existing terminology should scientific investigation demand it in order to gain a new perspective and novel understanding of a problem or question. Consequently, it seems natural to conceive of neural activity, which is elicited by the appearance of a stimulus in one's visual field, as a description of the stimulus or its underlying cause. But this is mistaken. Neural activity, i.e. the firing of nerve cells within a specific neuronal circuit in the brain, cannot be said to be a form of symbols, as for something to qualify as a symbol, it must have a *rule-governed use*. For something to be a symbol there must be correct and incorrect ways of applying it (i.e. the concept of a symbol is correctly only applied to those forms of expression to which a standard of correctness can be applied). The neurons of the brain, however, do neither know n do they not know what any array of symbols means. Agents can be said to employ symbols only in so far as they can be said that they are aware of the rules for the use of the symbols they use, i.e. only if they are aware of the grammar of the symbols, which is given by the explanations of the use of the symbols. Neurons on the other hand, do not know any grammar they could deploy in order to provide descriptions of stimuli in the outside world. It constitutes a transgression of the bounds of sense to think of patterns of electrical activity in the brain as descriptions of features of the (outside) world, which can be correct or incorrect, accurate or inaccurate, objective or

subjective. From the observation that our perceptual faculties allow us to know what is where in the world, it does not follow that this information *must* be represented in the human brain. In fact, there is no such thing as the brain *representing information* in the ordinary sense of that expression. In the context of the Aristotelian view of the mind, it would not be false to postulate that certain neural activities can be thought of as representations of, e.g. orientation or depth, if all this means is that one can correlate certain neural firings with certain features in one's visual field. However, the received conception in cognitive science goes much further than that. In the context of the cognitive view mental representations are bundles of information, i.e. formal systems for making explicit certain entities or types of information. Together with a specification of how the system does this, the result of using a representation to describe a given entity (is) a *description* of the entity in that representation[144]. The incoherent idea underlying this line of thought is that there is a description of what one sees (perceives) inside one's head, which rests on the notion that the brain must be capable of representing information by forming internal *mental* representations. As a result of this, representation is conceived of as a method of describing something by means of the brain's rule-governed physical symbol system. David Marr, for example, writes that

> 'a representation…is not a foreign idea at all – we all use representations all the time. …the notion that one can capture some aspect of reality by making a description of it using a symbol and that to do so can be useful seems to me to be a powerful and fascinating idea.'[145]

According to this statement, however, it is not a new and extended sense in which the concept of representation is used. Rather, the use is its customary common one. Representations here are not the causal correlates resulting from the exercise of our perceptual abilities, but the symbols of a system the grammar and meaning of which is determined by rule-governed use. This however, constitutes a transgression of the bounds of sense.

[144] see e.g.: Marr D, (1980) Vision, p.20f
[145] ibid.,p.21

4.2 Representations and Maps

In cognitive science the concept of representation and the concept of a map or mapping often occur in related closely linked contexts. For example, JZ Young argues that

> 'the nerve cells that analyse the information of the senses are laid out on the brain to make actual physical maps of the surface of the body or retina. What goes on in the brain must provide a faithful representation of events outside, and the arrangement of the cells in it provides a detailed model of the world….the information from the sensory surfaces of the retina, or skin, or from the ear is laid out in a topographically precise way on the surface of the brain. Moreover for each such sense there is a series of such maps, each recombining in a new way the words of information provided by the cells. So the grammar of this language has something to do with spatial relations. It communicates meanings by topological analogies'[146]

By implication, the information contained in these maps is then used by the brain to formulate hypothesis as to what is where in the world, hypothesis which in turn provide the basis for human action[147]. The problem with the idea of maps in the brain, which represent information perceived by our senses, is not altogether unfamiliar as it runs into similar difficulties as the idea of there being symbolic representations in the bain. Maps, in the ordinary sense of the term, are tools generated by cartographers, geologists, astronomers etc., to depict an array of features ranging from the topography of a piece of land or the demographic distribution of the population of a given town or country, to the location and relative distances of the stars visible in the night sky. In this sense, maps are indeed systems of representation. But it is important to note that such systems of representation imply the agreement upon and the usage of conventions, i.e. rules of representation. Without these, there are no maps representing anything. And, as in the context of the notion of symbolic representation, there exist no conventions of representation, which are not vindicated by their intentional use by creatures of sufficient intelligence, and which makes them able to apply these conventions (of course, this can be done either correctly or incorrectly). But as pointed out in the previous section, neither brains nor neurons can be said to employ symbols (or maps for that matter) as they neither know nor are

[146] Young JZ (1978) Programs of the Brain, Oxford, Oxford University Press, pp.11, 57
[147] ibid.p.60, see also: Phillips CG, Zeki S, Barlow HB (1984) Localization of Function in the Cerebral Cortex, *Brain 107:338-347* p.345

ignorant of what the symbols or maps mean. Similarly, both brains and neurons can neither be said to follow nor can they be accused of failing to follow rules for the use of symbols. The point here is that it is not an intrinsic feature of a pattern of neural activity in the cerebral cortex whether it is a map or not. It is a conventional one. Thus, certain features or stimuli in ones sensory field can be mapped (i.e. causally correlated with) onto the firings of cortical neurons as long as nothing more is implied than to indicate an inductive correlation. If, however, talk of mental maps implies the neural "communication of meanings by topological analogies", the bounds of sense have once more been violated.

4.3 The Storage of Representations

Another questionable idea, which pops up again and again in discourse about mental representations, is that mental representations are supposedly *stored* in the brain. For example, when we drive a car or take a walk our brains match up and compare the current representation of our environment with the one stored and by doing so direct and re-direct our behaviour (see above). This misconceived notion has a long tradition within philosophy and has been propounded, above all, by empiricist philosophers. Although, the confusions regarding the notion of storage are manifold and run criss cross in all directions (in particular, they are closely associated with the concept of memory) the following will concentrate only on those aspects of the problem which are of immediate relevance to the representationalist conception under scrutiny.

It is tempting to assume that unless what one perceives is stored in one's brain in the form of a symbolic representation one would not be able to direct and re-direct one's behaviour: It seems to be the perfect mechanism enabling human beings to constantly adapt their actions to changes in their environment. But apart from the fact that neither brains nor neurons use symbols (see above), the idea of a neural storage of representation in a symbolic or semantic sense *per se* does not make sense. The idea of a stored representation makes only sense if the representation can be accessed and is available *to a person* which could read and recognize the representation and potentially tell somebody what it is a representation of. Furthermore, it is a person who *does* the storing. This is how the concept of storage would normally be applied. I can pack my books in a box and store the books in the

room of a friend while I am leaving college for the summer. The Tate Modern can store the paintings and other works of art currently not under exhibition in its cellar. Similarly, my friend Paul might store the plan for the guitar he is building for me (which can be said to be a representation of the dimensions and layout of my future instrument) in a cupboard along with the other plans and blueprints of the instruments he is working on. And on opening his cupboard in the morning, Paul discriminates between the various plans and blueprints and chooses the one he wants to work on that day (e.g. in doing so he might recognize the plan of my guitar as the one in the right upper corner of the cupboard). But neither brains nor neurons can be said to be doing any of this, nor could anyone put a piece of paper into my head for it to be stored in my brain. Importantly, neither brains nor neurons are in the business of storing things. The only sense in which it can reasonably be said that there is a neural representation is in a *non-symbolic, i.e. non-semantic,* sense. In this sense a neural representation would be a causal correlate of the exercise of one's perceptual capacities. But exercising my ability, to see, hear, smell, etc. has got nothing to do with the storage of a representation, let alone with encoding the information contained or conveyed by a representation. Unfortunately, this is not how the idea of representation and storage is applied in cognitive science. Cognitive scientists, have yet to specify the criteria of what the storing of a landscape representation, for example, would look like. They have not provided us with a novel rule for the use of the concepts of storage and representation in the context of the brain. Instead, they use the concepts of representation and storage in their customary way and in doing so unwittingly transgress the bounds of sense.

4.4. Some Clarifications regarding the Notions of Mental State, Process, Activity and Event

By now it will not be surprising to find that like the ideas of (mental) maps and storage of representations the views about the nature of mental states, processes, activities and events are also closely associated with the conception of mental representations. For example, cognitive scientists think of mental states as being characterized what the internal representations currently specify, and of mental processes by how the representations are obtained and how they interact. From the preceding discussions it should be clear that these views are nothing but inevitable extensions of the confused ideas underlying the reductionist conception of mind

endorsed by contemporary cognitive scientists and philosophers. By effecting the collapse of the foundations of the cognitive view, one also effects the collapse of its peripheral architectural ornaments. If the idea of symbolic mental representations is a chimera, then surely so must be the idea of a mental process as the interaction of symbolic mental representations. However, for the sake of conceptual clarity, it is still worthwhile to gain an overview of the conceptual landscape covered by the concepts of (mental) state and (mental) process.

When applied according to its customary use, the concept of state denotes, at its most general, a combination of circumstances or attributes, which pertain for a certain amount of time to a person or thing. As such, states may be temporary and short-lived, or they may persist for a considerable period of time. My room, for example, may be in a state of disorder, but only as long as I chose not to clean up my mess. Similarly, my friends and I are in a state of excitement as we wait for the final of the Professional Latin Championships at Blackpool to begin, and we find ourselves in a state of confusion when Michael and Beata are not recalled to dance the Paso Doble. Furthermore, the concept of state is also applied in relation to being in a certain stage or form, as relating to structure, growth, or development, e.g., to be in an embryonic state or fetal state or economic state, and it is applied in relation to the condition of a physical system, e.g. its phase form, composition or structure as in “Ice is the solid state of water”. Finally, the concept of state can also refer to a specific mode of government, like in the welfare state, the communist state, or a federal state. The concept of state contrasts with the concept of a process to the extent that while states consist of patterns of relationships between elements pertaining at a time, processes denote dynamic goings on. These go on for a certain length of time, and normally consist in a sequential transformation of states. Thus, we can speak both of economic, political or physical states and economic, political and physical processes. The processes involve the action or activities of people or substances that bring about the change underlying the transformation of one particular state into another (hot to cold, sweet to sour, bitter to sweet, rich to poor, stable to instable, liquid to gaseous). As such, they may be the result of people exercising their intellectual powers, or they may be the result of the powers of certain substances or processes.

It is important to note that when these concepts are moved from their natural habitat, as outlined in the preceding overview, and applied in the context of

psychology and talk of the mind subtle modulations in the meanings of these concepts follow in the wake. The notion of a *mental state* is primarily applied in association with talk about emotions and moods. We speak of somebody being in a state of anxiety, depression, love, happiness, despair, excitement, agitation, enthusiasm or tranquillity, and beyond that, also of someone being in state of intense concentration or fatigue. Such states of mind normally only persist for a certain time, and they are distinguished by the tendency to display various characteristic patterns of behaviour in certain circumstances. These behavioural patterns and dispositions of people to behave in a certain way, are the *criteria* for the respective mental state. Furthermore, emotional states are directed towards a certain object (i.e. I dislike Paul Killick's dancing, I am in love with Cécile etc.), whereas moods, on the other hand, usually lack such object directedness. However, in all these cases, it would be confused to invoke the idea of internal representations in order to account for the intentionality and object directedness of these states. Mental states are not characterized by asserting what the internal representations currently specify (see Marr). Rather, they are characterized by a variety of behavioural criteria, which specify the emotion, or mood a person is experiencing (see also chapter III).

In contrast to the concept of a mental state the concept of a mental process is mainly applied in talk and expressions indicating goings on in the cognitive domain. Importantly, the notion of mental processes does, unlike the notion of non-mental processes, not involve the notion of a sequential transformation of states although the notion of development is retained. One may conceive of thinking through a problem or formulating a plan in one's head as a mental process, as one is thinking first of this, then of that and then of that before arriving at a conclusion (importantly, this does not involve a transformation of states). The idea of a mental process may also intelligibly be invoked in the context of loss or grief and other painful coming to terms. For example, overcoming a broken heart may be a painful process, as may be the attempt to go on living and overcome the death of a loved one. But, it would be mistaken to conceive of a mental process as a matter of interaction between and manipulation of various internal mental representations, as in such cases one has ventured from the intelligible to (despite appearances) the unintelligible.

5. Concluding Remarks

The present investigation has revealed the deceptive simplicity of the questions 'what is the mind?' and 'what is the nature of mind?'. They imply a picture of mind as a concrete entity and obscure the fact, that the mind is best thought of as the wide range of characteristic human (intellectual) powers and their exercise and the extensive repertoire of human character traits. It is only after one has gained an overview of the concept of mind that the temptation to look for something which corresponds to it weakens, as the increasing understanding of the grammar of the concept gradually reveals the true nature[148] of mind. A glance at the history of philosophy and cognitive science shows that both Cartesians and materialists have been taken in by these questions and consequently adopted an entity view the mind. The mind-body problem and the problem of the explanatory gap are but two examples, of the confusion this notion has caused. These problems are inevitable consequences following in the wake of adopting the cognitive view. Yet, as the entity view of the mind collapses, so do these questions, and their true nature is revealed: they are mere muddles, which have been elevated to mysteries. One general lesson to be learned from this discussion is, that in philosophy *nouns* are to be treated with care. They are *substance hungry*, and without a clear overview of their use tempt us into hypostatisation, i.e., in the case of the mind the ascription of a concrete reality to a linguistic chimera.

On the other hand, viewing the mind as the ability to acquire intellectual abilities illustrates what kind of insight empirical investigations into the human mind can yield: Neural events and processes can be correlated with the exercise of a human beings ability to think, perceive, imagine etc (if conceptual clarity with regard to these concepts has been achieved before the start of an empirical investigation, of course). Moreover, cognitive scientists can try and discover what neural states, events and processes are empirical necessary conditions for human beings to be able to exercise their intellectual abilities. Thus, if minds are capacities, they not only are not entities, but most importantly, they can't be *physical* entities. And naturally, if it is wrong to think of the mind as a physical entity, it is wrong to conceive of the mind as biological information processing unit.

[148] Or *essence* (See: PI§371)

In addition, the preceding discussion has demonstrated that the misconceptions about the nature of mind extend far beyond its conception as a biological entity. The idea of the mind as a biological computer, which operates on the basis of symbolic mental representations to provide the mind with descriptions of the "outside" world is misconceived, as there the mind/brain can not be said to know what an array of symbols means or how to apply them according to a given standard of correctness. For something to count as a symbol it must have a rule governed use, i.e. there must be correct and incorrect ways of applying it. However, the only way to make sense of the notion of mental representations is to think of them as correlations between neural firings and certain features in one's visual field, for example. For similar reasons, the conception of mental representations as descriptions inside one's head as suggested by the cognitive view, does not make much sense either, nor does the idea that maps in the brain (i.e. mental maps) convey meanings by topological analogies. Maps are representational systems and, like symbols, imply the application of a system of conventions, which the brain cannot be said to make use of. Finally, the present chapter also investigated the notion of storing representations in the brain, a central assumption within the context of the cognitive view (see above). Yet, as the investigation demonstrated the notion of a stored representation can only make sense if the representation can be accessed and is available *to a person*, which could read and recognize the representation and potentially tell somebody what it is a representation of. Cognitive scientists and philosophers have yet to specify, however, the criteria for identifying what counts as storing a landscape representation, for example, in the brain. As they use the concepts of representation and storage in their customary way when talking about the storage of representations in the brain, however, they unwittingly transgress the bounds of sense.

Chapter V

On the Nature of Thinking and the Cognitive Conception of Thought

1. Preliminary Thoughts

What is thinking? What is the nature of thought? By asking these questions, human animals characteristically mark themselves out as thinking beings. It is interesting to observe, however, that although these questions seem to arise naturally in the course of a "thinking life", as it were, even the most feeble attempt at providing an answer reveals, that there is something slightly queer about them. On the one hand, thinking never strikes us as mysterious when we think. On the other, it regularly does so when we start to reflect upon our thinking (PI§428). The experience one undergoes when one tries to give an account of one's understanding of thinking is not unlike the experience one undergoes when trying to provide an account of time. Augustine famously remarked that when nobody asks us, we seem to know perfectly well what time is, but as soon as we are asked about it, we are unable to give an account of it. The same seems to hold true for thinking. When being asked to provide an account of thinking, one tends to struggle, as it appears seems to possess some sort of elusive quality, which is difficult to capture in words. In this respect, thinking is strongly reminiscent of our musings on the mind (see chapter IV). On reflection, thinking (or thought) sometimes appears to be something immaterial, as when it operates on images, for example, while at other times, when it is framed in words, it seems to be a substance of some kind; yet in any case, it seems to be a complex process occurring in the mind/brain, which operates in a mysterious way (Z§125), as when a thought of Pimm's and Lemonade suddenly strikes me unannounced whilst writing this chapter, or when my inability to stop thinking about Cécile keeps me distracted from writing my thesis. In these cases, thought appears beyond our control, and it looks as if our actions are at the mercy of some enigmatic process. Ideas like these not only pervade the scientific and philosophical study of thinking, but are also present in contemporary popular culture. For example, we find the headmaster of *Hogwarts School of Witchcraft and Wizardry*, Albus Dumbledore, retrieve his thoughts, i.e. a silvery glistening, misty kind of substance, from his

temples for storage in a pensive[149]. But what are we to make of this host of views and opinions regarding the nature of thought? How are we to understand thinking? What is thinking? What is the nature of thought?

At first glance, the cognitive view promises rescue from this quandary. It *seems* to provide insight in the form of an elaborate theory, which many contemporary cognitive scientists and philosophers believe to constitute the most plausible explanation of thinking:

> 'Thoughts are inner representations; thinking is the processing of inner, mental representations. These representations have a double aspect…their role within the mind depends on their…syntactic properties…and they are representations in virtue of relations with the world. Mental states represent in virtue of *causal* relations of some kind with what they represent.'[150]

According to this *cognitive conception of thinking*[151], thoughts are *mental representations* and thinking is to be understood as a *mental process,* i.e. as the processing of mental representations. However, there is reason to be deeply suspicious about this explanation. First, because these ideas are an inevitable extension of the information processing view of the mind, the significant flaws of which have been exposed in the previous chapters. Even if one was to assume that the flaws of the latter do, despite the obvious parallels, not impinge on the validity and sense of the cognitive conception of thinking, due diligence nevertheless dictates a thorough investigation of this particular account of thinking and the nature of thought. Second, from a Wittgensteinian point of view philosophical theories of any kind should be treated with the utmost suspicion (see chapter II). As Peter Hacker points out, it is a distinguishing mark of philosophers (and nowadays also of cognitive scientists) that after having tied a knot in their musings about thinking, they respond to the knot by constructing a theory.[152] Due to a misguided belief in the powers of scientific technology, and a general unawareness of the distinction between factual and conceptual questions, it is readily assumed that the mystery and puzzlement surrounding thinking can be resolved by devising a sophisticated theory of the nature

[149] See: Rowling JK, (2000)
[150] Sterelny (1990), p.39
[151] I.e., the conception of thought inherent in the cognitive view.
[152] Hacker (1997b), p.144

of thought[153]. Yet, the features of thinking that make it seem mysterious are not amenable to empirical investigations, as they are the product of conceptual confusion and grammatical misunderstanding (PI§110) (see also: chapter II):

> ‘Das Denken ist ein rätselhafter Vorgang, von dessem vollen Verständniss wir noch weit entfernt sind.” Und nun stellt man Experimente an. Offenbar, ohne sch bewußt zu sein, *worin* das rätselhafte des Denkens für uns liegt. ...’ (RPPI§1093)

> ‘Thinking is an enigmatic process, and we are a long way off from a complete understanding of it." And now one starts experimenting. Evidently without realizing *what* it is that makes thinking enigmatic to us. …’ (RPPI§1093)

As in the case of the human mind, the explanation of thinking which the cognitive view offers, does not constitute so much a profound insight, as it does betray deep seated illusions resulting from a distinct lack of conceptual clarity. Wittgenstein, used to warn his students of *“What is…?”* questions as characteristic utterances of puzzlement and unclarity. The mental discomfort expressed in such questions is, he held, comparable with the question "Why?", often asked by children. Importantly, while these "Why?" questions are (also) an expression of a mental discomfort, they do not necessarily ask for a cause or a reason (BB26). Similarly, “What is…?” questions (usually) do not ask for definitions and explanations, but for grammatical clarification.

It has already been indicated that the account of the nature of thought as symbolic representations, and of thought processes as interactions of such representations, is a quasi-inevitable step, which follows in the wake of endorsing the cognitive view. Within the course of the present chapter, this account of the nature of thinking will be subjected to a thorough investigation in order to show, once again, how inattention to conceptual hygiene, can lead to nonsensical conclusions. Although the previous discussion of the entity view of the mind, should have robbed the association of the brain (and thus by implication the present account of thinking) with a physical symbol system much of its allure, the misconceptions regarding the nature of thinking are only indirectly linked to the misconceptions underlying the entity view

[153] The tendency to respond to philosophical puzzlement by constructing and elaborate theory on the model of the natural sciences, is due the prevailing influence of what Russell called the ‘scientific method in philosophy’.

of the mind. First and foremost, they are the result of a general ignorance pertaining to the polymorphousness of the concept of thinking and ignorance of the wide variety of phenomena, which are associated with thought. As Gilbert Ryle put it:

> 'Only some thinking is excogitation; only some thinking is work; only some thinking has a topic or a problem. … Some thinking is just reverie or musing, intellectual doodling or strolling [154],

As the puzzlement and confusion surrounding thinking does not spring from the same source as the confusions surrounding the concept of the mind, the grammar of the concept of thinking requires independent scrutiny. Failing to do so would be wantonly negligent, as the confusions following in the wake of the main body of misconceptions underlying the cognitive view may, if not eradicated, attach themselves to or even give rise to other misconceptions of mind and mental phenomena. These may then distort the understanding of philosophers and cognitive scientists in the future. After all, the modern representationalist stance is merely an updated version of the classical representationalism endorsed by Descartes and Locke. Conceptual confusions, if not eradicated, clearly have the potential to distort our understanding again and again, and the latest hypothesis and theories devised to provide relief may prove to be mere variations on the same theme, as it were, while being fuelled by similar grammatical misconceptions. Hence, investigating the cognitive conception of thought is a necessary step on the way to completely dispelling the myth of the cognitive view.

Target Identification

Among the dubious presuppositions informing the cognitive conception of thinking are the ideas of thoughts as mental representations and of thinking as a mental process. The idea of thinking as a mental process is partly rooted in the misconception of introspection as a form of perception (see chapter III). Hence it is thought that introspection grants us access to our thinking and the underlying mechanism by means of self-observation. This idea plays right into the hands of the puzzlement we feel when confronted with the question "What is thinking?". One feels that one can't point to anything in reply, yet one *feels* that one ought to (BB1). In the wake of the demise of behaviourism and the development of cognitive science

[154] Ryle (1971b) p.258; See also: Ryle (1979) p.5

(cognitive psychology, in particular) the notion of unobservable cognitive processes underlying cognitive activities regained prominence[155]. Consequently, since the cognitive revolution of the 1950s, cognitive psychology has favoured the idea that speaking and acting are the public representation of a private domain of thought processes, i.e. the mental going ons in which cognitive activity really resides. Such private and sometimes hidden processes of thought were and still are assumed to accompany those actions we take to be "rational"[156]. Among cognitive scientists as well as among philosophers, it is widely believed that it is the accompanying thought, which endows these processes with their rationality. Consequently, the following investigation will also scrutinize the notion that it is an independent process of thinking, which is the soul of language, as it were, i.e. a process whose presence or absence renders speech thoughtful or thoughtless.

Despite its "*scientific"* appeal and deceptive explanatory power, the cognitive conception of thought itself leads to a variety of questions and conundrums, in response to which more confused theories and hypothesis have been conjured up by philosophers and cognitive scientists. The language of thought (LOT) hypothesis, for instance, constitutes the most significant and widespread of these misunderstandings. If left alone, conceptual misunderstandings may, as in the present case, grow to the size of a formidable avalanche of confusions and conundrums with the potential to sweep even the soundest and most careful of minds of its feet. For example, the notion of thinking as the processing of mental representations implies the question as to how the physical implementation of meaningful information processes is actually achieved. If there is a symbolic description in the brain of the outside world, as implied by the cognitive conception o thinking (a description cast in symbols which stand for the various aspects of the world of which our senses make us aware), what *language* is this description in? What is the *medium* of thought? In the light of his examination of numerous cognitive theories, Jerry Fodor (1975) concluded that, because almost all theories about cognitive psychological processes implicitly assume these to be computational, there must be a medium in which sensory information is computed in order for these theories to be true. Hence, he suggested the existence of an innate language of

[155] In the pre-behaviourist phase of psychology the notion of unobservable mental processes was endorsed by both William James and Wilhelm Wundt.
[156] See e.g.: Sterelny (1990)

thought (LOT)[157] as a representational system in which the computation of sensory information takes place. Fodor[158] conceived of the LOT as a semantically complete language, which differs from natural language, but which contains all the conceptual resources necessary for any of the propositions that humans could ever understand, think or express. The LOT, according to Fodor, constitutes the foundation not only of thinking[159], but of all cognition.

Although few have followed Fodor in adopting this extreme hypothesis, some weaker form of a language of thought (LOT) view, i.e., that there is a mental language that is different from human spoken languages, is held by many philosophers and cognitive scientists. The cognitive conception of thought bears witness to this fact. Besides the necessity to account for a medium of computation, i.e. the nature of the physical symbol system, the LOT hypothesis is also inspired by the apparent similarity of thought and language with respect to systematicity and normativity[160]. This refers in particular to the structural similarities, which exist between language and thought, and the similarity that exists between linguistic meaning and mental content. Both are understood as resulting from the compositional structure of sentences. The compositional structure of language is conceived to be paralleled by the compositional structure of inner mental representations in the LOT, which are regarded as the "linguistic vehicles of thought".

---

[157] This innate language is sometimes referred to as *mentalese.*

[158] Jerry Fodor is by no means the first to entertain the idea of a language of thought. In fact, the conception of thought which the young Wittgenstein propounded in the *Tractatus*, i.e., the idea that thinking is a kind of language, can be seen as an ancestor to the Fodorian idea of a LOT. In a psychological/cognitive science context we also find ideas strongly reminiscent of a LOT in William James's *The Principles of Psychology*:
`*...And in states of extreme brain-fag the horizon is narrowed almost to the passing word,- the associative machinery, however, providing for the next word turning up in orderly sequence, until at last the tired thinker is led to some kind of conclusion. (...) Nothing is easier than to symbolize all these facts in terms of brain-action.´(*James W, (1890), Ch. IX , p.247ff).
Similarly, Noam Chomsky holds: `*A naturalistic approach to linguistic and mental aspects of the world seeks to construct intelligible explanatory theories, taking as ´real´ what we are led to posit in this quest, and hoping for eventual unification with the `core´ natural sciences... . (...) In the present case, theories of language and mind that seem best established on naturalistic grounds attribute to mind/brain computational properties of a kind that are well understood ...'* See: Chomsky N (1995) `Language and Nature´ ,p.1, in: Mind,Vol.104.413

[159] The fully fledged language of thought hypothesis is that thinking consists, quite literally, in computational operations informed upon sentences of mentalese, an internal language with which human beings are innately endowed. For a creature to think on this view is for it to have rational-symbol manipulation processes occurring in the mind/brain.

[160] While some aspects of the systematicity and normativity of thought will be commented on in the present chapter, a full appraisal of these features of thought can only take place in the context of considering the intentionality of thought (see chapter VI).

The above outline provides a sketch of the manifold believes and presumptions, which inform not only our home baked ideas about thinking but also their more scientific counterpart. The various believes and presumptions constitute a tightly woven and closely interconnected web of conceptual confusions and grammatical misunderstandings. Entangled in a web of confusions, philosophers and cognitive scientists were forced to adopt a point of view (PI§308), which itself gives rise to a host of "bad philosophical questions" and which leads to the further perpetuation of confusion. In order to dispel the conceptual confusion, which prevents a proper understanding of thinking and the nature of thought, the following discussion will subject these various beliefs and presumptions to a philosophical investigation. The main assumptions to be investigated are,

a) the idea that thinking is to be understood as a *mental process*,

b) the associated idea, that we always think *in* something (i.e. the LOT hypothesis),

c) the idea that thinking goes on independently of speech, i.e. the picture we have of thinking as an inner accompaniment of speech, an inner process which can go on alone independently of the expression of our thoughts.

2. The Polymorphous Character of Thinking: Thought in its Natural Habitat

In Z§112, Wittgenstein reminds us of the widely ramified network of the individual uses of psychological verbs. Consequently, when striving for an overview of the concept of thinking we must not expect a uniform employment, but rather the contrary (Z§112; see also: RPPII§194). An initial scan of the conceptual geography of the concept of thinking reveals a stunning variety of uses, ranging from the ratiocinative to the expression of opinions and beliefs, and from instances of association and reflection to idle rumination and imagining[161]. However, philosophers and cognitive scientists generally fail to acknowledge the "varieties of thinking" in

[161] The following overview is partially indebted to Gilbert Ryles investigations into the concept of thinking (see: Ryle 1979) as well as Peter Hacker's discussion of thinking in Hacker (1997b) and Bennett and Hacker (2003).

favour of the misguided investigation of the *general case*. By studying the general case, however, it is falsely assumed that the universal features of thinking and thought are investigated. This approach mistakenly combines one's craving for generality with a flawed belief in the methods of science (BB18), i.e. by subscribing to the belief that an understanding of the phenomenon in question can only arise if one reduces the explanation of natural phenomena to the smallest possible number of laws (see Chapter II). By focusing on one variety of thinking in favour of others, cognitive scientists tend to give undue prominence to one kind of thinking, while failing to acknowledge all the other instances which constitute thinking. Although the acquisition of the concept of thinking during the course of our lives is (usually) rather unproblematic, problems arise when we try to describe the circumstance under which the concept is acquired (Z§114). In order to facilitate the description of the manifold variety of thinking, it is helpful to start the investigation off by reflecting on the most familiar variety of thinking as ratiocinative problem solving. This picture of thinking has been perpetuated in Rodin's famous sculpture *Le Penseur.*

Le Penseur, Ratiocinative Thought and related Forms of Thinking

At the heart of ratiocinative thinking, which could also be described as reasoned problem solving, lies the attempt to find the answer to a question or the solution to a problem. It has the characteristics of an activity, in so far as it can be engaged in continuously or intermittently, and in so far as it may be conducted methodologically, efficiently, promptly, swiftly or slowly, or formally or leisurely. It culminates characteristically, in the success or failure to find an answer or solution. Finally, ratiocinative thinking may be done silently, as during occasions when one is "sunk" or "lost in thought", but it may also involve speaking one's thoughts aloud (e.g.: "Now, what would happen if I didn´t…").

It is important to note, however, that silent thinking *per se* need not be ratiocinative. For example, I may try to remember who I was supposed to meet this afternoon, what Cécile asked me to pick up from Sainsbury's, when I was supposed to call my mum, or where I was to supposed to go in case the fire alarm goes off. Similarly, I may think silently yet with eager anticipation of the skiing holiday Cécile and I are going on. Alternatively, I may think about what my life would be like had I

never met Cécile. These instances of thinking, which involve neither reasoned problem solving nor remembering, may be thought of as imagining something, idle rumination or daydreaming. In such instances, thinking does not warrant the ascription of adjectives like methodological, effective, slow or swift as it is not directed towards a particular goal like answering a question or solving a problem. Depending on the individual, topic and context, these ruminations may be voluntary or involuntary (e.g. I may not be able to stop thinking about Cécile), or even compulsive.

Non - Ratiocinative Thought and related Forms of Thinking

Moving away from the reasoned problem solving variety of thinking, there is the kind of thinking, which varies with the degree of complexity of the task at hand (i.e., thinking associated with non-meditative activities). Mechanical tasks and activities, like sweeping the floor, sorting out rubbish, or cleaning the bathroom may be engaged in without paying much attention to the task at hand, or without concentrating much on what one is doing. Thus, when one is engaged in such activities, one frequently does so *without thinking*. However, this does not mean that thinking has no place here. Rather, it means that one characteristically thinks of other things, as when cleaning my bathroom I think of the movie I will watch afterwards. To say that I clean my bathroom without thinking, simply is to say that I do so without paying much attention to the task at hand, and that my mind is occupied with other things. When I am conducting a neuroscientific experiment, however, I am carefully thinking of all the eventualities, which may endanger the success of the experiment, and I focus all my attention on the task(s) at hand. Thus, one might say that I put a lot of thought into the experiment. On the other hand, if I fail to do so, one might say that I am thoughtless, and go about my duties as a DPhil student mechanically, without due care and attention. The more complex or delicate a task, the greater the level of thought and attention it demands. But in taking a thoughtful approach to the conduct of neuroscientific experiments, I do not talk to myself about the difficulties or problems, which may arise in the course of the experiment, for example. Rather, it just means that I am paying due attention and care.

Related to thought, which is associated with non-meditative activities, is the kind of thinking, which is involved when one attends to certain tasks or pursues

certain activities with cunning, skill, creativity or intelligence. This form of thinking is instantiated by the skilful diplomat contributing to the peaceful resolution of a crisis through the swift and intelligent reaction to changing circumstances, for example, or by the creative football player who can “read the game” and thus is able to play the decisive match-winning pass at the right moment. In other circumstances, this form of thinking may be exemplified by the sensitivity with which I recite a poem, or the originality of my rap rendition of “O sole mio”. A form of thinking that is related to thinking as attending to a task or activity at hand, is speaking with or without thought. To be speaking with thought, for example, may be the careful presentation of an argument, *viz.* an occasion where one has considered all the factors relevant to the issue one is speaking of. It is not, however, to accompany one’s presentation of the argument with an inaudible activity called thinking. By the same token, speaking without thought may be a case of failing to take into account all the factors pertaining to a certain issue, or a careless, unconsidered remark as when one compares the look effected by one’s girlfriends latest head gear to a racoon peering out from under a flower pot. In both cases, it is not the absence of a process of thought, which might go on in the brain in parallel to one’s statements, and which renders them thoughtless. Rather, it is the fact that theses statements have been voiced without due attention to their cogency or effect.

Beside those forms of thought that may be linked to activities of one form or another, there is the variety of thinking which takes the form of voicing an opinion, forming a judgement, formulating an assumption or making a supposition. Thus, depending on how one has arrived at what one thinks and the reasons one has for thinking it, one may think this and that according to one’s belief, opinion, judgement, supposition, assumption, assessment, evaluation or conclusion. “I didn´t think she was married”, voiced apologetically after one unsuccessfully tried to woe one’s supervisor’s wife at high table, indicates that one assumed or supposed, i.e. thought, that things are thus and so, for example. Of course, such a statement does not mean that one has given the matter due attention, and that after a prolonged period of reflection one has come to the conclusion that it is appropriate to pursue this rather than that lady. On the contrary, it indicates a distinct absence of thought.

The kind of thought that comes in the form of thinking of something, as when one daydreams, imagines something or ruminates about something for example, has already been discussed above. However, there is also a form of thinking of something *as* something. For example, I may think of Dega's painting *Dancers in Blue* as the perfect depiction of female grace and beauty without being chauvinist, or I may think of my brother Klaus as one of the best doctors around. Similarly, I may think of my forthcoming DPhil *viva* either as an excruciating grilling about my work of the last two years, or alternatively as a pleasant exchange of ideas on a topic of interest (depending on the time of day and level of Pimm's sedation). As Peter Hacker points out, thinking of something as something is a way of viewing it or a manner of conceiving it, which may either be illuminating and helpful, or misguided and confusing[162]. Finally, there is a variety of thinking which consists in meaning something or other by what one said. This may take the form, for example, of making explicit to whom or what one was referring to, as in "When I said 'his Jive is divine', I was referring to Bryan Watson", or "When I said 'that's nice', I was thinking of your extended holiday leave". Alternatively, it may also take the form of elaborating what one has thought, as when I said that "they suck", I didn't mean Oxford's but Cambridge's dancesport team.

An important form of thinking, which has not yet been touched upon, is the form of thinking as believing. In certain circumstances, thinking and believing, can be used as synonyms when expressing an opinion, judgement or point of view. For example, to say that 'pink becomes you' is to say ' I believe that pink becomes you', and the statement 'I believe it's going to rain tomorrow', is tantamount to saying that 'I believe it is going to rain tomorrow'. However, there also exist significant differences between the uses of think and believe. First, certain uses of thinking stand in close proximity to certain kinds of activities. I.e., while I can be engaged in thinking through a problem, I can not be said to believe through a problem. To try and think of a answer is not to belief of an answer, and while I can think swiftly or slowly, ably or efficiently, these attributes have no place in the case of believing. On the other hand, while I can believe in Santa Claus and the Easter Bunny, I cannot think *in* Santa Claus or the Easter Bunny. Second, in certain circumstances the use of thinking rather than believing may indicate a degree of apprehension. Bede Rundle, for

[162] Bennett and Hacker (2003), p.177

example, has pointed out that the statement 'I believe Kate is sincere', may be thought of as conveying more faith in Kate than does the statement 'I think that Kate is sincere'. While it may be granted that this suggestion weakens significantly if the emphasis is placed on the 'I' rather than 'think', there nevertheless seems to be a general difference. In contrast to thinking, believing seems to have a solemn ring to it. Thus, it appears to express a commitment to a position, opinion, or point of view that leaves no room for empirical proof, but signals an acceptance of something on the grounds of faith alone:

> '"I believe that God will provide" – not a happy context for 'think'. It is a word *(believe)* favoured by politicians as well as by the clergy, there being a suggestion that it is to one's credit that one believes, is prepared to take the plunge. Here we note that we can believe a person as well as what is said, a matter of placing faith in the speaker's veracity.[163],

3. Consequences for the Cognitive Conception of Thinking and the Empirical Study of Thought

The above overview of the varieties of thinking not only highlights the polymorphous character of thinking, but also outlines the many manifestations of the (mental) life of human beings, with which the concept of thinking is profoundly entwined. The phenomena of thinking prove themselves to be widely scattered indeed (Z§110). The grammatical investigation highlights, that as regards the concept of thinking we must, like in the case of the concept of mind, avoid falling victim to the temptation of supposing that *"where there is one word, there must be correspondingly, just one kind of thing."* Bearing this principle in mind, is an essential preliminary for gaining insight into what the word means, and thus, for gaining insight into the essence of thinking (see e.g.: PI§371). The failure to acknowledge this, however, has lead many cognitive scientists and philosophers to illegitimate generalisations from the particular to the general case, as a result of which, they falsely concluded in favour of a unitary neurological basis of thinking[164] (see chapter II). Highlighting the polymorphousness of thinking also illuminates the flaw in the cognitive science approach to the study of thinking, i.e. viewing one or two examples

[163] Rundle (1997), p.77
[164] See e.g. Greenfield (2000); *Brain Wise, BBC1 ,UK*

of thinking as representative of the entire variety of phenomena of thought. Of course, once under the influence of cognitive psychology (see below) and once one has started to look for just *one* thing that corresponds to the word thought or thinking, one quite naturally tends to attend only to those manifestations of thinking that *can* be described as a process or activity (of the brain). However, in the light of our grammatical observations, the cognitive conception of thinking seems overly broad as even our most elementary investigation into the grammar of the concept thinking fails to provide any substantial support for a unitary phenomenon of thinking.

The above overview of the various uses of the concept of thinking enables us to free ourselves from overhasty generalizations and conceptual confusion. What counts as thinking (like what counts as rehearsing, obeying, or practising, or fighting, farming or playing, which are also polymorphous verbs) depends not only on what is said or done, but on the context or circumstances, the manner in which it is said or done, the purpose with which it is said or done, and the forms of evaluation (if any) of success or failure appropriate to what is said or done. Thinking and the expression of thought are inextricably linked. In addition, it must be noted that the polymorphous character of thinking also highlights the gross violation of the bounds of sense, which occurs in the wake of ascribing thought to the brain. In the light of earlier mereological observations (chapter III), however, cognitive scientists and philosophers, must remember that it is the human being, which thinks, not the brain. Moreover, the brain is not the *locus* of thought either. In contrast to the grandiose claims of neuroscientists like Gerald Edelman, for example, the crucial link between the concept of thinking and the expression of thought[165] (see section 5 below, for details) implies that thoughts are *not* to be found in the brain. On can write down one's thoughts in a diary (or convey them in any other form of written text for that matter), express them in poetry or song or in the form of graffiti on the wall of some toilet stall, but one can neither find them in heads nor in brains[166]. A thought is just what is expressible by an utterance or other representation, and from the fact that human beings may chose not to disclose their thoughts, it does not follow that what they think is to be found in their brains (see section 6 below for details).

[165] Edelman & Tononi (2000), p.200
[166] Note: Of course, one can speak of the thoughts in one's heads *metaphorically.*

Finally, cognitive scientists and philosophers need to be careful with the interpretation of metaphorical phrases like “Use your brains”. Otherwise, they may be easily tricked into thinking that the brain is the organ of thought in the same sense in which the ear is the organ of hearing or the eye is the organ of sight. To say `Use your brains´ is tantamount to saying “Think!”, and is comparable to a metaphorical expression of affection like “I love her with all my heart”. Of course, it is true that without the very specific neural activities one could not think- but equally, without very specific neural activities one could not walk or talk either. Yet, no one would imply that we walk or talk with our brains. What goes on in the brain are neural processes, which are necessary conditions for the person, whose brain it is to be going through the relevant mental processes.

4. The Idea of Thinking as Mental Process

The cognitive view conceives of the interaction of mental representations as mental processes, and of thinking as *the* paradigm of a mental process. Cognitive scientists and philosophers subscribing to the cognitive view hold, that the process of thinking is constituted by the manipulation, i.e. the transformation, of symbolic mental representations in the LOT. Given its close connection with the LOT hypothesis, the picture of an inner process of thought warrants examination. A general sketch of the concept of a mental process has already been provided in the previous chapter (Ch.IV, Sec.4.4). The following discussion will concentrate on the specific notion of an inner process of thought.

In ordinary discourse, we are familiar with the use and application of the concept of a process. We competently apply it in the physical, chemical or biological domains as well as in the characterisation of certain human activities which involve, e.g., the sequential transformation of states or sequences of actions. Similarly, we are familiar with the use of the concept in the context of judicative, legislative, industrial or artistic contexts. However, the use of *this* picture of a process in the psychological domain leads by way of metaphor to metaphysics. The following discussion aims to show that this picture of an inner process does not so much constitute an illuminating analogy bur rather is the product of many misunderstandings.

First, contrary to the assumption of cognitive scientists a process of thought need not involve the manipulation and transformation of mental representations. Unlike the process of legalizing marihuana, which involves an ordered array of consecutive legal actions or the process of dialysis which involves the sequential separation of smaller molecules from larger molecules or of dissolved substances from colloidal particles in a solution by selective diffusion through a semipermeable membrane etc., *the process of thinking* need not involve transformations, changes or manipulations of any sort. Even if the notion of inner mental representations was intelligible[167]. What happens or takes place in someone's head when thinking need not necessarily involve a consecutive assortment of stages in an inner process of thought. In fact, it is characteristic of thought that it is not accompanied by process-like inner going ons. Although, there are some narrow confines in which the notion of a mental or psychological process can intelligibly be invoked, e.g. in the context of painful going ons or phases of depression and anxiety, confusion tends to ensue when philosophers and cognitive scientists invoke the notion of an inner process. Thinking, Wittgenstein points out, is not an incorporeal process, which lends life and sense to speaking, and which it would be possible to detach from speaking (PI§305). It is important to note that to question the link between thinking and the conception of an inner process is not tantamount to denying that people do think. Furthermore, it does also not imply that thinking is equivalent to speaking or doing as behaviourists might do (or any behavioural disposition for that matter) . To deny that thinking is an inner process is not to refute that a variety of words, images etc. may cross one's mind. But to describe these goings on is neither a report of *what* one was thinking, nor of the reasoning of the conclusion one has reached as a result of thinking, nor a depiction of an inner process.

The genesis of the idea of a mental process of thought lies partly in the peculiar imagery conjured up by certain turns of phrase employed in everyday discourse, and partly in the speculative metaphysics underlying the cognitive view. First, it should be noted, that it is perfectly sensible to suggest that we can speak with or without thought, or that we can speak mechanically or with empathy, compassion or understanding. However, the fact that we can speak thoughtfully or mechanically as a machine or a parrot might do, for example, does *not necessitate*

[167] The flaws of this notion have been exposed in Chapter IV.

the presence or absence of a process of thought, which sometimes accompanies speech and sometimes does not. As the overview of the concept of thinking illustrates, to speak with and without thought is tantamount to comparing someone's utterance to the playing of a piece of music with understanding and empathy, i.e. whether someone gives an emotional rendition of a piece of music or just a mechanical performance (PI§341). This analogy highlights the fact that whether I do something thoughtfully or mechanically, whether or not I say something thoughtful or thoughtless or whether I have thought about a problem or question *is not determined by reference to an internal process of thought,* but by the way I behave. Importantly, this is not meant to negate the existence of a myriad of mental or neural accompaniments of thought. But such going ons are merely causal preconditions for a person having a thought. In a logical, i.e. conceptual, sense they are neither necessary nor sufficient for a person having a thought.

These observations suggest that, contrary to the believes of many contemporary cognitive scientists and philosophers, it is not an inner process of thought, which *warrants* the ascription of thoughtful or thoughtless speech to a person. Rather, it is the *expression of thought*, i.e. the various *behavioural criteria*, which warrant the application of the attributes thoughtful or thoughtless to a person. One might say that to express one's thoughts, to say what one thinks, cannot be a description, report or communication of an inner process or activity, because no such description of an inner process could have the consequences in the language game-game of the expression of thought. Thought and its expression are crucially linked. If this sounds puzzling, it helps to remind oneself of the myriad of things a man must do in order for us to say he thinks (RPPI§563).

The content of thought is also critically linked to the expression of thought. *What* we think, is evident from what we say and what we do[168]. It is neither evidenced by the occurrence of an inner process of thought nor by entertaining mental images. Of course, mental images may cross my mind when I am thinking. But they are neither sufficient nor necessary for me to think. Again, this is in line with the understanding of the polymorphousness of thought, as evidenced by the overview of the concept of thinking. When I am toddling home in a drunken stupor after a long

[168] If we are sincere, of course!

night down the college bar, many a mental image may cross my mind, but I may no longer be able to think[169]. Moreover, I may think about a philosophical or mathematical problem without any images crossing my mind. Thus, although mental images and neural processes for that matter, can be accompaniments of thought, they need not be. They may or may not give rise to thoughts. They also may or may not serve as mnemonic devices. In any case, however, they do not determine what I think, nor whether I think. And having them is not a logically necessary condition in order for me to think that thus-and-so[170].

5. The LOT Hypothesis and the Idea of a Medium of Thought

The idea of a LOT is not a Fodorian invention. A brief look at the history of ideas reveals that a number of illustrious minds have endorsed the idea in the past in one form or another. Traces of the idea of a LOT can already be found in the writings of Augustine, for example, of whom Wittgenstein remarks that he

> '…beschreibe das Lernen der menschlichen Sprache so, als käme das Kind in ein fremdes land und verstehe die Sprach des Landes nicht; das heißt: so als habe es breits eine Sprache, nur nicht diese. Oder auch: als könne das Kind schon *denken*, nur nicht sprechen. Und „denken" hieße hier etwas, wie: zu sich selber reden.' (PI§32)
>
> '… describes the learning of human language as if the child came into a strange country and did not understand the language of the country; that is, as if it already had a language, only not this one. Or again: as if the child could already *think*, only not yet speak. And "think" here means something like "talk to itself".' (PI§32)

Similar ideas have also been voiced, for instance, by Thomas Hobbes who remarked that

> 'the general use of speech, is to transferre our Mentall Discourse, into Verbal; or the Trayne of Thoughts, into a Trayne of words; and that for two commodities; whereof one is, the Registering of the Consequences of our thoughts; which being apt to slip out of our memory, and put us to a new labour, may again be recalled, by such words as they (the thoughts) were marked by.´ (LN, Ch.IV),

and by John Locke who held that

[169] This, for example, may be evidenced by the fact that I propose marriage to the college porter on duty as well as several other people I pass in the quad whilst stumbling home.
[170] For a more detailed disussion of the content of thought see chapter VI.

> ‘(1.)... Man, therefore, had by nature his organs so fashioned, as to be fit to frame articulate sounds, which we call words. But this was not enough to produce language, for parrots and several other birds, will be taught to make articulate sounds distinct enough, which yet, by no means are capable of language. (2.) ...Besides articulate sounds, therefore, it was necessary that he should be able to use these sounds as signs of internal conceptions, and to make them stand as marks for the ideas within his own mind, whereby they might be known to others, and the thoughts of men’s minds be conveyed from one to another.’(E, Bk. III, Ch.I, Sec.1&2)

Locke’s lasting influence can be traced down to modern theoretical linguistics as exemplified by the writings of Ferdinand de Saussure[171], for example. According to the British empiricist tradition, thinking is operating on the ideas or images of the mind, and speaking is quite literally a translation from this imagist language of thought into word language. This *picture* of thought found its way into modern empirical psychology via the writings of William James who held, whilst engulfed in the `study of the mind from within´[172], i.e. introspectionist psychology, that the stream of thought is dissociated from speech. Thought according to James

> ‘always appears to deal with objects independent of itself.’[173]

Similar ideas have also been propounded by Frege, who thought that sense (in this case conceived as an abstract entity distinct from the sign) is logically independent of language but gives language life. Finally, the young Wittgenstein of the *Tractatus* believed that it is meaning or thinking which accompanies thoughtful non-mechanical speech and that this breathes life into otherwise dead and meaningless signs.

As this brief historical sketch shows, some form of the LOT hypothesis always seems to have exerted some influence over the most brilliant of minds. Its modern appeal, for example, is partly due to the fact that it appears to provide an explanation for a variety of linguistic phenomena, which human beings encounter in everyday

[171] see e.g.: Hacker (1993), `Meaning and Mind´, Part I, p.179 (Footnote)
[172] James, `The Principles of Psychology´, Ch.VII p.185
[173] ibid., Ch.IX p.219f

life[174], while also providing an account of the structure and systematicity(normativity) of thought (see Chapter VI):

> 'The theory of propositional representation (i.e., the LOT hypothesis) corresponds to the frequent introspective sense of a distinction between, having an idea, and putting it into words. The distinction emerges `behaviourally´ in the `tip-of the tongue phenomenon´ where we have a concept clearly in mind, but experience a delay in retrieving the appropriate word for it.
> Even in cases where we easily generate the English name for a concept (dog, chair, and so on) we often are not very good at giving its English definition. When asked to define even the most familiar concepts, we often find ourselves labouring.' [175]

It is expressions like, `*So what you really wanted to say was…´* (PI§334), which create the impression that while we look for the right words to express our thoughts, the actual `thinking´ is going on behind the words. Similar ideas are inspired by questions of the following kind: If a native English speaker converses in a foreign language, for example, does he *think his thoughts* in English or in another foreign language? As in occasions in which language users struggle to find the right words, such questions invite contemplation and hypostatisation with regard to the "something" a language seems to be thought in (PG66, PI§335). For what one meant was already seemed to be present somewhere in one's mind, even before we gave it expression. But as Wittgenstein pointed out, nobody would ask whether the written multiplication of two decimals runs parallel to the thought of the multiplication (PG§66). These examples highlight one of the distinguishing features of contemporary ways of thinking and contemplation, in so far as they are indicative of the mood of our times: One immediately responds to these kinds of problem by devising theories and hypotheses, rather than as Hertz[176] and later Wittgenstein urged to aim for (conceptual) clarification (see Chapter II).

---

[174] It should be noted, however, that originally the appeal of a representational theory of mind was not due to its plausibility as a explanatory theory of psychological phenomena, but rather, because it was compatible with the demands of a foundationalist epistemology (e.g. Descartes).

[175] Stillings et al (1995), p.26f

[176] Heinrich Hertz famously remarked *'that some kinds of vexing problems are to be resolved, not by scientific explanation and hypotheses, but by clarification. Unclarity is often expressed by questions about the nature of a phenomenon, but what is needed, in some such cases, is not fresh information or sharper definitions, but a clearer understanding of existing information and definitions´.* In the end, the questions troubling us often dissolve after such clarification. (See: Hertz (1956), p.iiif.)

In a similar spirit Ludwig Boltzmann argued for the importance of gaining an overview, an argument that might have had some bearing on the Wittgensteinian idea of 'Übersicht': *`...a division of labour certainly helps greatly to promote rapid progress in science and is indeed indispensable for it; but just as certainly it harbours great dangers. For we lose the overview of the whole, required for any*

The two main flaws of the idea of a medium of thought are the notion that there are symbolic representations of the (outside) world in the brain, and the failure to acknowledge the intimate link between the concept of thinking and the expression of thought (see above). First, although it is true that the capacity to manipulate symbols is bound up with the capacity of thought, this is not because thought must take place in a medium of any kind. Rather, it is because the *expression of thoughts in speech* must be bound up like this. If the expression of thoughts was not intimately connected to symbol manipulation, the concept of thinking could be given meaning without reference to the rules of a language community (i.e. thinking could, for example, acquire meaning by means of ostensive definition). This however, is not possible. It is the rules for the use of the concept of thinking and the criteria associated with these rules, which give the concept of thinking its meaning. Yet, in the absence of these rules and criteria, nothing *counts* as thinking.

Second, the fact that the capacity for thought is bound up with the capacity to manipulate symbols is not because thoughts must be expressed in a language. Rather, it is because the expression of thoughts in speech must be. Events like the 'tip of the tongue phenomenon', for example, deceive us into thinking that we are thinking *in* something. But, when I say that 'the word is on the tip of my tongue', all I am saying is that the right expression escapes me momentarily, and that I hope to find it soon. In the meantime the verbal expression does no more than certain wordless behaviour (PI II, p.219e). Similarly, to say "*well, I do know exactly what I want to say, but I cannot find the right words to say it*" when we struggle to give the right definition for a word or expression, is tantamount to saying "*Just give me a sec, I just need a moment for the thought to come together*". We tend to compare the occasional search for the right expression to the efforts of one who is trying to make an exact copy of a line which only he can see. But this analogy is flawed, and the situations are not comparable. In the case of searching for the right expressions,

---

mental activity aiming at discovering something essentially new or even just essentially new combinations of old ideas.' (See: Boltzmann (1979), p.77)

there is no line, no mental train of thought, which guides the verbal expression of what one wants to say (RPPI §580).

Third, while the LOT hypothesis would indeed explain the tip of the tongue phenomenon, it is neither the only explanation nor the most plausible one. When I am struggling to find the right word, I am not facing a translation problem. I am not asking myself which German, English or Italian phrase is the correct pendant to a mentalese expression which I entertain in my mind. Rather, in the first case I am trying to remember a sentence or phrase, which I believe captures the situation I am trying to describe adequately. In the second case, I am trying to come up with such an appropriate sentence or phrase. Thus, even these seemingly favourable cases do not *force* us to adopt the LOT hypothesis. In addition, the idea of a LOT seems also entirely implausible for the majority of cases in which we express our thoughts without hesitation.

Finally, must we always think *in* something, be it in the words of a particular language, images or mentalese, i.e. in some sort of symbolism or picture? While the fact that we can talk to ourselves *in foro interno*, as it were, may lend some appeal to this idea, talking to myself inwardly is, however, not the same as to think in a particular language. An interior monologue is neither necessary nor sufficient for thinking (see above). The question “what language do you think in?” arises only with respect to those individuals who, in addition to their native tongue, have also mastered a foreign language. To ask, “what language do you think in?”, is tantamount to asking whether you speak this foreign language fluently or hesitantly. Can you simply express what you want to say in the foreign language, or do you first have to be sure what you want to say, and then try to remember and express the equivalent in the foreign language. But it is important to realize, that this case provides no ground for extrapolating from the particular to the general. It does not necessitate the idea that one must first think in some sort of symbolism. This idea, which when taken to its extreme tends to precipitate in the form of pseudo scientific theories of the Fodorian variety, implies that one might always be mistaken about even the most simple and basic of one’s thoughts. This, however, is utterly implausible.

6. Mental Content, Thinking and Conceptual Abilities – A Revised Understanding of the Nature of Thinking and the Content of Thought

The previous chapters have excavated, exposed and resolved the numerous confusions and misunderstandings underlying the notion of the mind as a biological information processor. In addition, the view of the mind as the capacity to acquire intellectual skills, *viz.* mental abilities, has been introduced as a plausible alternative. In accord with this proposed shift in perspective, and in the wake of the deconstruction of the cognitive conception of thought, thinking needs to be accommodated within this proposed (neo-)Aristotelian framework. To do this, the present section will investigate the role of concepts in thinking and suggest an understanding of thinking as the capability to exercise an interlocking network of rule-governed abilities. This account of thinking is developed out of the critical link, which exists between thinking and the expression of thought.

The ability, or rather the set of abilities, which characterises thinking are manifest in those *practices* in which a human being masters a concept[177] (or symbolic representations of any kind). These practices range from linguistic to non-linguistic behaviours. Thus, thinking is to be understood as the mastery of a set of rule-governed abilities, which enable the subject of thought to pick out common features of the situations in which they are exercised. If thinking is understood in this way, we not only avoid the confusions and misunderstandings entailed by the cognitive conception of thinking, but we also make room for a plausible account of a *structured* and *generalizable* way of thinking about the world, which does not invoke the notion of a LOT (see also chapter VI). Both the structure of thought and its exercise are grounded in those practices in which a human being masters a concept. The rules operative in these practices, determine not only the application of concepts, but also the role that the concept plays in thought. In grasping a concept, one becomes competent in a subset of articulated abilities, which link situations and combine to shape our activity in the world. In order to illuminate the connection between the ability to think and the ability to use concepts according to rule governed ways, and above all, of these abilities to the structure of thought, it is necessary to begin with an examination of the link between concepts and

[177] Note: This can include verbal and non-verbal forms of communication and behaviour.

judgements. Although the following section will briefly touch upon questions relating to the content, structure and normativity of thought a detailed examination of these issues will be deferred to the subsequent chapter.

6.1 Mental Content: On Concepts , Judgements and Structure

The following account of concepts and conceptual structure is based on both Kantian as well as Wittgensteinian ideas. In the *Critique of Pure Reason (CrPR)*, Immanuel Kant argued that all human experience is conceptual. Thoughts, are *about* things, and formed by the combination of concepts. For example the thought 'this dog is an English Bulldog' involves the concepts 'dog' and 'English Bulldog'. Without these concepts, the subject of experience, i.e. the thinker, cannot frame the thought. Thus, in order to *think about* a given object one needs to possess concepts, which apply to a given object. Concepts, it could be said, group objects according to whether they count as instances of this or that. According to Immanuel Kant, for example, concepts are based on judgements, which represent items of experience of discursive thought. For a human being to apply a concept is, according to Kant, tantamount to making a judgement that something counts as an instance of this or that:

> 'Concepts are functions based on acts of unity of judgement' (CrPR B93) 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… The knowledge yielded by understanding, or at least by the human understanding, must therefore be by means of concepts and so is not intuitive but discursive (CrPR B93)[178]

In other words, for me to recognize the English bulldog chewing on my trainers as Giacomo, is for me to judge (on the grounds of coat colouring, size, behaviour etc.) that the English bulldog misbehaving is indeed the bulldog Giacomo and not, for example, the bulldog Achilles (although the bulldog Achilles is known to indulge in similar misbehaviours). Furthermore, Kant's idea that functions unify different experiences by "*bringing them under one common representation"* implies, that a structured body of knowledge can be generated from these functions. Within this

[178] *Intuition* is a Kantian *terminus technicus* denoting the content of perceptual experience, *viz*. sensory matter structured according to the forms of space and time.

structured body of knowledge conceptual connections between different situations can be discursively traced. Consequently, human experience obtains structure by virtue of its link to a conceptual structure, which itself constitutes a structured system of representation. As a result, once an object is related to a conceptual structure, the object can be used in discursive thought and feature in discourse. Thus, human thought is permeated with judgements as they figure both in the application of concepts, and in the determination or acknowledgement of the truth of a thought, *viz.* proposition. Judgements can thus be said to have a dual role in thought. The notion of judgement also played a role in Wittgenstein's thought. In the *Investigations* he states for example that

> 'Zur Verständigung der Sprache gehört nicht nur eine Übereinstimung in den Definitionen, sondern (so seltsam dies klingen mag) eine Übereinstimmung in den Urteilen. ...' (PU§242)

> 'If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments. …' (PI§242)

Kant´s and Wittgenstein's proximity regarding their views on judgements provides the basis for linking the Kantian notion of concepts with the Wittgensteinian ideas about following rules. This in turn allows for an account of the systematicity of thought, which is independent of a conception of a LOT[179], while also making room for an account of the content of thoughts, i.e. mental content, which does not invoke the notion of mental representations.

In both the Kantian and the Wittgensteinian context, the subject judges that this or that element of experience counts as an instance of a particular concept. It thus actively categorizes experience. Put in purely Wittgensteinian terms, the application of a concept follows distinctive norms, instantiated and regulated by the rules governing their *use*. Baker and Hacker, for example, point out that:

> 'What we call 'communication' must be partly determined by a certain consensus about what is true and what is false. Why? Because definitions (or explanations) of meanings are rules for the use of words and because the understanding of a rule is manifested in *two* ways, namely

[179] It is important to remind oneself here, that the modern cognitive view conceives of the workings of the mind, the mental processes, as the result of causal impingements of the outside world. These causal impingements take the form of inner mental representations. However, this empiricist notion (derived from Locke and Hume) neglects the Kantian insight that human experience is conceptual.

> in formulating or paraphrasing it *and* in applying or following it in practice (e.g. in making empirical judgements or giving descriptions (PI§240)).'[180]

Thus, judgements are a means by which human beings assert whether a concept applies to an object or situation. The application of a concept implies a normatively constrained judgement, and to grasp a concept is to understand that there is a way of going right or wrong. Consequently, a human being can be said to have a concept if it displays a rule-governed way of behaviour, i.e. response, and if it can participate in a given language game according to a rule governed technique. The defining characteristic of a thinker who has mastered a concept is, when his natural pattern of responses becomes principled or rule-governed. Usually, this development takes place in an intersubjective network of activity provided by a language community.

The ability of human beings to act in conformity with a rule has a dual aspect to it. On the one hand, a particular concept – once mastered – can figure as a structured element of thought, and on the other, by being incorporated in the already existing conceptual scheme, the concepts user's ability to think about his experience and to communicate with other becomes richer and more sophisticated as the conceptual structure expands. Each newly acquired concept is connected to the already existing conceptual structure and assumes it's own particular role. Thus,

> '…Das Licht geht nach und nach über das ganze auf' (ÜG §141)
>
> '…Light dawns gradually over the whole' (OC §141)

The present focus on concepts and judgements allows for an understanding of the structure of thought as resulting from the complex network of interconnected rule governed abilities, which allow human beings to pick out common features of situations in which they are exercised. The role a given concept plays in thought, is thus given by the way in which it characterises a number of *different abilities*:

> 'The exercise of a given concept in an act of judgement is not in general a definite uniform sort of mental act; it does not even make sense to ask just how many concepts are exercised in a given judgement. Our chess analogy may be of service, in showing why this question is unreasonable. Playing chess involves a number of abilities, which are not only distinguishable

[180] Baker & Hacker (1985), p.259

> but can actually exist separately; for one way of teaching chess would be to play first with the kings and the pawns and then add other pieces successively in later games. It would, however, be absurd to ask just how many of these abilities were exercised in a particular move. Our language about the concepts exercised in a given act of judgement makes sense or does not make sense in much the same way.'[181]

6.2 Mental Content and Generalizability of Concepts

In Kantian terms, the grasp of a concept is conceded by acknowledging that the concept user unifies a number of different objects under one common representation, as it were. To acknowledge the mastery of a certain concept requires the ability to subsume objects under concepts in judgement. This ability is constitutive of mastering a concept. It is, as Wittgenstein would put it, part of mastering the grammar of a concept. Consequently, the grasping of the meaning of a concept, i.e. to master the rules for the use of a concept such as *yellow*, for example, commits the user of a concept (i.e. the thinking subject) to a particular response in an indefinite number of interrelated judgements. For example, if my friend's three year old son Jaime can identify my Levi's jacket as being yellow when asked whether there is something yellow in the room, but fails to identify as being thus my Fender Stratocaster, my Lakers sweater, and a rubber duck, which are in the room too and which also have a yellow colour, Jaime cannot be ascribed the mastery of the concept *yellow*. On the other hand, had he been able to pick out my Strat, my Lakers sweater and the rubber duck as being yellow, as well as my Levi's jacket, his mastery of the concept could hardly be questioned. This example demonstrates that the structural role of concepts is crucially linked to their ability of functioning in generalizable ways when a subject thinks about the world. In the present context, the generality of concepts can be further underlined if we ask what Jaime is thinking when he identifies the Levi's jacket as being yellow, yet fails to identify the other objects as being of the same colour. His failure to identify all the yellow objects in the room implies that Jaime is unable to make (general) judgements about an object's "yellowness", as it were. Consequently, Jaime cannot be ascribed[182] the thought 'this jacket is yellow', as the thought's content implies the ability to make correct judgements about the yellowness of objects (or lack thereof).

---

[181] Geach (1957), p.15

[182] Nor can he self-ascribe.

For any subject to be ascribed the mastery of the concept of yellow, the subject must display appropriate responses to the colour of yellow and non-yellow items across a wide range of situations and contexts. This implies the *ability* to respond selectively to only those features or aspects of a situation or context, which contain instantiations of yellowness. The ability to discriminate between instances of yellowness and the lack thereof involves various judgements concerning the features of objects, and presupposes that the subject is capable of flexible control of attention as well as selectivity of response. Importantly, colour classifications are not mechanical as they do not just involve extrapolation from a paradigm, but rather may require the active search for the critical features in the objects a subject encounters. Consequently, it can be said that while the *vehicle* of a subject's *conceptual abilities* is the brain, i.e. they are built on natural faculties, the conceptual abilities are not entirely produced by it, as subjects develop and elaborate their conceptual abilities through their exposure to a community of language users. That is, thinking subjects are trained in the application of concepts through the adoption of rule governed patterns of response, which themselves are founded on criteria circumscribing the correct use of the respective concepts. To grasp a concept is an ability to respond at will to a rule shaped through multiple applications. As such responses are feature selective according to the given inter- or intrasituational contexts, they have a potentially general use. Once a subject has developed a conceptual ability, and displayed proficiency in the use of concepts the subject can be ascribed (and self-ascribe) the grasp of the conception in question. The mastery of a concept implies proficiency to express his judgements concerning a concept. It is in this way that a thought may be conceived of as what is expressible by an utterance or other representation. It is manifest in the variety of linguistic and non-linguistic behaviours associated with the master of a given concept. As the connective analysis of the concept of thinking has highlighted, these manifestations have a distinctly polymorphous character.

In the wake of the present investigation, three main features emerge as characteristic of the content of thought, (i.e. mental content). First, the normative constraints on the content of thought imply an intersubjective nature of content. The normative constraints connect one's self-ascription of a concept with the general

practice in which the general grasp of a given concept is ascribed. The manifestation of the correct use of a concept constitutes the basis for this ascription. This aspect is also underlined by Strawson's generality principle:

> '…the idea of a predicate is correlative with that of a range of distinguishable individuals of which the predicate can be significantly, though not necessarily truly, affirmed.'[183]

For example, to predicate of myself the grasp of the concept 'English bulldog', involves making a judgement, which I ought also be able to make about other subjects. Consequently, the ability to judge that I grasp the concept 'English bulldog', is tied to my ability to judge that *Klaus* grasps the concept 'English bulldog' etc.. The link between self-ascription and the ascription of a concept to others, thus follows from the fact that mental ascriptions are a particular class of generalizable predicative concepts. Strawsons generality principle, implies that if one grasps a grasp of the concept 'English bulldog', for example, as predicative, then one ought in principle be able to judge whether this is instantiated by a range of subjects. Second, manifestability, is an essential feature of the grasp of a concept. Klaus can only be ascribed a grasp of the concept 'English bulldog' because his grasp is manifest in his application of the concept. In addition, the link between the self-ascription of a concept and the ascription of a concept to someone else implies, that although the grasp of a concept need not be manifest on every occasion it is applied by a subject, manifestation is still the ground on which mastery of a concept is ascribed[184]. Finally, the present analysis provides an account of the structure of thought, independent of the notion of a medium of thought. This is in line with a recommendation made by Gareth Evans who suggests that the structure of thought should not accounted for in terms of a medium of thought, but in terms of thoughts *'…being a complex of the exercise of several distinct conceptual abilities.'*[185] Structured thought about experience is manifest in and constituted by the exercise of those abilities, which allow for the selection of aspects of presentations and which allow for their generalization at will, so as to ensure the subsequent use of the outcome of these generalizations in an action guided manner. The structure of thought is reflected in considerations about what a subject must do if he is to be ascribed a mastery of a

[183] Strawson (1985), p.99. See also: Strawson (1999)
[184] Note: Although once mastered, the exercise of the relevant ability might be concealed.
[185] Evans (1980), p.101

given concept. Thus, the structure of thought is evident in the close connection between thought and its expression. To judge whether a concept counts as an instantiation of this or that, is to judge that it is both like and unlike various other things.

7. Summary and Concluding Remarks

Absence of conceptual clarity is one of the main sources of speculative metaphysics. The preceding discussion underlines that, in moments of puzzlement and perplexity, philosophical speculation must be avoided in favour of conceptual investigations. Failing to do so will invariably lead to the construction of “houses of cards” (PI§118). The present investigation into the cognitive conception of thinking has shown this conception to be overly broad and simplistic, as it cannot be supported by the most rudimentary grammatical analysis. It neglects the polymorphous character of thinking, and consequently fails to provide a genuine insight into the nature of thinking and thought. Most importantly, the cognitive conception of thinking fails to acknowledge that thinking and the expression of thought are inextricably linked through the normative constraints governing the use of concepts and the various judgements involved in ascribing a concepts. While it seems ill advised to attempt giving a general definition of thinking in the light polymorphous character of this concept, one can nevertheless specify what has to be true of an individual for him or her to count as a thinking being. That is, an individual needs to be able to participate in an interlocking network of rule-governed abilities, which pick out common features of the situations in which they are exercised. This stands in stark contrast to the metaphysical conception of thinking as the processing of internal mental representations as promoted by the cognitive conception of thinking and the cognitive view. Thinking is not an inner process, as even within the narrow confines in which an association with a psychological process makes sense (e.g. during periods of grief, depression or anxiety), a description of the words or images that may cross one’s mind at that time does not count as a report of what one was thinking, nor is it a depiction of an inner process. And although mental images can be accompaniments of thought, it is important to note that they need not be. Similarly, although the specific neural activity of the brain is a necessary condition for a person to undergo mental processes and to entertain thoughts, it does not follow that these processes constitute thinking, anymore than neural activity in the visual cortex constitutes seeing.

In the wake of the flawed cognitive conception of thinking, the idea of a medium of thought in general, and the idea of a LOT in particular, are inevitable

extensions of the confusions informing the cognitive view. These ideas constitute a *scientistic myth* founded on numerous misunderstandings and misconceptions. It exemplifies how cognitive scientists and philosophers are taken in by conceptual confusion, and how this results in the construction of metaphysical theories, which give the impression of providing a picture of the deep underlying reality of things. In defusing and deconstructing these misleading pictures, however, one destroys the very foundation on which the cognitive conception of thinking and the LOT hypothesis are built. In contrasts to Fodor's claims, the cognitive conception of thinking and the LOT hypothesis do not constitute the foundation of thinking or the explanation of cognitive phenomena in general for that matter. Thinking does not consist in computational operations upon sentences of *mentalese* but in the various forms of linguistic and non-linguistic behaviours, which characterise the expression of thought. For a creature to think, is not to have rational-symbol manipulation processes occurring in the mind/brain, but to exercise conceptual judgement according to rules governing the use of concepts. In the context of the Aristotelian conception of mind introduced in the preceding chapter, thinking is best understood as an interlocking network of rule-governed abilities, which pick out common features of the situations in which they are exercised (see above). To have a concept of x we must not only have a formal understanding of what x is…but also knowledge of what x is to be applied to." The rules operative in these practices determine not only the application of concepts, but also the role that the concept plays in thought ("what it is for something to be an x"). The abilities, which constitute thinking, are structured and depend on a flexible selectivity of attention and response to the world. The exercise of these abilities comprises thought with content rather than merely a complex of causal influences, which explain the subject's behaviour. Both the self-ascription, and the ascription of mental content to others have their origins in these abilities. All human beings are endowed with the capacity to develop these (congruent) abilities, which provide the foundation for the learning and exercise of rule-governed techniques, and which imply an agreement in judgements. This relation between thought, the expression of thought and rule governed abilities, allows for structured and generalizable ways of thinking, and provides the basis for a non-causal account of thought as well as linguistic content (see: chapter VI).

Chapter VI

Intentionality and the Cognitive View

1. Representationalism and Intentionality

The modern representationalist stance embodied by the cognitive view has developed as a result of the inability of functionalist theories of mind to account for mental content. While functionalist theories of mind allowed mental states like being in pain, for example, to be multiply realized by defining types of mental states according to their functional role (thus improving substantially upon identity theories of mind), they failed to provide an account for intentional mental states as well as for sensations (see Chapter I).

> 'The functionalist thinks of mental states as causal intermediaries between perceptual inputs and behavioural outputs. This is an advance on thinking of them simply as physical states. But, for all that, functionalism still presents mental states as part of a system of causal pushes and pulls inside the head.'[186]

Modern representationalist theories of mind recognized and addressed this problem by postulating, that in addition to their functional roles, mental states also *encode meanings*. This assumption made a significant contribution to the picture of the mind as a physical symbol system, and the LOT hypothesis. Because the content of mental representations can be described as attitudes or relations towards propositions it is also called propositional content, and the content laden mental states are often referred to as propositional attitudes (PAs). Within the context of the cognitive view, the possession of PAs by human beings is accounted for by the claim that the internal mental states (i.e. tokenings of inner mental representations) encode the respective propositional attitudes. Fodor´s language of thought hypothesis, discussed in the previous chapter is widely conceived of as the most thorough account of a physical symbol system within which inner mental representations are realized:

> 'At the heart of the theory is the postulation of a language of thought: an infinite set of `mental representations´ which function both as the immediate objects of propositional attitudes, and

[186] Papineau (1987), p.69

> as the domains of mental processes.' More precisely, the representational theory of mind is the conjunction of the following 2 claims: 1. (the nature of prepositional attitudes): For any organism O, and any attitude A toward the proposition P, there is a (computational/functional) relation R and a mental representation MP such that MP means that P, and O has A iff R to MP… . 2. (the nature of mental processes): Mental processes are causal consequences of tokenings of mental representations.'[187]

Moreover, it has also been pointed out that

> '…the feature of `propositional attitudes´ known as `*intentionality´* itself forces upon us a theory of mental representations. The intentionality of a thought consists in the fact that verbs of propositional attitude are about or directed, in a curious way, upon the situation specified in the proposition they contain.' [188]

The conception of mental processes as the causal consequences of tokenings of inner mental representations, i.e. mental states, as envisaged by Fodor (see above) is important because it provides the crucial link between mental representations and a causal reductionist explanation of human action and behaviour. Mental representations are conceived as the inner termini of the causal chains that instigate action. Consequently, they assume a causal functional role in the explanation of action, which is analogous to the role microscopic unobservables play in the explanation of the behaviour of macroscopic objects. The property of intentionality is the decisive feature in virtue of which mental representations are able to fulfil their causal role. Without the property of intentionality, it is thought, there can be no causal explanation of behaviour (see: Chapter I, Sec. 3.1, 3.2).

Intentionality is a property of both content laden mental states, like beliefs, desires, hopes etc., and linguistic tokens like (written) sentences or utterances. They can be about distant or non-existent affairs or they can be descriptions of, or utterances about, actual goings on. Although there is a close connection between the mind and intentionality[189] it is important to note that not all content-laden mental states are intentional states. Mental states, like sensations of pain or pleasure, lack intentionality. In the present context, intentionality is to be understood as a property

[187] Fodor (1987), p.16f
[188] Preston (1997), p.8
[189] See e.g.: Brentano (1874)

of certain content-laden mental states[190], as well as of linguistic entities, like utterances, expressions etc.. Moreover, because intentionality is characteristic of both certain content-laden mental states and linguistic tokens it appears to provide a link between mental content and linguistic meaning. Because content-laden mental states can be expressed as propositional attitudes, and the expression of propositional attitudes by human beings is a way of communicating mental states through language, the expression of PAs provides crucial evidence for the ascription of mental states. It is in this way that mental content and linguistic meanings are crucially linked up.

To develop a satisfactory account of the intentionality of mental states and linguistic meaning has been a long standing problem within the philosophy of mind and contemporary cognitive science. To many, intentionality appears to be an utterly mysterious property. How is it that mental states, thought and language can be about something? The fact that intentionality seems to escape our usually highly successful explanations in terms of physical properties makes this feature even more puzzling. The sciences can easily explain a person's height, mass, weight, colour of hair, eyes or skin etc., but they can't explain my hope that I may dance with Cécile for OUDT this year, or my desire to strangle my new bulldog puppy Giacomo, because he has chewed up the body of my favourite Fender Stratocaster and left significant 'bite-marks'. And while the quality and character of my voice (i.e. its pitch, frequency etc.) can be analysed in physical terms, the fact that my wish to dance with Cécile concerns a non-existent state of affairs, and that my desire to strangle Giacomo is about a hypothetical action that would never actually be carried out, can not be explained in terms of natural science. A characteristic divide seems to exist between the properties of human beings and their minds. Consequently, the question how the property of intentionality fits into our naturalistic framework of nature, and how it can be integrated into our general scientific view of the world, forces itself upon us.

In order to address these questions in a naturalistic manner, modern representationalism adopts a strategy of *explanatory priority*. It attempts to explain

[190] Note: Because content can be stated using a that-clause, it is also called propositional content. Content-laden mental states are propositional attitudes because they can be individuated as attitudes or relations towards propositions.

linguistic meaning as resulting from mental content and then to give a reductionist account of the latter. Naturally, this account gravitates around the notion of inner mental representations, which are thought to stand in causal relations to the world. Thus, mental content is naturalized by way of providing a causal explanation of content:

> 'What we want at a minimum is something of the form 'R represents S' is true iff C where the vocabulary in which condition C is couched contains neither intentional nor semantic expressions.'[191]

This approach is exemplified in the cognitive view in general and Fodor´s language of thought hypothesis in particular, where the explanation of mental content takes the form of a purely descriptive causal theory. The adoption of this explanatory strategy implies the assumption that inner mental representations possess natural, underived or intrinsic content (as opposed to conventional or derived content/intentionality) in virtue of the causal relations in which they stand. The meaning of language is then explained as deriving from the content of mental states[192], which itself is explained by reduction to non-intentional phenomena.

In addition, this reductionist explanatory strategy is also thought to provide an account for two further characteristic properties of mental content, besides intentionality: Its structure and normativity. The normativity of content is linked to the conceptualisation of intentionality by means of PAs and propositional content, as PAs specify a condition for the world to satisfy, if the attitude is to be fulfilled. In addition, mental states stand in logical and rational relations to each other[193]. Furthermore, it has been claimed that understanding a thought, i.e. entertaining a PA, is a structured ability in as much as it presupposes connections and links to other thoughts (or PAs). That is, if one is able to understand the thought like <this bulldog is naughty>, one must also be able to understand other thoughts which contain the concepts of bulldog and naughty, which convey, for example, that there are bulldogs that aren't naughty or that there are other dogs besides bulldogs, which are naughty too. Both the normativity and the systematicity of content, and thus the normativity of thought

[191] Fodor (1991) p.32
[192] I.e., from tokenings of inner mental representations in a language of thought.
[193] See e.g. Frege (1964)

are explained by postulating the existence of causal relations between mental representations.

The following discussion will examine the plausibility of the representationalist account of intentionality and the normativity and structure of content, which has been outlined above. Particular emphasis will be placed on examining the notion of *explanatory priority* and the hypothesized causal role of mental states. In the course of this investigation, some of the criticism and ideas regarding the conceptions of a medium of thought, mental states and the structure of thought discussed and introduced in preceding chapters will be revisited to the extent that they have a direct bearing on the issues examined in the present chapter. The alternative account of intentionality and mental content will draw heavily on Wittgenstein's conception of language and his ideas about grammar and rules.

2. Examining the Representationalist Account of Intentionality, and the Normativity and Structure of Mental Content

Content-laden mental states are normative in that they prescribe what should satisfy them in advance. As Wittgenstein points out,

> 'Der Wunsch scheint schon zu wissen, was ihn erfüllen wird, oder würde; der Satz, der Gedanke, was ihn war macht, auch wenn es gar nicht da ist! Woher dieses Bestimmen, dessen, was noch nicht da ist? Dieses despotische Fordern? („Die Härte des logischen Muß?.")' (PU§437)

> 'A wish seems already to know what will or would satisfy it; a proposition, a thought, what makes it true--even when that thing is not there at all! Whence this *determining* of what is not yet there? This despotic demand? ("The hardness of the logical must.").' (PI§437)

But, can the *normative relations* that these states have to the facts or events that satisfy them still be accounted for, if mental states are identified with internal states of the mind or body, which posses non-relational properties? In order to settle this question two key premises of the cognitive view need to be investigated. First, the claim that mental content can be explained by the possession of internal mental representations. Second, the idea that there is a distinction between intrinsic and

derived intentionality. The distinction between intrinsic and derived intentionality, provides a suitable starting point for the following discussion.

2.1 On the Plausibility of Distinction between Intrinsic and Derived Intentionality

The motivation for adopting the distinction between intrinsic and derived intentionality arises out of the assumption that while thought requires no interpretation, signs, expressions and behaviour do. While thought is deemed to possess *intrinsic intentionality,* signs, expressions and behaviour are considered to possess merely derived intentionality, which necessitates interpretation. This distinction corresponds to a picture of language, which divides language into an inorganic part – the handling of signs – and an organic part – the interpretation of signs. It is believed, that it is the organic component of language – the interpretation of signs - in virtue of which live is breathed into otherwise dead signs of meaning (PI§431). Let's examine each half of this distinction in turn.

The notion of derived intentionality is highly problematic. First, in the tradition of cognitive psychology, for example, one might want to suggest that interpretation is achieved through the workings of some mental thought like process. Yet, the implausibility of such an explanatory strategy has been highlighted in the last chapter within the course of the discussion of thoughtful and thoughtless speech (see chapter V). Second, problems for this explanatory strategy also arise as the consequence of the assumption that signs, expression or behaviour are always in need of interpretation is, that the sharing and conveying of meaning is *a.* either a lucky guess, or *b.* entirely impossible. For example, on this account it is only possible for someone to share or convey the meaning of a sign or an expression by invoking more signs. Yet, these could also be given multiple interpretations. Consequently, explanation after explanation of the meaning of the previous explanation is required, leading to an *infinite regress.* It thus appears impossible for the meaning of someone's words, to be fully conveyed to others. There is always room for guesswork, always room for error and mistake. Wittgenstein, describes the dilemma thus:

> 'Wenn wir einen Befehl geben, so kann es scheinen, als ob das Letzte, was der Befehl wünscht, unausgedrückt bleiben muß, da immer noch eine Kluft zwischen dem Befehl und

seiner Befolgung beleibt. Ich wünsche etwa, dass einer eine bestimmte Bewegung macht, scheint unzweideutig; bis auf die Frage: wie weiß er, dass *er diese Bewegung machen soll?* – Wie weiß er überhaupt, wie er die Zeichen, welche immer ich ihm gebe, gebrauchen soll? – Ich werde nun etwa trachten, den Befehl durch weitere Zeichen zu ergänzen, indem ich von mir auf den anderen deute, Gebärden der Aufmunterung mache, etc. hier schient es, als finge der Befehl zu stammeln an.

Als trachtete das Zeichen mit unsicheren Mitteln in uns ein Verständnis hervorzurufen. – Aber wenn wir es nun verstehen, in welchem Zeichen tun wir das?' (PI§433)

'When we give an order, it can look as if the ultimate thing sought by the order had to remain unexpressed, as there is always a gulf between an order and its execution. Say I want someone to make a particular movement, say to raise his arm. To make it quite clear, I do the movement. This picture seems unambiguous till we ask: how does he know that he is to make that movement?--How does he know at all what use he is to make of the signs I give him, whatever they are?--Perhaps I shall now try to supplement the order by means of further signs, by pointing from myself to him, making encouraging gestures, etc.. Here it looks as if the order were beginning to stammer.

As if the signs were precariously trying to produce understanding in us.--But if we now understand them, by what token do we understand?' (PI§433)

In other words, if signs always have to be interpreted the full meaning of one's words is never available to others, but rather has to be guessed. Explanations of meaning would never convey the full meaning of a sign or an expression. This, however, is absurd.

The problems associated with the intrinsic/derived distinction do not stop here, however. The problems regarding a representationalist explanation of intrinsic intentionality non-withstanding (see below), the flip-side of the argument that one always falls short of conveying the full meaning of a sign or expression is, that oneself knows something more than one can actually explain. The conception of intrinsic intentionality seems to imply a form of *private understanding* of what it is that is supposed to be explained. We seem to have a deeper understanding of the *explanandum* than we can convey in the *explanans* (PI§209). Yet, this intuition cannot be substantiated. Our understanding of the meaning of a word is connected to our understanding of how to apply it (see chapter V). But this understanding is not derived by way of entertaining a form of private[194] understanding in my head that

[194] For a discussion of the notion of privacy at issue here, see chapter III.

comes to me, e.g., in a flash. Consequently, any belief that someone can have a deeper understanding of one's own meaning than other's could possibly have is absurd:

> '"Aber erklärst du ihm wirklich, was du selber verstehst? Lässt du ihn das Wesentliche nicht *erraten*? Du gibst ihm Beispiele, - er aber muß ihre Tendenz erraten, also deine Absicht." – Jede Erklärung. Die ich mir selbst geben kann, gebe ich auch ihm.. – „ Er errät, was ich meine" würd heißen: ihm schweben verschidene Deutungen meiner Erklärung vor, und er rät auf eine von ihnen, Er könnte also in diesem Falle fragen; und ich könnte, un dwürde, ihm antworten.' (PI§210)

> '"But do you really explain to the other person what you yourself understand? Don't you get him to *guess* the essential thing? You give him examples,--but he has to guess their drift, to guess your intention."--Every explanation which I can give myself I give to him too.--"He guesses what I intend" would mean: various interpretations of my explanation come to his mind, and he lights on one of them. So in this case he could ask; and I could and should answer him.' (PI§210)

There is no private token which one could have in mind, like a flash of understanding, for example, which governs the correct application of a sign or expression in this way. It is the rules for the use of sign or expression, which govern its application and which are learnt by exposure to a community of language users. Consequently, there is no reason to assume that one can have a deeper understanding of he meaning of the signs and expressions one uses than others. The meaning of a sign or expression is accessible to anyone, and intelligible to the same extent. The upshot of the present discussion is that the distinction between intrinsic and derived intentionality is implausible. On the one hand, it seems impossible for the signs and expressions, which are thought to possess derived intentionality, to acquire any content. On the other, the content of thoughts, which are supposed to possess intrinsic intentionality, remains a mystery, as it never can be fully conveyed.

2.2 On the Philosophical Foundation of Intrinsic Intentionality: Investigating the Plausibility of the Representationalist Account

The strategy contemporary representationalists employ in the attempt to explain the intrinsic intentionality of mental states is to postulate, that the intrinsic

intentionality of the inner mental representations encoded in the LOT underlies the intrinsic intentionality of mental states. That is, the content of internal mental representations is thought to explain the content of mental states. Yet, this raises the following question: What gives the inner mental representations their content? In the light of the previous discussion of an interpretive account of (mental) signs, it seems that an interpretative account of the intrinsic intentionality of symbolic inner mental representations would equally lead to an infinite regress. If mental content is encoded in a symbolic LOT which requires interpretation, then the correct method of interpretation would also have to be encoded. However, if we assume that any form of encoded content[195] also stands in need of interpretation, this kind of explanation could only succeed via invoking another inner mental representation, which would also require interpretation and an infinite regress ensues.

It seems that a way out of this dilemma would be to take the view that mental representations are contents themselves, and as such do not require interpretation. For example, one might claim that mental representations do not require interpretation, as the *representations are the contents* of signs and symbols, rather than signs and symbols themselves. Thus, one would avoid an infinite regress simply by invoking the possession of inner mental representations. Yet, the flaw with *this* explanatory strategy is that on close scrutiny, it fails to explain anything. Representationalism is an explanatory strategy, i.e. the cognitive view is a framework of explanation, which is thought to explain the possession of psychological attributes by human beings. However, merely taking the possession of inner mental representations as an *explanans* of mental content, fails to shed any light on the problem how mental content is possible. A mere wish, does neither provide a philosophically nor a scientifically satisfactory account of intentionality. One simply substitutes one set of words with another. Thus, as far as explanatory power and value is concerned, the modern representationalist account explanation leaves as much to be desired as its *classical counterpart*. Although, at this point one may be tempted to object that these considerations are mere straw men, serving no greater purpose, it is noteworthy that Wittgenstein in the *Blue Book* found it worthwhile to spend some time on examining this line of thought:

[195] In the present context the rules of interpretation.

‘…was wir erwarten, ist nicht die Tatsache, sodnern ein Schatten der Tatsache; gleichsam das, was der Tatsache am nächsten kommt. Wir haben gesagt, daß man damit die Frage nur einen Schritt weiter zurückführt. Diese Idee eines Schattens hat mehrere Ursprünge. Einer davon ist dieser: wir sagen: „Gewiß können zwei Sätze aus verschiedenen Sprachen denselben Sinn haben“; und wir argumentieren: deßhalb ist der Sinn nicht daßelbe wie der Satz“, und fragen: „Was ist der Sinn?“ Und wir machen aus „ihm“ ein Schattenwesen, eines der vielen, die wir erschaffen, wenn wir den Substantiven, denen kein körperlicher Gegenstand entspricht, Bedeutung geben wollen.

Ein anderer Ursprung der Idee, daß ein Schatten der Gegenstand unseres Denkens ist, ist dieser: Wir stellen uns vor, daß der Schatten ein Bild ist, dessen Intention nicht *in Frage gestellt werden kann*, das heißt ein Bild, das wir nicht deuten, um es zu verstehen, sondern das wir verstehen, ohne es zu deuten. Nun gibt es Bilder, von denen wir sagen würden, daß wir sie unmittelbar verstehen, ohne jede weitere Deutung. ...Der Schatten wie wir ihn uns vorstellen, ist ein Art Bild; er ist in der Tat einem Vorstellungsbild, daß sich vor unser geistiges Auge schiebt, sehr ähnlich; und dies wiederum ist einer gemalten Wiedergabe im gewöhnlichen Sinne nicht unähnlich. Einer der Ursprünge der Idee vom Schatten ist gewiß die Tatsache, daß in einigen Fällen das Sagen, Hören oder Lesen eines Satzes Vorstellungen vor unser geistiges Auge bringt; Vorstellungen, die dem Satz mehr oder weniger deutlich entsprechen, und die deßhalb in einem gewissen Sinne Übersetzungen dieses Satzes in eine Bildersprache sind. – Aber es ist absolut wesentlich, daß das Bild, als das wir uns den Schatten denken, das ist, was ich ein „Bild aufgrund von Ähnlichkeit“ nennen werde. Ich meine damit nicht, daß das Bild dem ähnlich ist, was es einer Absicht entsprechend Darstellen soll, sondern dass es ein korrektes Bild nur dann ist, wenn es dem ähnlich ist, was es darstellt. Für diese Art Bild könnte man das Wort „Kopie“ gebrauchen. Beiläufig gesprochen sind Kopien gute Bilder, wenn sie leicht mit dem verwechselt werden können, was sie darstellen.’ (BB_D, p.63f.)

‘…what we expect is not the fact, but a shadow of the fact; as it were, the next thing to the fact. We have said that this is only pushing the question one step further back. There are several origins to this idea of a shadow. One of them is this: we say "Surely two sentences of different languages can have the same sense"; and we argue, "therefore the sense is not the same as the sentence", and ask the question "What is the sense?" And we make of 'it' a shadowy being, one of the many which we create when we wish to give meaning to substantives to which no material objects correspond.

Another source of the idea of a shadow being the object of our thought is this: We imagine the shadow to be a picture the intention of which *cannot be questioned*, that is, a picture which we don't interpret in order to understand it, but which we understand without interpreting it. Now there are pictures of which we should say that we interpret them, that is, translate them into a different kind of picture, in order to understand them; and pictures of which we should say that we understand them immediately, without any further interpretation. …The shadow, as we think of it, is some sort of a picture; in fact, something very much like an image which comes before our mind's eye; and this again is something not unlike a painted representation in the ordinary sense. A source of the idea of the shadow certainly is the fact that in some cases saying, hearing, or reading a sentence brings images before our mind's eye, images which more or less strictly correspond to the sentence, and which are therefore, in a sense, translations of this sentence into a pictorial language.--But it is absolutely

essential for the picture which we imagine the shadow to be that it is what I shall call a "picture by similarity". I don't mean by this that it is a picture similar to what it is intended to represent, but that it is a picture which is correct only when it is similar to what it represents. One might use for this kind of picture the word "copy". Roughly speaking, copies are good pictures when they can easily be mistaken for what they represent.' (BB, p.36f)

These remarks shed some light on the inherent appeal of the above lines of argument. By assuming that the last interpretation is a picture, which represents by virtue of its similarity to the *real thing*, this line of thought side steps the question as to the possibility of representations. Thus, the dilemma here is that while an explanatory model based on invoking the idea of an inner picture leads to an infinite regress, an explanation that does not involve a surrogate for mental content[196] fails to be explanatory.

Finally, the central representationalist claim that inner mental representations are the inner termini of the causal chains of action is incompatible with the identification of mental representations with content. This is because representationalism assumes that the causally relevant properties of inner states are due to their first-order physical properties. In order for mental representations to possess these properties, however, they are to be conceived of as concrete states of or within a body. Contents, on the other hand, do not have any physical properties. Only the bearers of content-laden mental states – i.e., human beings – do. Thus, if the above line of argument is pursued, mental representations could not fulfil their hypothesized causal role in the explanation of thought processes.

On the Notion of Self-Interpreting Inner Mental Representations

In order to escape the problems following in the wake of the explanatory strategy outlined above, most contemporary forms of representationalism conceive of inner mental representations as *self-interpreting*. This approach is inspired by modern computer technology[197]. The general idea is that mental representations can be ascribed a role analogous to the role of the internal tokens or symbols which constitute machine languages. These internal tokens are bound by rules of syntax.

[196] I.e., Inner mental representations.
[197] Because computer technology provides the basis for this explanatory strategy contemporary representationalist theories of mind are often referred to as computationalist theories of mind.

Like the higher-level input/output languages, which programmers use to program computers (e.g., Basic, Pascal, Fortran, C^{++}), they possess a representational character. Yet, in contrast to these programming languages they do not necessitate interpretation by a compiler. If this was the case, one would end up with a 'compiler regress' because compilers are written in machine languages. However, due to the engineering design of computers, the internal tokens of machine languages are "understood" by computers without compilation. As a consequence, the causally relevant properties of the internal tokens or symbols correspond to their syntax by design. The correspondence between symbols and syntax is, as it were, assumed to be design immanent. As a result, it is the causal relations existing between the internal tokens or symbols of machine languages which underlie the syntactical interpretation of the higher level input. In *The Language of Thought* Fodor describes this scenario as follows:

> 'Real computers characteristically use at least two different languages: and input/output language in which they communicate with their environment and a machine language in which they talk to themselves…`Compilers´ mediate between the two languages…What avoids an infinite regression of compilers is the fact that the machine is built to use the machine language. Roughly, the machine language differs from the input/output language in that its formulae correspond directly to computationally relevant physical states and operations of the machine. The physics of the machine thus guarantees that the sequences of states and operations it runs through in the course of its computations respect the semantic constraints on formulae in its internal language.'[198]

Although this fancy bit of speculative psychology provides an outline of how mental representations could be governed syntactically, and also intimates how mental representations could instigate action by invoking the existence of causal relations between them, the computer analogy still leaves intentionality essentially unexplained. While it depicts how mental representations might be syntactically governed, it fails to answer questions regarding their semantics or their content. The central problem – to explain how inner states can be *about* anything – remains unexplained. Importantly, this issue is not solved by extending the computer analogy in suggesting a simple causal theory of mental content[199]. Even if one granted the

[198] Fodor (1975) p.65f

[199] While gong beyond the basic computer analogy, such a theory assume that mental representations not only to stand in causal relations to each other (e.g. internal causal relations, which assumedly constitute the syntax of mental representations), but also to objects in the external world.

inclusion of the assumption that mental representations also stand in causal relations to the external world, the model would still fail to explain how the mental representations come to possess intentionality. In addition, it has also not demonstrated how the systematicity or normativity of content could be accounted for.

Fodor´s LOT hypothesis takes the form of a descriptive causal theory in order to address these issues further. Causal descriptive theories of content are inspired by the idea that effects can sometimes carry information about their causes. For example, the yellowish tone of David's skin indicates liver dysfunction, or Samantha's yellow teeth are indicative of excessive smoking and immoderate caffeine consumption. Yet, a simple causal theory, which identifies mental content with whatever causes, will not do as it fails to account for the normativity of content. In particular, causal theories fail to provide an account for situations in which entertaining a mental representation corresponds to a false belief. In the context of simple causal theories, misrepresentation would be impossible. In order to address this problem, Jerry Fodor has proposed that the causal relations, which constitute the content of mental representations, are to be defined as those relation upon which the causal relations which correspond to error, depend asymmetrically[200]. Here, the content of mental representations is determined by *independent* causal relations, while errors correspond to *dependent* causal relations. But is this really enough to account for the normativity of content? Not it seems, if one focuses on the aspect of systematicity. Despite the fact that Fodor's theory fulfils the requirement's of Evan's Generality Constraint[201] by invoking the idea of a LOT, it still seems to fall short with regard to the fact that contents stand in relations of support, compatibility and contradiction. This aspect of systematicity is problematic, as it provides further normative constraints. Consequently, a descriptive causal theory[202], which only explains how misrepresentation is possible, is not enough. Rather, it also needs to offer an explanation of how one belief can imply, contradict or support another. Thus, if I observe that there is one bulldog (Achilles) sleeping at the top end of my bed and another bulldog (Giacomo) curled up at the bottom of my bed, I can rightly draw the

[200] See: Fodor (1987, 1991)
[201] Evans defined his notion of generality constraint as follows: '*If a subject can be credited with the thought that a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception. This is the condition that I call "The Generality Constraint".' (see Evans (1980), p.104ff).'*
[202] I.e., an asymmetric dependence theory.

conclusion that there are *two bulldogs* enjoying the comforts of my nightly refuge. Can Fodor´s descriptive causal theory provide an explanation for this *rational systematicity*? In the present case of logically related thoughts, it should be able to give an account of how the right logical conclusion is possible. But on close scrutiny, Fodor´s theory seems to lack the resources to do so. Although the distinction between independent and dependent causal connections among mental representations is useful in so far as the explanation of the correct result is concerned (because it allows to distinguish between the correct conclusion and random error), the idea that the demands of rationality and independent causal relations must coincide is not justified. This belief is implicit, however, in the distinction between independent and dependent causal relations between mental representations as it presumes that independent causal relations track the right norm. It is not unreasonable, for example, to imagine an individual who is disposed to making systematic errors in drawing the wrong conclusions from her thoughts (beliefs, views, convictions etc). There is no guarantee that causation and rationality will go hand in hand. Furthermore, a similar criticism also applies to the rule governing the representation or naming of observable states of affairs. The notion that whatever is the independent cause of a mental representation is part of its content is the initial driving force behind the asymmetric dependency theory. But its plausibility is questionable as there is no reason to assume that the independent cause of a mental representation is part of its content. Thus, correctness cannot be guaranteed merely in virtue of assumed characteristics of independent causal relations. Consequently, in the overall scheme of things, the reductive explanation of mental content is questionable as purely causal processes consistently fail to account for the normativity of content and related aspects like its systematicity.

There are two lessons to be learnt here: On the one hand, explanations of content must not merely presuppose the very notion they aim to explain by presupposing the content of a mental sign or symbol. On the other hand, the kind of causal theory suggested by Fodor fails to account for both the normativity as well as the systematicity of content. Although, Fodor's computational theory of mind is just one example of a causal theory of content[203] (although a very poplar and influential

[203] There are other causal theories of content, e.g. teleological theories. However, teleological theories share with causal descriptive theories of content the assumption that mental content can be explained by postulating internal mental representations that stand in causal relations to things in the world.

one), and the previous discussion has not shown that reduction is utterly impossible, the lessons learnt from this investigation give such an endeavour very little chance for success. If anything, the burden to demonstrate that such an account is possible lies with the champions of such an endeavour.

3. On the Representationalist Account of the Intentionality of Linguistic Meaning

The problem of intentionality has troubled philosophers of the analytic as well as the continental traditions. Yet, it is within the tradition of analytic philosophy in particular, that intentionality has attracted considerable attention not only in the context of the mind, but also in the context of language. Analytic philosophers have come to think of intentionality as a property of language and linguistic behaviour as well as regarding it as a defining feature of mind and thought (beliefs, desires, etc). Statements, no less than beliefs can be about a something and posses content. For example, both my belief that "Edgar Dega's painting *Dancers in Blue* is the perfect depiction of female grace and beauty without being chauvinist", and my brother's gleeful remark that (our puppy bulldog) "Giacomo has been chewing on your Strat again", are characterised by being about something[204]. Consequently, the question arises how the intentionality of thought and the intentionality of language are related?

In the representationalist setting of the cognitive view, it is assumed that linguistic meaning derives from mental content, and that the latter can be naturalized (see above). By explaining the intentionality of language via the intentionality of thought one arrives at a reductionist account of both mental and linguistic intentionality. Descriptive causal theories of intentionality, as exemplified by Fodor (1975)[205], hold that words stand immediately for inner mental representations in the mind of the speaker or hearer, and mediately for the items in reality, which are the original causes of the ideas with which the mind is furnished. On such an account,

[204] A state of mind and a matter of fact, respectively.

[205] Note: Fodor's position with regard to the explanatory strategy of linguistic and mental content is in fact somewhat ambivalent. Consequently, Fodor can't quite be put into either the mentalist (e.g. Hume, Russell, Evans and Peacocke) nor the lingualist camp (e.g. Davidson, Dummett and McDowell). On the one hand, Fodor explains even simple behaviour by reference to a rich variety of complex thoughts and calculations, on the other he holds that these thoughts are sentences in a language of thought, a language which is not public but consists of physical tokens of computational types in the brain.

the intentionality of language is derived from the intentionality of thought and thought-constituents.

It is best to start the investigation into this question by examining the driving force behind the causal representationalist account. Two sources of puzzlement can be readily identified: First, bewilderment arises out of the fact that while we recognize words and sentences as spatio-temporal items, and thus as part of the furniture of the material world, we wonder how it is that a thing like a sound or a mark on paper can represent, i.e. be about something? Second, even more baffling seems the fact that we are not only able to talk about things that exist, but also about things that do not exist. And, not only can we represent things in language which are true, but also things which are not (e.g. PI§518). Thus, the question arises how it is that statements can be false, yet still be meaningful?

> ‘Wenn man fragt “Wie macht der Satz das, dass er darstellt?“ so könnte die Antwort sein: „Weißt du es denn nicht? Du siehst es doch, wenn du ihn benützt.“ Es ist ja nichts verborgen. Wie macht der Satz das? – Weißt du es denn nicht? Es ist ja nichts versteckt.
>
> Aber auf die Antwort „Du weißt ja, wie es der Satz macht, es ist ja nichts verborgen“ möchte man erwidern: „Ja, aber es fließt alles so rasch vorüber, und ich möchte es gleichsam breiter auseinander gelegt sehen.“ (PI§435)

> ‘If it is asked: “How do sentences manage to represent?” – the answer might be: “Don´t you know? You certainly see it, when you use them.” For nothing is concealed. How do sentences do it? – Don’t you know? For nothing is hidden. But given this answer: “But you know how sentences do it, for nothing is concealed” one would like to retort “Yes, but it all goes by so quick, and I should like to see it as it were laid open to view.” (PI§435)

Bewildered, yet driven by a burning desire to give an account of intentionality, which can acclaim scientific credibility (PI§571), cognitive scientists readily postulated some kind of mental process, a *causal mechanism,* in order to explain this mysterious property of mind and language. On the one hand, this move is partly due to the subscription to the cognitive view. On the other, as psychological predicates of the kind I think, I believe, I hope, I want, I expect, I fear etc., indicate various different states of consciousness, and because these obtain independently of the use of language, cognitive scientists assume that signs *per se* are lifeless. It takes the activity of the mind (e.g. processes of thinking, meaning, understanding or intending)

by means of which signs obtain their intentionality[206]. Wittgenstein illustrates the appeal of the idea of mysterious mental processes with characteristic clarity:

> 'Es scheint, daß es *gewisse definitive* geistige Vorgänge gibt, die mit dem Arbeiten der Sprache verbunden sind, Vorgänge, durch die allein die Sprach funktionieren kann. Ich meine die Vorgänge des Verstehens und Meinens. Die Zeichen unserer Sprache erscheinen tot on diese geistigen Vorgänge; und es könnte der Eindruck entstehen, dass es die einzige Funktion der Zeichen ist, solche Vorgänge hervorzurufen, und dass diese Vorgänge eigentlich das sind, wofür wir und interessieren sollten....Wir sind versucht zu denken, dass die Aktion der Sprache aus zwei Teilen besteht; einem inorganischen teil, dem handhaben von Zeichen, und einem organischen teil, den wir verstehen, Meinen, Deuten und denken dieser Zeichen bezeichnen können. Diese letzteren Tätigkeiten scheinen in einer seltsamen Art von Medium stattzufinden, dem Geist.' (BB_D, p. 18f.)

> 'It seems that there are *certain definite* mental processes bound up with the working of language, processes through which alone language can function. I mean the processes of understanding and meaning. The signs of our language seem dead without these mental processes; and it might seem that the only function of the signs is to induce such processes, and that these are the things we ought really to be interested in…We are tempted to think that the action of language consists of two parts; an inorganic part, the handling of signs, and an organic part, which we may call understanding these signs, meaning them, interpreting them, thinking. These latter activities seem to take place in a queer kind of medium, the mind. (BB_E, p.3)'

Further puzzlement is added by the fact that while I do not have to interpret my thoughts in order to find out what I think, believe, expect, fear etc. the situation appears to be different when I am at the receiving end. The same appears to be true with regard to the interpretation of words. While I do not have to interpret my own words, when I express what I think, believe, expect, fear etc., it appears that my interlocutor needs to do so. That is, when I hear and understand the remarks of another person (e.g. when I break out in panic after my brother has reported Giacomo's mutilation of my Strat), in the context of the cognitive view, it seems that I do so in virtue of an underlying mechanism of understanding.

> ' "Wenn ich sage 'Ich habe Schmerzen', weise ich nicht auf eine Person, die Schmerzen hat, da ich in gewissem Sinne gar nicht weiß, *wer* sie hat." Und das lässt sich rechtfertigen. Denn vor allem: Ich sage ja nicht, die und die Person habe Schmerzen, sondern „Ich habe...". Nun,

[206] See: Hacker (1996) p.3

damit nenne ich ja keine Person. So wenig wie dadurch, dass ich vor Schmerzen stöhne. Obwohl der Andre aus dem Stöhnen ersieht, wer Schmerzen hat.

Was heißt es denn: wissen, *wer* Schmerzen hat? Es heiß z.B., wissen, welcher Mensch in diesem Zimmer Schmerzen hat: also, der dort sitzt, oder, der in dieser Ecke steht, der Lange mit den blonden haaren dort, etc. – Worauf will ich hinaus? Darauf, dass es sehr verschiedene *Kriterien* der *Identität* der Person gibt.

Nun, welches ist es, das mich bestimmt, zu sagen *ich* hab Schmerzen? Gar keins.' (PI§404)

'When I say 'I am in pain', I do not point to a person who is in pain, since in a certain sense I have no idea *who* is." And this can be given a justification. For the main point is: I did not say that such-and-such a person was in pain, but "I am....." Now in saying this I don't name any person. Just as I don't name anyone when I *groan* with pain. Though someone else sees who is in pain from the groaning.

What does it mean to know *who* is in pain? It means, for example, to know which man in this room is in pain: for instance, that it is the one who is sitting over there, or the one who is standing in that corner, the tall one over there with the fair hair, and so on.--What am I getting at? At the fact that there is a great variety of criteria for personal '*identity*'.

Now which of them determines my saying that '*I*' am in pain? None. (PI§404)

'Muß ich einen Befehl verstehen, ehe ich nach ihm handeln kann? – Gewiß! Sonst wüsstest ja nicht, was du zu tun hast. – Aber vom *Wissen* zum Tun ist ja wieder ein Sprung! (PI§505)

'Must I understand an order before I can act on it?--Certainly, otherwise you wouldn't know what you had to do!--But isn't there in turn a jump from *knowing* to doing?' (PI§505)

Importantly, while all understanding appears to involve some sort of interpretation within the shadow of this picture, misunderstanding is thought to be the result of misinterpretation, and the failure to understand is thought to be the result of a failure to interpret.

Throughout the history of philosophy, various suggestions have been made with regard to the nature of a mechanism of understanding. Classical empiricist philosophers favoured an imagist account of communication[207], in contrast to the computationalist account of modern cognitive science. The modern version involves, for example, the attribution of implicit knowledge by virtue of a theory of interpretation to members of a language community. Such a theory, it is believed, is grounded in

[207] I.e., by conceiving of speech as a means to engender images or idea in the mind of the hearer.

the abstract computational mechanism by means of which the meaning of an utterance is computed. Common to both traditions, however, is the notion that in order for mere signs to be conveyed successfully interpretive acts of understanding accompany the hearing or reading of those signs.

> “Zwischen dem Befehl und der Ausführung ist eine Kluft. Sie muss durch das *Verstehen* geschlossen werden.“ „Erst im verstehen heißt es, dass wir DAS zu tun haben. Der *Befehl* – das sind ja nur Laute, Tintenstriche.-“ (PI§431)
>
> "There is a gulf between an order and its execution. It has to be filled by the act of *understanding*." "Only in the act of understanding is it meant that we are to do THIS. The *order*--why, that is nothing but sounds, ink-marks.--" (PI§431)

In the grip of this picture, the pressing question seems to be by what mechanism does the mind generate and exercise these representational powers? Yet, the futility of descriptive causal theories has been outlined above. Again, we are also reminded of the implausibility idea of a process thought, which was thought to underlie thoughtful and thoughtless speech. The prospect for success of the endeavour to give a *general* account of linguistic content by reference to another hypothesized mental process (i.e. understanding) thus seems doubtful at best. However, for the sake of clarity and completeness it is not only necessary to track down the confusions underlying the modern account of linguistic content, but also to rob them of their appeal by dispelling the underlying misunderstandings.

3.1 Understanding the Life of Signs – On Mental Symbols and Physical Processes

The representationalist characterisation of mental states as ontologically independent freestanding states of mind and body, applies to the understanding no less than to intentional mental states in general. This conception raises the question what the connection between my understanding a certain concept and my correct application of the concept consists in. In other words, when we understand an expression that has been explained to us, what happens that enables us to go on and use the expression correctly? As outlined above, representationalists attempt to explain, linguistic meaning – the life of signs – through the execution of (causal) mental processes such as, e.g., intending and understanding, implemented in the

physical symbol system of the mind. But is it plausible to assume that the possession and processing of mental symbols provide the foundation of understanding? If we imagine, for example, that we were trying to continue a mathematical series of the sort n+2, our understanding of what the numbers in this series should be could not consist in having a formula constituted by mental symbols in the mind. Neither could it consist in reminding oneself continuously that the differences between successive mathematical series should be the same. This is because both the formula, a well as the reminder could be misinterpreted and misapplied. The correct use of such mental signs is in the same need of specification as the correct use of written signs. Consequently, it is not possible to explain the understanding that governs the continuation of a written series by invoking the notion of mental signs and symbols. It is not possible to account for the understanding of a rule, which governs the use of a concept or (as in the example above) of a written mathematical series, by invoking the idea of a physical symbol system in the mind. This is a key objection against the representationalist endeavour in as far as it posits the question what would determine the correct interpretation of a particular mental state (i.e. how would I know what state I am in) if the understanding is identified with entertaining an inner mental representation or the interaction of various mental representations? The very question one sets out to explain remains unanswered.

> ‘Wir versuchen nun, den seelischen Vorgang des Verstehens, der sich, scheint es, hinter jenen gröbern und uns daher in die Augen allenden Begleiterscheinungen versteckt, zu erfassen. Aber das gelingt nicht. Oder, richtiger gesagt: Es kommt gar nicht zu einem wirklichen Versuch. Denn auch angenomen, ich hätte etwas gefuden, was in allen jenen Fällen des Verstehens geschähe,- warum sollte *das* nun das Verstehen sein? Ja, wie konnte denn der Vorgang des verstehen versteckt sein, wenn ich doch sagte „Jetzt verstehe ich“, *weil* ich verstand? Und wenn ich sage, er ist verstckt,- wie weiß ich denn, wonach ich zu suchen habe? Ich bin in einem Wirrwarr.’ (PI§153)

> ‘We are trying to get hold of the mental process of understanding which seems to be hidden behind those coarser and therefore more readily visible accompaniments. But we do not succeed; or, rather, it does not get as far as a real attempt. For even supposing I had found something that happened in all those cases of understanding,--why should *it* be the understanding? And how can the process of understanding have been hidden, when I said "Now I understand" *because* I understood?! And if I say it is hidden--then how do I know what I have to look for? I am in a muddle.’ (PI§153)

‘Aber halt!- wenn “jetzt verstehe ich das System” nicht das gleiche sagt, wie “mir fällt die Formel....ein“ (oder „ich spreche die Formel aus“, „ich schreibe si auf“, etc.) - folgt daraus, daß ich den Satz „jetzt verstehe ich...“, oder „jetzt kann ich fortsetzen“, als Beschreibung eines Vorgangs verwende, der hinter, oder neben dem des Aussprechens der formel besteht?

Wenn etwas „hinter dem Aussprechen der Formel“ stehen muß, so sind es *gewisse Umstände,* die mich berechtigen, zu sagen, ich könne fortsetzen,- wenn mir die Formel einfällt.

Denk doch einmal gar nicht an das Verstehen als „seelischen Vorgang“! – Denn *das* ist die redeweise, die dich verwirrt. Sondern frage dich: in was für einem Fall, unter was für Umständen sagen wir denn „Jetzt weiß ich weiter“? Ich meine, wenn mir die Formel eingefallen ist. –

In dem Sinne, in welchem es für das Verstehen charakteristische Vorgänge (auch seelische Vorgänge) gibt, ist das verstehen kein seelischer Vorgang. (Das Ab- und Zunehmen einer Schmerzempfindung, das Hören einer Melodie, eines Satzes: seelische Vorgänge.) (PI§154)

‘But wait--if "Now I understand the principle" does not mean the same as "The formula.... occurs to me" (or "I say the formula", "I write it down", etc.)--does it follow from this that I employ the sentence "Now I understand....." or "Now I can go on" as a description of a process occurring behind or side by side with that of saying the formula?
If there has to be anything 'behind the utterance of the formula' it is *particular circumstances*, which justify me in saying I can go on--when the formula occurs to me.
Try not to think of understanding as a 'mental process' at all.--For *that* is the expression which confuses you. But ask yourself: in what sort of case, in what kind of circumstances, do we say, "Now I know how to go on," when, that is, the formula *has* occurred to me?--
In the sense in which there are processes (including mental processes) which are characteristic of understanding, understanding is not a mental process.
(A pain's growing more and less; the hearing of a tune or a sentence: these are mental processes.)’ (PI§154)

Furthermore, even if understanding a concept was accompanied by some mental process or other, understanding itself could not consist in that process, because such a process would be neither necessary nor sufficient in meeting the normativity constraint of content. Rather, the justification for the claim that someone has grasped a rule must be founded on the success in applying the rule. This can be clarified by considering the case of reading. The ability to read cannot be defined by any characteristic mental accompaniment or process, as it is an unmediated relation of text and speech only. The driving force of this claim rests on the fact that the ability to read is justified on the grounds of repeated manifested success. And so is the

understanding of a word. Whether, I have understood a word or not is shown in the way I apply it, i.e. whether I consistently apply the word according to the rules for its use, or whether I fail in doing so. Neither a mental phenomenon nor a mental process, which someone who learns to read may entertain, is necessary or sufficient in settling the question what determines someone's ability to read. In other words, rule following cannot be accounted for in terms of accompanying mental processes.

Finally, to think of the understanding as a (mental) process, constitutes a transgression of the grammar of the concept to understand. The mistake under scrutiny here is, in Rylean terms, a *category mistake*. And, as Wittgenstein himself points out, it is as mistaken to conceive of the understanding as a mental process as it is to conceive of the number three as an object (PG§42). The nature of this error is best highlighted by investigating the use of the concept in everyday language[208]. It has been pointed out earlier, for example, that processes are characterised by a sequence of events or actions, which may or may not be interrupted, repetitive, elegant etc. (see chapter IV). The quintessence of these reflections was that processes *take time*. It has also been acknowledged that Wittgenstein was far from denying the existence of mental processes, if such talk was meant to refer to a process like, for example, reciting E.A. Poe's *The Raven* in my mind on my way to the library, or assessing the size for the market of golf balls in the US (in my mind), whilst waiting for Cécile at the cinema. These few remarks suffice to underline the claim that understanding is not a (mental) process, in the sense implied by the cognitive view. Although, it is true that when I attend a philosophy lecture, or watch a dancesport competition, I undergo a different experience (or set of experiences), which are crucially different from those experienced by someone who has not studied philosophy, or who is not well acquainted with the sophistications and intricacies of Modern Ballroom and Latin dancing. Importantly though, these experiences are not *characteristic marks* or criteria of understanding. They may vary from case to case, and from person to person. Even if such experiences were characteristic of understanding, it would not constitute understanding, as others do not ascertain themselves of my understanding of philosophy or dancesport by discovering *what*

[208] Note: In the *Philosophical Investigations,* Wittgenstein's identifies and discusses a number of misconceptions regarding the understanding (e.g. the conception of understanding as an experience, disposition etc.). However, the following will only survey examples of understanding, which pertain to the conception of understanding as a mental process, which is of direct relevance to the investigation of the cognitive view.

processes accompanied my listening to the lecturer or my watching the couples on the floor. A mental process is neither necessary nor sufficient for understanding (see also above).

What indicates whether I possess an understanding of philosophy may be determined, for example, by my ability to answer questions pertaining to the subject coherently and intelligently, or whether I can assess and explain the relative significance and virtues of different philosophical positions (or lack thereof). Alternatively, whether I possess any expertise with regard to Ballroom and Latin dancing, may be displayed in conversations about the quality and proficiency of Beata's backstep(s) in the Rumba, or the rhythmical expressivities of Bryan's cha-cha-cha or Pino's Tango, which I frequently engage in with my friends. Moreover, whether Nick, Tim, Dave or I have any understanding of Ballroom and Latin dancing may (to the connoisseur) also be evident in the aspects of the dancing, which we appreciate versus the aspects, which we may fail to appreciate[209]. These examples, underline the crucial difference between the criteria indicating understanding and the criteria indicating what is going on in my mind or brain. Somebody who is interested in finding out whether I understand anything about philosophy or dancesport, for example, is not interested in what is going on 'inside me', as it were, i.e. what I am undergoing. What that somebody is interested in lies in what I can do. In contrast to a process, understanding something is not interruptible, like my recital of *The Raven* or my assessing of the US golf ball market (which is). Wittgenstein points out that,

> 'Man sagt wohl überhaupt kaum, man habe etwas seit gestern "ununterbrochen" geglaubt, verstanden, beabsichtigt. Eine Unterbrechung des Glaubens wäre eine Zeit des Unglaubens, nicht z.B. die Abwendung der Aufmerksamkeit von dem Geglaubten, z.B. der Schlaf. (Unterschied zwischen „knowing" und „being aware of".) (Z§85)

> 'One hardly ever says that one has believed, understood, intended something "uninterruptedly" since yesterday. An interruption of belief would be a period of unbelief, not e.g. the withdrawal of attention from what one believes--e.g. sleep. (Difference between 'knowing' and 'being aware of'.) (Z§85)

[209] Although it is practically impossible not to note Beata's fine aspects, i.e. qualities, which are evident even to those uninitiated to the world of competitive dancesport.

The same goes for understanding. The best approximation to an interruption of the process variety in the case of understanding is the temporary loss of understanding, or failure to understand, but not an interruption of the process due to withdrawal of attention[210]. With these examples in mind, one is able to appreciate the folly in the association of the understanding with a process more clearly. My sudden understanding of Cécile's rapid French, which I usually struggle with, is neither an articulated process like Cécile's speaking in French (e.g. the utterance of a sentence), nor is it an unarticulated process. It is not a process at all. Similarly, understanding the intricacies of Modern Ballroom and Latin dancing is an ability, which has been gradually acquired over time, rather than a process with a characteristic beginning, middle and end. Finally, my sudden understanding of the rules of the game, which the children outside my window are playing and which I failed to grasp on initial observation, is not something, which goes on, as processes do. It may happen in an instant, and last as short or as long as I care to watch or care to remember.

Although the present discussion of the representationalist strategy regarding the conception of understanding as a mental process provides some crucial insights into the nature of the misunderstandings informing this view, an investigation into the confusions of this conception would not be complete without trying to understand the origins of this illusion. First, as Peter Hacker points out, the temptation to conceive of understanding as a process is partly due to the similarity between the beginning of a process and the dawning of understanding. Yet, although the genesis of one's understanding of something or somebody may be accompanied by a characteristic experience as on occasions when "the penny drops", this event does not mark the beginning of a process. Furthermore, while the gradual acquisition of an ability may be process like[211], as in the case of mastering a complex technique, for example, the

[210] See: Baker & Hacker (2005), p.370

[211] For example, in cases where this involves the mastery of a calculus such as the multiplication table, a system of rules like those underlying the conjugation of Latin verbs. In these cases, the relative state of proficiency can easily be described as insufficient, not bad, moderate, good, excellent etc, and thus like a process can be partitioned into distinct phases. In other cases, however, as in the acquisition of expertise in Ballroom or Latin dancing, the gradual acquisition of understanding is less process like. Although, my understanding of aspects of Ballroom and Latin dancing may equally be said to be insufficient, moderate, good, or excellent, the transition between the respective phases is a lot less clear cut than in the case of mastering a calculus. Binary value, I either have it, or not. And while there may be a definite beginning to my acquisition of expertise in Ballroom and Latin dancing, on the day when I too my first steps on a dance floor, for example, the development of my understanding is an ongoing thing.

process is not the understanding. Rather, the understanding is the successful outcome of the process[212]. Second, there is the temptation of conceiving of understanding other people as involving interpretation, and hence to think of understanding as a process (see 2.1 above). Yet, it is not possible to *interpret* a sequence of signs that is unintelligible to one. One can only *ask* for it to be translated or deciphered by someone familiar with the underlying 'grammar' of the symbolism. Interpretation thus presupposes understanding, but it cannot explain. While interpreting is an activity or process, understanding is not[213].

3.2 On the Nature of the Understanding

The concept of understanding is a fluid one, Wittgenstein observed early on in his reflections on the subject (PG§5). The broad ramifications and opaque boundaries of the use of the concept alluded to by Wittgenstein in this remark, call for patience when investigating its grammar, and stipulate a careful scrutiny of those concepts, which stand in close proximity. The first steps towards gaining an insight into the nature of the concept of understanding have already been taken, by undermining its association with a mental process. This part of the present investigation has highlighted the proximity of the concept of understanding with the concept of ability:

> 'Die Grammatik des Wortes "wissen" ist offenbar eng verwandt der Grammatik der Worte „können", „imstande sein". Aber auch eng verwandt der des Wortes „verstehen". (Eine Technik „beherrschen".) (PI§150)

> 'The grammar of the word "knows" is evidently closely related to that of "can", "is able to". But also closely related to that of "understands". ('Mastery' of a technique,).' (PI§150)

There exists a definite affinity between the words "can", "is able to" and "understands". Yet, while this remark goes some way in underlining the suggested relationship between understanding and ability, the identification of understanding with an ability turns out to be a overhasty generalization on close scrutiny, which is thus to be avoided[214]. Like abilities, understanding what a word means lacks genuine

[212] Baker & Hacker (2005), p.370
[213] ibid. p.371
[214] See also: Baker (2003), p.357

duration. This is in contrast to the fore mentioned processes as well as to mental states or experiences, for example. A person who understands, is able to do certain things, Baker and Hacker[215] remind us. Yet, they emphasize that to simply categorize understanding, as ability is misleading. Rather, they highlight the fact that Wittgenstein was more comfortable with the idea of thinking of understanding as being *akin* to an ability. This formulation acknowledges that understanding indicates ability in the majority of its uses, but leaves plenty of room for the exemptions to the rule, as it were[216].

First, if understanding is conceived of as an ability proper, it is not an ability to do one single type or class of things. This is an important differentiator, which sets understanding apart from a number of abilities, even though this does by no means extend to abilities in their entirety. As German is my mother tongue, I obviously have little difficulty in understanding sentences formulated in German. This is exhibited, by my ability to paraphrase them, explain them, translate them, respond to them in appropriate ways etc.. Likewise, a politician who understands people can make himself popular, can appreciate their wants and needs, can recognize their motives and win their favour in an election, for example. This variety of abilities instantiates the various shapes and forms which understanding may take, and thus highlights the numerous criteria signalling understanding.

Second, it has been pointed out that understanding tends to be more passive than some of the abilities who have the same object[217]. While I am able to understand Italian, my ability to speak Italian is, comparatively speaking, non-existent. Thus, while I can understand an Italian sentence, and explain what it means, I may still be unable to speak Italian. Similarly, the fact that there are more high profile Ballroom and Latin judges than there have been World Champions in these discipline indicates, that understanding Ballroom and Latin dancing is not the same as the ability to dance a World Class Tango, Foxtrot, Cha-Cha or Jive. Moreover, it is important to note that there are degrees of understanding, which are not manifested in the degrees to which an ability is possessed. For example, it is unfortunate but true that despite my continuously increasing understanding of

[215] Baker and Hacker (2005), p.381
[216] ibid. p.381
[217] ibid. p. 381

Ballroom and Latin dancing, is not matched by my ability to dance a (better) cha-cha-cha, or Jive. Degrees of understanding do not necessitate increased proficiency in some activity. Thus, while the grammar of the concept of understanding runs parallel to the concept of ability for a while, the understanding is not identical with an ability. Understanding is not a mental process[218], but is best conceived of as enjoying a close relationship with being able to do something[219].

4. Mental and Linguistic Content Revisited

Thus far, the present investigation has shown the implausibility of the claim that the correct application of concepts over time is based on a mental process of understanding. In the previous chapter an outline of a positive account of mental and linguistic content has been sketched. In the context of this account, concepts were identified as the building blocks of content and the crucial role of the subject of experience as a concept user has been highlighted (see chapter V). The following discussion will underline and further substantiate these insights by suggesting an understanding of the normativity of mental and linguistic content as being shaped in interpersonal contexts of rule guided activities.

4.1 Rules and the Normativity of Content

Following in the footsteps of Immanuel Kant and Ludwig Wittgenstein, the previous discussion of judgements and concepts has emphasized the fact that a concept is a *normatively* constrained function of judgement, which are tied to rule-governed practices involving the terms of a natural language. For someone to grasp a concept, is for someone to respond in a principled or rule-governed way, and to acknowledge that there is a way of going right and going wrong. The mastery of a concept is displayed by exercising one's ability to judge the applicability or inapplicability of a given concept across various situations and circumstances. On this view, concepts are independent of the mind of any given subject. They are neither Platonic entities nor private Cartesian ideas. Above all, they are not inner mental representations processed in a language of thought. Rather, a subject

[218] Nor a metal state, or experience (see footnote 24, for details).
[219] See e.g.: Baker & Hacker (2005), p.385

masters a concept, when he applies rule-governed techniques of responding to the world. For example, if in responding to the behaviour of my English bulldog Giacomo I use the word 'Giacomo' to describe what he is doing, and my use of the word is correct and displays the correct links to other areas of discourse, I can self-ascribe the mastery of the word 'Giacomo'[220]. The mastery of a concept, is essential to the ascription and self-ascription of *propositional attitudes (PAs)* involving a given concept. These constrains imply not only that there is a link between understanding and manifestation (on the basis of which the subject of experience has been characterized as an identifiable entity engaged in interpersonal activity) which accounts for the normativity of content, but also suggest a non-causal view of mental explanation and account of human action (see 4.2 below). One's tendency to apply a certain concept can be well or ill grounded, and in using a concept one acknowledges and is answerable to a norm which governs whether one is right or wrong in applying the concept in question. Norms also govern the judgements in which a thought (comprising conceptual elements) is assessed for its truth. The subject judges that an item counts as an instance of a concept he is using, like 'guitar', for instance, and also judges whether the conceptually framed thought like e.g., 'this guitar is a 1952 Gibson Les Paul *Goldtop*', is true. In both cases the judgements implicitly answer to prescriptive norms.

At this point, it is important to be aware of the distinction between *following* a rule and *merely responding in conformity* with a rule. A subject who follows a rule, models his responses on the set of rules endorsed by fellow subjects of a given language community. A subject holds himself answerable to this set of rules, i.e. set of prescriptive norms. In doing so the subject not only acknowledges that there is a difference between going right or going wrong but also that his tendency to respond thus and so can be evaluated for its correctness. Consequently, a subjects behaviour (under normal circumstances) is generally aimed at fulfilling a normative regularity. In contrast, someone might just feel an overwhelming urge to utter the word 'guitar' whenever such an instrument captures his glance. However, unless the subject is able to provide other members of the same language community with additional

[220] Note: The links concerned do not constitute a set of necessary and sufficient conditions for (self-) ascription of the mastery of a given concept, since the fact that 'Giacomo' is the name of an 'English bulldog' may be part of the full grasp of the word 'Giacomo', but may not be known by a child who is not yet able to distinguish between different breeds of dogs (although a child may correctly identify Giacomo as a dog), nor may it be known by adult individuals who don't know a lot about dogs.

evidence[221] for his mastery of the word 'guitar', the automatic utterance of this word *per se* does not count as evidence for the subjects ability to apply the word 'guitar' correctly. Under such circumstances, it is quite plausible that while a subject acts in conformity with a rule, he does so unwittingly. Thus, to answer to a prescriptive norm implies a certain self-awareness and self-control, which enables a subject to respond according to a shared standard of correctness. To master a concept is to possess a certain form of understanding, e.g. to master the concept of 'English bulldog' is to understand what it is for something to be an 'English bulldog'. Consequently, to master the concept 'English bulldog', one must not only have an understanding of what an English bulldog is, what such a dog looks behaves like etc., but also know what the concept 'English bulldog' is to be applied to.

The Intentionality of Language - Linguistic Meaning

> 'Man kann für eine *große* Klasse von Fällen der Benützung des Wortes Bedeutung – wenn auch nicht für *alle* Fälle seiner Benützung – dieses Wort so erklären: Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache.' (PU§43)

> 'For a *large* class of cases – though not for all – in which we employ the word 'meaning' it can be defined thus: the meaning of a word is its use in the language. …' (PI§43)

The later Wittgenstein's reflection on language provided the grounds on which his insights about the philosophy of psychology cold flourish. Wittgenstein's ideas about concepts and rules were crucial to the earlier discussion of thinking and mental content as they not only provided us with an understanding of the intentional aspect of the mental, but also provided us with an account of the normativity and structure of mental content. In some of his lectures given in the early 1930s, Wittgenstein proposed an understanding of linguistic meaning, i.e. the meaning of words or concepts, as being determined by the rules for the use of that word (see: AWL 30), and as conceiving of the use of words in speech as a rule-governed activity. As outlined in chapter II, the rules for the use of a word constitute its grammar[222]. In

[221] An awareness (or unawareness) of mistakes of the application of the word 'guitar' , might count as such evidence for example.
[222] 'To grammar belongs everything that determines sense, everything that has to be settled antecedently to questions about truth.' Baker &Hacker (2005), p.145

other words, the meaning of a word is given by explanations of meaning, which in turn are rules for the use of the word or concept in question (see e.g. PG 23, 133). Thus, rather than deriving their meaning from mental representations, words, concepts, linguistic expressions and utterances are given meaning by the rules governing their use. By linking meaning to rule governed practices which are endorsed by a community of language users (sharing a range of human activities or form of life) it is possible to give an account of linguistic meaning and the intentionality of language that is free of the confusions informing the notion of derived intentionality emerging in the context of the cognitive view. Linguistic meaning, thus understood, is intersubjective. Its relation to truth and falsehood is constituted by those practices through which the use of a word is defined and refined. Both the acquisition and the refinement are public and regulated by norms which regulate and guide the response of the individual.

4.2 Mental Explanation, Human Action and the Cognitive View

The dominant model of mental explanations is based on the notion of causality. It aims to explain human agency by reference to antecedent mental states and events, which stand in a causal relationship to human action. This explanatory strategy follows from one of the basic tenets of the cognitive view, the idea that mental representations constitute the inner termini of the causal chains that instigate action, thus implying some sort mechanical interaction between mental representations. In the guise of methodological solipsism[223], for example, an agents mental events and states are thought of as the causes of actions. Furthermore, it is thought that the cognitive role of a given mental event or state determines its role in the explanation of action. Of course, the account of human agency proposed by the methodological solipsist, is incompatible with an analysis of mental content in which content ascriptions involve the world, and in which a subject's responses to objects is developed through the acquisition and application of rules for the use of concepts[224].

[223] Note: Methodological Solipsism has its origins in Hilary Putnam's *'The Meaning of 'Meaning''*. According to this view, the mental state of an individual does not presuppose the existence of another subject. It only presupposes the existence of the individual who is in a mental state of some sort. Thus, methodological solipsism ties mental ascriptions and their contents to (mental) states (i.e. tokenings of metal representations, in the context of the cognitive view) in the thinker.

[224] Note: The following discussion will only investigate the position of the methodological solipsist with regard to mental explanation and agency. While an important contribution to these issues has been made by Donald Davidson, a discussion of his ideas is beyond the scope of the present investigation.

As PF Strawson has pointed out mechanical transactions are fundamental to our notion of causality. Consequently, it is not surprising that whenever the notion of causality is invoked, the notion of mechanical interaction is not far behind. It is in this way that some advocates of causal theories of mental explanation have, by way of metaphor, been lead to metaphysics:

> '…we should regard mechanical transactions as fundamental in our notion of causality in general….It is not then to be wondered at that such transactions supply a basic model when the theoretical search for causes is on; that we look for causal mechanisms; that, even when it is most clearly metaphorical, the language of mechanism pervades the language of causes in general, as in the phrases "causal connection", "causal links" and "causal chain".[225]

This remark indicates that causal explanations only get a grip in cases where mechanical links between states and events provide a coherent story, e.g. in the context of an account where mechanical processes connect states and events. Yet, if one searches for corresponding mechanical links in the context of mental states and events a metaphysical mismatch ensues. In the preceding discussions of thinking and mental content it has been repeatedly highlighted that thoughts and their content depend on the way a subject uses concepts to structure his activity. But this aspect is not captured by an account in which (spatio-temporal) states and events interact with each other in a causal manner. This line of argument has already been adopted by Kant, who also refuted the idea that content laden mental states, i.e. PAs, are causal states and that they obey physical laws. Wolff, for example, has pointed out that according to Kant, the paradigmatic case of rational action is a case in which *a.* I form a concept of some event, object or state of affairs, which I choose to bring into being, and *b.* I do something, which I believe will actualise that which my concept represents. Thus, subjects act in order to realize a certain goal or end. A subject that acts thus, acts according to its thoughts. Importantly, however, it acts in accord with his thoughts *qua* representations with *cognitive significance,* and not in accord with thoughts *qua* mental events, which have a temporal location and thus, phenomenal causes and effects[226]. Furthermore, in the *Critique of Pure Reason Kant* argues that:

[225] Strawson (1985), p.124
[226] See: Wolff (1973), p. 111

'…in judging free actions wit regard to their causality we can only get as far as the intelligible cause, but not *beyond it*. (CrPR, B585)'[227]

This view is compatible with the results of the recent investigation into thought and thinking, which suggested that thoughts are not bits of reality, and that having a thought is not possessing some thing which stands in a material relationship to other extra-mental things, but is rather to be understood as an interlocking network of rule-governed abilities (if one feels the need to give a "general account" of thinking). On the other hand, Kant disputes that explanations of action necessitate references to rational determinations or reasons of the acting subject. Rather, in the light of Kant's reasoning an explanation of action needs to put the reasoning subject at the centre, by focusing on his judgements, i.e. his use of concepts, as the source of action.

A correct explanation of action, necessitating reference to the actual reasons underlying a subjects behaviour, highlights the connection that exists between the ascriptions others make of an acting subject and the subject actions as such. This connection is not mechanical, as the subject must freely hold those reasons, whose force is normative and not physical or deterministic. Just as in thought, judgements are made according to the normatively constrained applicability of concepts rather than according to mere causal conditions, so, in action, the subject's physical activity is formed and directed by his conception of the world. These conceptions are, of course, constructed out of concepts. Consequently, if concept use proceeds without compulsion and if it is responsive to the claim of reason, it imparts this same feature to action. If one accepts that the ability to use concepts is governed by prescriptive norms, one implicitly concedes that both theoretical and practical reason involve 'oughts' and not just dispositions to move thus and so. This awards a special status to mental ascriptions, in so far as they have properties which are unlike those of physical states. To a certain extent ascriptions of knowledge or of beliefs and desires, for example, involve a subject's 'making up his mind', as it were. In contrast to physical states, which can be discovered by evaluating and analysing mind-in dependent evidence, mental attributes come into being when a subject 'makes up his mind'. In the case of belief, for example, one not merely reacts, but rather has to

[227] Note: '*By freedom, on the other hand, I mean the power to begin a state on one's own. (…)reason creates for itself the idea of a spontaneity that can, on its own, start to act – without, i.e., needing to be preceded by another cause by means of which it is determined to action in turn, according to the law of causal connection.' (CrPR B561f.)*

answer to a given norm, e.g. to form a belief on the basis of (true) evidence, as one takes it to be. Consequently, whenever one forms a thought, belief, intention, or desire one could say to oneself 'I do feel inclined this way, but is that how it should be?'. Importantly, if one experiences a degree of uncertainty as to where one stands then one will not experience thus because one is unable to obtain certain facts about certain states inside oneself. Rather, in the light of certain norms, one was not yet able to come to a decision. This aspect points at an interesting connection with moral judgements. In self-ascriptions and mental ascriptions (to others) like in moral judgements, there are implicit prescriptive norms to which the subject's inclination must be sensitive. Because one's recognition of such a norm is part of mental self-ascription, one finds reason or spontaneity, i.e. the ability to act thus and so for reasons rather than as a result of antecedent conditions (CrPR B561f.), at the centre of our mental life. When acting, a subject structures his activity in accordance with rational determinations.

The present analysis emphasizes the fact that talk of thought and action concerns the rational control, by a thinker, of his own activity, and thereby explains how it reveals his mental life. A mental explanation appeals to the ways an agent reasons and thus it concerns the rules which articulate his activity. The nature of reasoning displayed by human beings, and consequently the structure and content of mental explanation only emerge when one considers them as rational and social beings. Mental explanation tells us which concepts are being used to shape an action. Concepts involve rule-governed links between a subject's behaviour and the world and thus determine the way that an action is sensitive to that world. The same concepts make linguistic interaction with a subject possible. By suggesting an explanation of action which centres on what an agent thinks about things rather than providing a description of a causal chain, the ascriptions involved and the ways they fit together appeal to a far richer conception of human beings and their relations than those that would be allowed within the constraints of the cognitive view.

5. Concluding Remarks

The present chapter investigated the reductionist account of intentionality inherent in the cognitive view, which attempts to explain linguistic content as resulting from mental content and then to give a reductionist account of the latter. In addition to scrutinizing this notion of explanatory priority and the hypothesized causal role of mental states inherent in this account, the plausibility of this explanation with respect to the normativity and structure of content was examined. It was demonstrated that the representationalist distinction between intrinsic and derived intentionality is implausible as it involves either an infinite regress or implies that one could never fully convey the full meaning of a sign. Furthermore, the discussion highlighted the normative structure of language and mind. Participating in a language game is tantamount to acquiring a (human) mind. Linguistic meaning cannot be explained as the result of the animation of otherwise dead signs by acts of understanding as has been postulated by Fodor, for instance. Fodor's attempt to naturalize mental content through the provision of a causal explanation[228] in the context of his LOT hypothesis fails. Mental content cannot be explained as a result of freestanding internal mental representations. Words are not injected with meaning through acts of understanding. Instead, their meaning is their use. Mental states are not internal freestanding states of the mind, which have to be connected via mechanisms with the world, but are intrinsically relational states. Freestanding internal representations cannot account for the normativity of content. They either presuppose what they set out to explain or fail to sustain normativity.

[228] In Fodor's particular case this took the form of a purely descriptive causal theory.

Chapter VII

Concluding Remarks

1. The Cognitive View Revisited

The present investigation subjected the basic tenets and fundamental premises underlying the explanatory framework adopted by cognitive scientists - the *cognitive view* – to a detailed philosophical investigation. The philosophical ideas of the later Wittgenstein provided the backbone of this investigation. In the light of the numerous flaws, confusions and misconceptions, which this study has uncovered, the cognitive view emerges as an entirely inadequate framework within which the possession of psychological attributes by human beings could be explained. The discussion in chapter III excavated the most basic claims which cognitive scientists and philosophers make about mind and brain, before proceeding to demonstrate that many of these common and accepted ways of thinking about and explaining mental phenomena make use of a degenerate form of Cartesianism (crypto-Cartesianism, brain/body dualism). This degenerate form of Cartesianism is exemplified, for instance, by the widespread tendency to ascribe psychological attributes to the brain and parts of the brain. However, by investigating the grammar of psychological predicates it has been shown that such ascriptions constitute a violation of mereological principles, which imply a transgression of the bounds of sense. Importantly, these violations tend to conceal the fact that the proposed *explanans* does in fact not explain anything, thus leaving cognitive scientists and philosophers with an illusory understanding of many mental phenomena. Following the discussion of mereological errors in cognitive science chapter III also highlighted the profound interconnection existing between the tendency to violate the principles of mereology and the numerous misconceptions and conceptual confusions inherent in the intuitively appealing and widely spread picture of the human mind as a private entity, to whose contents one has privileged access through introspection. It was shown that the mythology underlying this Inner/Outer picture, is the result of profound conceptual confusions regarding the nature of introspection, privacy, and the nature and foundation of language. In the course of investigating this mythology, it was highlighted that claims like *“I know that I am in pain”,* for example, are not epistemic but grammatical claims. Thus, contrary to popular believe different people can have

the same experience, share the same feeling etc., as inner states stand in need for outward criteria. The analyses and investigations conducted in this chapter were based around examinations of Kenny (1984, 1989) and Bennett and Hacker (2003) and thus covered territory, which may have been familiar to some. However, in order to provide a comprehensive picture of the variety of confusions and misconception underlying the cognitive view as well as their subtle yet intimate interconnections, it was of the utmost importance to start the present investigation by following in the footsteps of these principal investigators.

Chapter IV, continued the conceptual investigation initiated in the preceding chapter, by investigating the concept of mind and highlighting the fact that contrary to the cognitive view, which views the mind as a biological information processor, minds are not entities of any kind. Rather, if one feels pressed to give a general account of mind, it is best to adopt an Aristotelian position and think of the mind as the capacity to acquire intellectual skills. To think otherwise constitutes a transition from the metaphorical to the metaphysical. Furthermore, the discussion of chapter IV identified the mind-body problem and the problem of the explanatory gap as two examples of the type of confusion and problem the entity view of the mind gives rise to. In abandoning the entity view of the mind, however, the mind body problem as well as the explanatory gap, dissolve and vanish from view. After concluding the examination of the entity view of the mind, the notion of mental representations was subjected to a thorough scrutiny. This part of the discussion revealed that the cognitive scientist's understanding of mental representations as (symbolic) descriptions (in the ordinary sense of this term) is fundamentally flawed, as descriptions in this sense are not to be found in the brain. For something to qualify as a symbol, it must have a rule-governed use. Despite intuitions to the contrary, neural activity in the brain, however, does not qualify as a symbolic description of any kind as it lacks a rule governed use. For something to be a symbol there must be correct and incorrect ways of applying it (i.e. the concept of a symbol is correctly only applied to those forms of expression to which a standard of correctness can be applied). The neurons of the brain, however, do neither know nor do they not know what any array of symbols means. Similar difficulties were highlighted and discussed with regard to the notion of representations and maps. As in the case of symbolic representations such systems of representation imply the agreement upon and the usage of conventions, i.e. rules

of representation. And, as in the context of the notion of symbolic representation, there exist no conventions of representation, which are not vindicated by their intentional use by individuals, and which makes them able to apply these conventions. Yet, neither brains nor neurons can be said to employ symbols (or maps for that matter) as they neither know nor are ignorant of what the symbols or maps mean. Similarly, both brains and neurons can neither be said to follow nor can they be accused of failing to follow rules for the use of symbols. Thus, while certain features or stimuli in ones sensory field can indeed be mapped, (that is causally correlated with), onto the firings of cortical neurons, indicating an inductive correlation, talk of mental maps in the sense of neural "communication of meanings by topological analogies", constitutes a transgression of the bounds of sense. Finally, chapter IV also examined the notion of storing representations in the brain, a claim that is integral to the cognitive view (see above). However, the discussion showed that the notion of a stored representation can only make sense if the representation can be accessed and is available *to a person*, which could read and recognize the representation and potentially tell somebody what it is a representation of. But cognitive scientists and philosophers have yet to specify the criteria for identifying what counts as storing a landscape representation, for example, in the brain. Yet, as they use the concepts of representation and storage in their customary way when talking about the storage of representations in the brain they unwittingly transgress the bounds of sense.

Chapter V focused on examining the notion of thought and thinking inherent in the cognitive view, viz. the cognitive conception of thought, which portraits thinking as the processing of mental representations in a language of thought (LOT). However, as the conceptual investigation of this chapter showed, the cognitive conception of thinking fails to take into account the polymorphous character of thinking, instead providing an account of the nature of thought which could not be supported by the most rudimentary philosophical, i.e. grammatical, analysis. Above all, it ignores the crucial link between thinking and the expression of thought, a link which is forged through the normative constraints governing the use of concepts in thought, and the various judgments involved in the ascription of concepts. In line with the Aristotelian conception of mind proposed in chapter IV, it was suggested to conceive of thinking as the ability to participate in an interlocking network of rule-governed abilities.

Through the exercise of this ability, i.e. by “thinking” an individual is able to pick out common features in those situations in which the rules are applied and exercised. Thinking does not consist in computational operations upon sentences of *mentalese,* but in the various forms of linguistic and non-linguistic behaviors, which characterize the expression of thought. For a creature to think is not to have rational-symbol manipulation processes occurring in the mind/brain, but to exercise conceptual judgment according to rules governing the use of concepts. The abilities, which constitute thinking are structured and depend on a flexible selectivity of attention and response to the world. The exercise of these abilities comprises thought with content rather than merely a complex of causal influences, which explain the subject’s behavior. Both the self-ascription, and the ascription of mental content to others have their origins in these abilities. All human beings are endowed with the capacity to develop these (congruent) abilities, which provide the foundation for the learning and exercise of rule-governed techniques, and which imply an agreement in judgments. It is this relation between thought, the expression of thought and rule governed abilities, which allows for structured and generalizable ways of thinking.

Chapter VI, examined the account of intentionality inherent in the cognitive view, which attempts to explain linguistic content as resulting from mental content and then to give a reductionist account of the latter. In addition to scrutinizing this notion of explanatory priority and the hypothesized causal role of mental states inherent in this representationalist account, the plausibility of this explanation with respect to the normativity and structure of content was examined. In the course of the following discussion, it was demonstrated that the representationalist distinction between intrinsic and derived intentionality is implausible as it involves either an infinite regress or implies that one could never fully convey the full meaning of a sign. The failure to recognize the normative dimension of language and mind was thus identified as the fundamental error of the representationalist stance. In the context of the cognitive view (linguistic) norms have mistakenly been regarded as metaphysical claims. This is how otherwise banal ordinary statements could be transformed into metaphysical pictures of a deep underlying reality hidden behind the everyday surface. By demonstrating that language is normatively structured, and by highlighting that participating in a language game is tantamount to acquiring a mind, the discussion emphasized that linguistic meaning cannot be explained as the result

of the animation of otherwise dead signs by acts of understanding as envisioned by Fodor, for example. Fodor´s infamous LOT hypothesis constitutes an attempt to naturalise mental content through the provision of a descriptive causal explanation. Yet, as the discussion of chapter VI showed, mental content cannot be explained as a result of freestanding internal mental representations. Words are not injected with meaning through acts of understanding. Instead, their meaning is their use. Mental states are not internal freestanding states of the mind, which have to be connected via mechanisms with the world, but are intrinsically relational states. Freestanding internal representations cannot account for the normativity of content. They either presuppose what they set out to explain or fail to sustain normativity.

As these discussions illustrate, the cognitive view has been built on a host of flawed and misconceived presumptions and premises. Furthermore, it proves itself to be not only the product of conceptual confusion and metaphysical speculation, but is also to be regarded as a source of further philosophical confusion and misunderstanding itself, thus distorting contemporary thought about the mind and human nature (see section 3 below). Consequently, it cannot be regarded as an adequate explanatory framework, within which the mind and mental phenomena could be profitably studied. Any “insights” derived within the context of this framework hitherto, should be re-considered and re-evaluated with the utmost care, as they are more likely to betray illusion than understanding. The preceding discussions also underline that at its most general, cognitive scientists endorsing the cognitive view tend to fail in achieving their ultimate goal, - the understanding and explanation of the possession of psychological attributes by human beings. This is mainly due to the uncritical adoption of a questionable philosophical legacy, and a general misunderstanding of the nature and limits of science (and philosophy). These flaws are not so much the result of failing intellectual capacity, but the detrimental effect of a misguided believe in the powers of science and a consequent spread of scientism, which has hampered the education of scientists of various disciplines. Bemoaning this fact in 1944, Albert Einstein wrote to a friend:

> “I fully agree with you about the significance and educational value of methodology as well as a history and philosophy of science. So many people today – and even professional scientists – seem to me like someone who has seen thousands of trees but never seen a forrest. A knowledge of the historic and philosophical background gives that kind of independence from prejudices of his generation from which most scientists are suffering.

> This independence created by philosophical insight is – in my opinion – the mark of distinction between a mere artisan or specialist and a real seeker after truth."[229]

As the present investigation highlights, Wittgenstein's later philosophy[230], by impressing on us the importance of conceptual clarity, enables us to seek this independence and free ourselves from the prejudices and misconceptions we have inherited from our scientific ancestors successfully.

2.On the Possibility of a Science of the Mind

If one accepts the critique of the cognitive view as an inadequate framework for the explanation of the possession of psychological attributes by human beings, has one thereby accepted the impossibility of a scientific study of mind? Not necessarily. The answer to this question largely depends on what kind of conception of mind provides the framework of investigation for scientists, and on whether the nature and limits of science (and philosophy) are understood correctly. What the present investigation of the cognitive view underlines, however, is that problems and questions about the mind, which were created by conceptual confusion and perpetuated through the creation of bad philosophical theories, cannot be addressed by empirical scientific means, regardless of how sophisticated these might be.

In the process of describing their discoveries, cognitive scientists presuppose a host of psychological concepts. It is important to bear in mind that, in order to determine what is *true or false* about any given phenomenon, it is necessary to start out with descriptions, which *make sense.* In order for empirical investigation to culminate in the genesis of understanding (and not illusion), it depends on conceptual clarity, for which philosophical speculation is no adequate substitute. Misunderstandings regarding the nature of science (and philosophy), the failure to acknowledge the crucial distinction between factual and conceptual questions

[229] Einstein (1944)

[230] Heinrich Hertz, who exerted a significant influence on Wittgenstein's thinking wrote in the preface to his *Principles of Mechanics* that *" …some kinds of vexing problems are to be resolved, not by scientific explanation and hypotheses, but by clarification. Unclarity is often expressed by questions about the nature of a phenomenon, but what is needed, in some such cases, is not fresh information or sharper definitions, but a clearer understanding of existing information and definitions´(see: Hertz (1956), p.1ff).*

combined with a flawed metaphysical conception of human mind and nature and the crippling effects of a latent crypto-Cartesianism, can be identified as the major flaws, which have thwarted the latest attempt to devise a *science of the mind*[231], - a cognitive science. The cognitive view bears witness to this fact.

In stark contrast to the cognitive view, Aristotle's characteristically biological (rather than metaphysical) conception of mind is uninfluenced by and free of the conceptual confusions and philosophical problems giving rise to the former and the explanations derived from it are free from their crippling effects. Conceiving of the mind as a capacity to acquire intellectual skills may thus be considered the first step towards creating such a "*science"* of the mind. Although endorsing Aristotle's conception, by itself, does neither prevent nor eliminate misconceiving the powers and limits of scientific investigation, it would certainly set such investigation on sure footing. Aristotle believed, that the attributes of all kids of individual beings are to be explained by the particular ways in which the essential form or structure of the species to which they belong is realized in the particular matter of which they are made. Any differences between kinds of beings are to be accounted for predominantly by differences in their forms or essential organization. Thus, investigations of the neural events and processes underlying the exercise of such characteristic human abilities as thinking, perceiving, remembering or imagining, can

[231] The project of creating a *science of the mind* has made several false starts, beginning with the first modern efforts to study the world of ideas in the same manner as the physicists of the era were studying the world of matter in the 17th century (e.g. Locke 1690, La Mettrie 1748, Hume 1749), and ranging to the efforts made by 20th century psychologists during the first cognitive revolution, which marked the end of a behaviourist psychology and created the *cognitive view*. These efforts also marked the abandonment of a positivist psychology in favour of a realist conception.

At the beginning of the 20th century, many theorists occupied with developing a science of the mind, were misled into thinking that the only permissible phenomena fitted to be material to be studied in a psychological *science*, would be publicly observable entities. They mistakenly supposed that such were the phenomena of the paradigmatic physical sciences, inorganic chemistry and Newtonian mechanics. This led to the restriction of legitimate categories of phenomena to those for which could be physically specified such conditions, states of the environment as were detectable with the five senses (usually only vision was employed), and bodily movements as could be registered by some inhuman apparatus. What is more, both classes of phenomena were thought of as capable of being partitioned into independent or dependent variables. This suggested that there ought to be some surveyable relationship between types of measurable events originating in the environment and equally measurable types of events originating in the organism and some, possibly complex but in principle discernible, set of laws that connected the two. The idea of explanation in psychology was also assimilated to modes and models of explanation assumed (wrongly) to be in use in physical science, apt to generate law like predictions linking simply specified antecedents and measurable outcomes. These links were often presented as *causal relations*, although in a truly positivist spirit all that was usually claimed was the existence of a statistical correlation between stimulus event types and response event types (See: Harre & Gillett (1994), p.2f).

shed light on the question which neural states, events and processes are empirical necessary conditions for human beings to be able to exercise their intellectual abilities. The neural processes of the brain, are necessary conditions for the person, whose brain it is to be going through the relevant mental processes. Such investigations, free from the kind of misconceptions the cognitive view is cursed with, have the power to help human beings understand why they are as they are, why they possess their characteristic powers and what goes on in their brains when they are exercised. And it is this kind of understanding, not the nowadays widespread *neuromythology,* which harbors the promise of enabling us not only to gain an insight into the neurobiological foundations of the mind, but also of devising successful treatments for the various dysfunctions of the central nervous system that currently haunt our kind.

Although, endorsing an Aristotelian conception of mind would be an important step towards a science of mind, free from conceptual confusion[232], a future successful science of mind would also need to be able to accommodate more than one model of explanation. The discussions in chapter V and VI, for example, suggest that there is a need to look at psychological phenomena dynamically, *viz.* as attributes of the unfolding of orderly patterns of meaningful action (directed by a network of rule-governed abilities). This opens up the possibility of there being other models for explaining the orderly progression of events characteristic of the psychology of human beings, than the mechanistic cause-effect relationship borrowed from a superficial conception of the natural sciences, as is currently the case. A further detailed discussion of such a *hybrid science* of the mind, is, of course, prevented by the limited scope of the present chapter and thesis, as it would involve a careful consideration of the necessary scientific method and underlying scientific theory. Such a discussion would need to include a discussion of the presuppositions that enter into the construction of a scientific taxonomy and the principles underlying the creation and use of models (both analytical and explanatory), hypothesis and theories, in order to provide a foundation for the analysis and understanding of public and private processes and procedures by which people use the available ‘symbolic

[232] Note: Although, this does not mean that such a science would be immune to any conceptual confusion that might rear its head in the future. As long as mankind continues to think, we are likely to fall victim to conceptual confusion. Thus, there will always be a need for critical conceptual investigation.

resources and techniques' (i.e. language and linguistic behavior) to interact socially and accomplish cognitive tasks. On the grounds of this analysis, abstract analytical or descriptive models (based on Wittgenstein's ideas of language) of the ways people accomplish these tasks could be developed. These would be based on abstractions from the task descriptions themselves. Importantly, however, unlike the metaphysical claims implicit in the cognitive view the models of mental processes to be developed would have no existential, i.e. ontological, implications. Finally, it would need to be shown how such models could be used to guide research in a new, and different "cognitive science". Such a revised and revamped cognitive science would thus be a *hybrid science*[233], which endorses an explanatory pluralism, leaving room for more than one type of explanation.

3. The Cognitive View and Human Nature

'There is an entire mythology embedded in our language' (BT, p.434), remarked Wittgenstein, who ascribed this insight to the German writer, dramatist and critic Paul Ernst. As the preceding chapters showed, this mythology is most clearly discernable in the words, idioms and turns of phrase employed in our talk about the mind and mental phenomena. However, as our conceptions of the mind and the mental have always constituted an integral part of our conception of human nature[234], it is of no surprise that the mythology surrounding the mind today also crucially influences not only our understanding of the mental but also our self-conception as human beings.

The cognitive view of the mind discussed throughout this thesis has retained a large part of its Cartesian and empiricist legacy, and in doing so contributed to the perpetuation of the respective mythology regarding the mental (see: Chapter I), while also significantly shaping our contemporary conception of human nature. Descartes infamous duality of mind and body has provided the framework of thought about human beings, their minds and their bodies since the 17th century. According to his conception of human nature, human beings were conceived to be mental substances

[233] Although never fully developed, a *hybrid psychology* had already been envisioned by Wilhelm Wundt towards the end of the 19th century.
[234] Human nature~ (dt.) menschliche Natur, Menschenbild

(minds) conjoined to material bodies. Relations of two-way causation were held to obtain between these two categorically distinct things. *Mental attributes (thought)* were believed to characterize the mind, *physical attributes (extension)* to characterize the body. Furthermore, Descartes famously held that the I is identical with my mind (since I can doubt the existence of my body, but not the existence of the *cogito*), identified the mind with consciousness (narrowly construed and misunderstood, for details see: Bennett and Hacker (2003)), associated consciousness with the private, while putting forward a conception of indubitable knowledge founded on the idea that the mind is a private realm to which the subject has privileged and infallible access. Although Descartes' conception was offered as a more correct representation of human nature and the principles that had guided the explanations of human thought, feeling and action hitherto, his ideas constituted a gross misconception and have forced a wholly inadequate framework for the representation of human nature on us. It was highly unfortunate that Descartes' vision replaced not only flawed empirical Aristotelian ideas (along with *Renaissance Vitalism* and *Neoplatonism*) but also the sound and sensible Aristotelian conception of mind[235], which had provided the first conceptual framework for the investigation of human psychology. In contrast to the unitary substance envisioned by (scholastic) Aristotelian thought, human beings became to be conceived of as a composite entity, and are now, largely due to the influence of cognitive science in the latter half of the 20th century, more or less identified with the brain. As our conception of human nature and of what it is to be a human being is closely linked to our conception of a person[236], which has a strong legal, moral and social dimension, the results and interpretation of research in cognitive science can have a profound impact on the moral and legal status of human beings. This is exemplified in current debates arguing about the status of human beings as free agents, that have ensued in the aftermath of neuroscientific experiments studying voluntary action, and which are hotly debated in academic circles as well as the feuilletons of the press. Some scholars of German criminal law, for example, have been particularly concerned about the suggestion that free will is an illusion, as the notion of free will is crucially linked to the notions of responsibility and guilt applied in this context (e.g., Kröber 2004, Lüderssen 2004 also: Roth 2004). Wolf Singer has put the dilemma thus:

[235] As Peter Hacker puts it: *"A classic case of throwing the baby out with the bath water!"* (personal communication)
[236] Note: To be a person is, among other things, to be a subject of moral rights and duties.

> ‘Die moderne Hirnforschung ist dabei, mit ihren analytischen Werkzeugen in die innersten Sphären des Menschseins vorzudringen. Das Fortschreiten auf diesem Weg bewirkt tiefgreifende Veränderungen unseres Menschenbildes, folgenreichere vielleicht als die kopernikanische Wende und die Darwinsche Evolutionstheorie. Denn diesmal werden nicht mehr nur unser Ort im Kosmos und unsere biologische Bedingtheit hinterfragt, sondern die Begründung unserer Selbstwahrnehmung als freie, geistige Wesen.’[237]

The following discussion will provide a brief example of how the cognitive view contributes to a distorted view of human nature, by reflecting on some aspects of the current debate about freedom of the will following the publication of neurophysiological studies by Benjamin Libet. Although this topic would merit an extensive discussion in itself, the following will contend itself with providing only a short sketch of the current debate. This will suffice, however, to demonstrate that due to its widespread acceptance as the explanatory framework of cognitive scientists, the cognitive view has exerted a detrimental influence on our self-conception as human beings. In addition, it will be shown that the debate, which ensued in the wake of Libet’s experiments, is a quasi-inevitable result of endorsing the cognitive view, as well as the product of conceptual confusions surrounding the thinking about voluntary action. Given the power, influence and appeal of modern cognitive science it is thus by no means a trivial matter whether we are getting our explanatory framework for the explanation of the possession of psychological attributes by human beings right.

The Cognitive View, Free Will and Voluntary Action

Over the last few years the philosophical discussion surrounding the topic of the freedom of the will, has been strongly influenced by the results of research conducted in cognitive science. The experiments, which instigated and provided the main momentum for this debate, were carried out by the neurophysiologist Benjamin Libet, who was originally interested in the relationship between neural activity and sensation thresholds (e.g. Libet 1999, Libet 2004). Libet’s initial experiments centred largely around the determination of the degree of neural activity necessary to trigger somatic sensations. However, this work soon included work on human consciousness. His most famous set of experiments seemed to show that

[237] Singer (2002), p.9

unconscious electrical activity in the motor-cortex (the so-called readiness potential (RP)[238]) precedes conscious decisions to perform volunatry acts, and thus has by many been taken as a scientific demonstration of the illusory character of free will, as unconscious neural processes appear to precede and cause willed acts. Importantly, these willed acts are retrospectively expreienced as consciously motivated by the subject[239].

In order to gauge the relationship between the readiness potential ('RP') and subjective feelings of volition and action experimentally, Libet required an objective method of marking the subject's conscious experience of the will to perform an action in time, and afterward comparing this information with data recording the brain's electrical activity during the same interval. Thus Libet used a cathode ray oscilloscope, which had been adjusted to act as a timer, and whose output was a single dot that could be made to travel in a circular motion, rather like the movements of a hand around a clock face. This timer was set so that the time it took for the dot to travel between intervals marked on the oscilloscope was approximately forty-three milliseconds. As the angular velocity of the dot remained constant, any change in distance could easily be converted into the time it took to travel that distance. To monitor brain activity during the same period, Libet used an electroencephalogram (EEG). Libet and co-workers would ask each subject to sit at a desk in front of the oscilloscope timer. They would affix the EEG electrodes to the participant’s scalp, and would then instruct the subject to carry out some small, simple motor activity, such as pressing a button, or flexing a finger or wrist, within a certain time frame. No limits were placed on the number of times the subject could perform the action within this period. During the experiment, the subject would be asked to note the position of the dot on the oscilloscope timer when "he/she was first aware of the wish or urge to act". Pressing the button also recorded the position of the dot on the oscillator, this time electronically. By comparing the marked time of the button's pushing and the subject's conscious decision to act, researchers were able to calculate the total time of the trial from the subject's initial volition through to the resultant action. On average, approximately two hundred milliseconds elapsed between the first appearance of conscious will to press the button and the act of

[238] A pre-motor potential, the readiness potential is considered to be a manifestation of the contribution of the neocortex to the pre-motor planning of volitional movement.
[239] See: http://en.wikipedia.org/wiki/Benjamin_Libet

pressing it. The scientists also analyzed EEG recordings for each trial with respect to the timing of the action. It was noted that brain activity involved in the initiation of the action, primarily centered in the secondary motor cortex, occurred, on average, approximately five hundred milliseconds *before* the trial ended with the pushing of the button. That is to say, researchers recorded mounting brain activity related to the resultant action as many as three hundred milliseconds *before* subjects reported the first awareness of conscious will to act. In other words, apparently conscious decisions to act were *preceded* by an unconscious buildup of electrical charge within the brain, the readiness potential[240]. Thus, some interpreters have come to think that, as the RP precedes the conscious experience of the intention to act, it seems as if the decision to act has *already been made* by the time the subject becomes conscious of this decision. It was concluded that the decision to act is made by unconscious neuronal processes, i.e. by the brain, rather than the conscious subject. Consequently, a host of philosophers and cognitive scientists have come to regard the notion of *free will* as a mere illusion[241], starting a fierce debate within academic circles as well as among members of the interested public[242]. Libet himself, on the other hand, proposed a less radical interpretation, suggesting that "free will" has only *veto power*. Thus while human being don't posses the ability to create and voluntarily change the world around them, they nevertheless are able to stop certain things from happening.

Crucial to the interpretation of these experiments was the "volitionist" conception of willing and human action subscribed to by Libet (and other cognitive scientists). According to this philosophical doctrine, which was also held by David Hume, William James, and Bertrand Russell, for example, a movement is caused by the mental phenomenon of willing. Willing was construed as a mental occurrence (i.e. the occurrence of an idea or image of the desired movement), commonly referred to as a *volition*. This conception figured prominently in the experimental design of Libet's experiments in the form of William James's ideo-motor theory of voluntary action. William James notably conceived of willing not as something one does but as something that occurs to one. According to James, willing is entertaining a representation or an idea, so called willed intentions, in the mind. Thus, an idea of a

[240] See: Libet 1999, 2004
[241] See e.g.: Wegner (2004)
[242] See: Gayer (2004)

certain action (e.g. raising my arm) is able to instigate one's motor centre. James deemed such ideas of movement, i.e. willed intentions or volitions, as causally responsible for voluntary action. Wiliam James synthesized these ideas in his ideo-motor theory of voluntary action:

> " In this vain, ideo - motor action is defined as the sequence of movement upon the mere thought of it (*i.e., upon the formation of willed intentions*), as the process of volition. Wherever movement follows unhesitatingly and immediately the notion of it in the mind, we have ideo-motor action. We are then aware of nothing between the conception and the execution."[243]

The Jamesian conception of voluntary action is embedded within the classical Cartesian/empiricist representationalist framework of mind, a trademark of which is the conception of representations as causal agents. Libet's thinking, in ascribing *phenomenal qualities* to voluntary action, is firmly rooted in this tradition, although the representationalist framework his studies are interpreted in is, of course, the cognitive view.

The discussion which ensued in the wake of Libet`s studies, soon spread from neuroscientific and philosophical forums over to the wider academic community and the public. Numerous news magazines and newspapers have published articles and interviews with authorities in neuroscience, philosophy and the law, in the attempt to provide an accurate account of the nature and implications of these findings and the scholarly discussions surrounding them. The suggestion that our experience of free will and free agents might merely be an illusion has (rather unsurprisingly) delivered a decisive blow to our self-conception as human beings. As remarked in chapter I, the questions which cognitive science pursues are indeed intimately tied to the very heart of our human self-conception, as the answers it strives to provide seem to hold the promise of a deeper understanding of human nature and rational agency.

The various critical voices featuring in this discussion have quite rightly (though it would seem with somewhat limited success) highlighted the numerous philosophical flaws inherent in the conception of voluntary action endorsed in these

[243] James (1983), p.1130

experiments. The most well-known (and powerful) of which is the *infinite regress* inherent in the conception of voluntary action[244]. In addition, various flaws in the design of Libet's experiment have been highlighted, the most prominent of which is the criticism that Libet actually failed to measure the variable he set out to measure (i.e. the experience of intending to act). Rather than measuring the conscious experience of the intention to act, Libet has by some been criticized for actually measuring the consciousness of the consciousness of the experience of intending to act[245].

However, I suggest that the debates surrounding the conception and nature of voluntary action or other design inherent flaws of Libet's studies touch only upon part of the problem. In fact, to a certain degree the discussion surrounding the interpretation of Libet's experiments and the underlying conception of willed action may be regarded as a mere side issue, although an important one. Yet, the real fuel which keeps the fire of this debate burning, is not so much comprised by the question as to how to interpret Libet's data, but by the flawed and utterly misconceived representationalist framework, the cognitive view, which accommodates Libet's misconceived experimental design and conception of willed action in the first place. This aspect has hitherto gone unnoticed. Viewed from this angle, the debate that ensued in the wake of Libet's studies is merely an inevitable extension of having adopted a flawed framework for the explanation of mental phenomena. Without endorsing a representationalist conception of the mind, there would be no room for a volitionist conception of willing and no room for the (mis-)interpretation of Libet's results as evidence for the illusory nature of our experience of free will. It is the prevalence of the cognitive view as the explanatory framework of cognitive science,

[244] If every act is preceded by a "volition" we may ask, what causes this volition. Not only bodily (like lifting a finger in a PET scanner), but also mental operations and activities (like performing mental arithmetic, reciting a poem in the mind or reflecting on a problem) may be voluntary These are on a par with willed intentions themselves. Thus, if the answer is no, they are not voluntary, then how can the acts be voluntary? If the answer is that they are voluntary, then the theory held by James, and Libet requests that these operation and acts must themselves be instigated by the formation of prior volitions, and those from other volitions and so on *ad infinitum*. (See e.g.: Ryle 1990, Kenny 1989)

[245] Even if one overlooks the infinite regress argument, the variable "Experience of intending to act" is not really the variable that Libet actually wanted to measure, because what is measured is not the conscious experience of the intention to act, but the consciousness of the consciousness of the experience of intending to act. Although, that may seem overly petulant, this fact is never the less of crucial importance given the minute time scale experimenters are dealing with in these studies. This error may be called the *"Cogito fallacy"*: The failure to notice, that the experience of enjoying a conscious experience presupposes that one already entertains a conscious experience!

which has given rise to a misguided debate, which does nothing but contribute to a distorted view of human agency and human nature. Given the fact that our conceptions of the nature of human agency and human nature influence central concepts of criminal law such as responsibility and guilt, it is not only important to point out misguided philosophical conceptions of action and flawed experimental design in Libet's studies[246] but also the flaws inherent in the explanatory framework within which these issues are discussed. Only by tackling these issues at their very root, can we ensure that our understanding of the brain advances, and avoid the misinformation of the public by continuing the cultivation of a deplorable *neuromythology*. The present investigation and deconstruction of the cognitive view and the subsequent suggestion of the Aristotelian alternative constitutes a significant step in this direction[247][248].

[246] And in follow up studies conducted in the wake of Libet's initial experiments, which may have avoided some crucial errors of design, but still endorse a flawed concept of willed action as well as the cognitive view.

[247] Aristotle characterized the mind in terms of the powers not only of the intellect but also in terms of the powers of the will. The mind is not only a cognitive capacity, but also a volitional „capacity". Thus, the human will can be understood *the ability to act for reasons*. Its freedom derives from the special characteristics of practical reasoning, which are tied to our conceptual structure.

[248] Viewed from the suggested Aristotelian perspective human beings are animals with a distinctive range of abilities (i.e. the abilities that are distinctive of human beings are abilities of intellect and will). Though they have a mind, they are not identical with the mind (or the brain) they have. Though they have a body, they are not identical with the body they have. Nor is a human being a conjunction of a mind and a body that causally interact with each other. Like other animals, human beings have a brain on the normal functioning of which their powers depend. But a human person is not a brain enclosed in a skull. A mature human being is a self-conscious agent, with the ability to act, and to react in thought, feeling and deed, for reasons

These capacities and their exercise give to human beings the status of persons. While *human being* is a biological category, *person* is a moral, legal and social one. To be a person is, among other things, to be a subject of moral rights and duties. It is to be not only an agent, like other animals, but also a moral agent, standing in reciprocal moral relations to others, with a capacity to know and to do good and evil. Since moral agents can act for reasons, and can justify their actions by reference to their reasons, they are also answerable for their deeds. To be a human being is to be a creature whose nature it is to acquire such capacities in the course of normal maturation in a community of like-natured beings. *(From personal communications with Peter Hacker)*

Bibliography

Ammereller E, Fischer E (2004) Wittgenstein at Work: Method in the Philosophical Investigations. *London, Routledge.*

Arrington RL, Glock HJ (1991) Wittgenstein's Philosophical Investigations – Text and Context. *London, Routledge.*

Arrington RL, Glock HJ (1996) Wittgenstein and Quine. *London, Routledge.*

Atkins P (2005) The Limitless Power of Science. In, *Nature's Imagination. Cornwell J (ed.), Oxford, oxford University Press*

Baars BJ (1997) In the Theatre of Consciousness – The Workspace of the Mind. *Oxford, Oxford University Press.*

Baker GP (2003) The Voices of Wittgenstein. *London, Routledge*

Bechtel W, Graham G (1999) A Companion to Cognitive Science. *Oxford*, *Blackwell Publishers Ltd.*

Bennett MR, Hacker PMS (2001) Perception and memory in neuroscience: a conceptual analysis. *Progress Neurobiology 65(6):499-543.*

Bennett MR, Hacker PMS (2002) The motor system in neuroscience: a history and analysis is of conceptual developments. *Progress Neurobiology 67(1):1-52.*

Bennett MR, Hacker PMS (2003) Philosophical Foundations of Neuroscience. *Oxford, Blackwell Publishers*

Bennett MR, Hacker PMS (2005) Emotion and cortical-subcortical function: conceptual developments. *Progress Neurobiology 75(1):29-52.*

Baker GP, Hacker PMS (1985) Rules, Grammar and Necessity, *Vol.2 of an Analytical Commentary on the Philosophical Investigations, Oxford, Blackwell Publishers Ltd..*

Baker GP, Hacker PMS (2005) Wittgenstein – Understanding and Meaning, Vol. I – Essays; (2nd revised edition), *Oxford, Blackwell Publishers Ltd.*

Block N, Flanagan O, Güzeldere G (1997) The Nature of Consciousness Philosophical Debates. *Cambridge (Mass.), MIT Press1997.*

Boltzmann L (1979) Theoretical Physics and Philosophical Problems. (McGuiness B, ed.) *London, D Reide, Publishing Company 1979.*

Brentano F (1995) [1874] Psychology from an Empirical Standpoint, *London, Routledge & Kegan Paul*

Chomsky N (1966) Cartesian Linguistics, *Cambridge Mass., MIT Press*

Chomsky N (1995) Language and Nature. *Mind, 104:413ff.*

Churchland PS (1986) Neurophilosophy. *Cambridge (Mass.), MIT Press*

Churchland PM (1992) A Neurocomputational Perspective. The Nature of Mind and the Structure of Science. *Cambridge (Mass.), MIT Press*

Churchland PS (2002) Brainwise. Studies in Neurophilosophy. *Cambridge (Mass.), MIT Press*

Cottingham J (1984) The Philosophical Writings of Descartes. *Cambridge, Cambridge University Press 1984, Volumes I-III*

de Charms RC, Zador A (2000) Neural Representation and the Cortical Code. *Annual Review of Neuroscience 23:613-647, p.613*

Descartes R (1984) [1649] The Passions of the Soul. In, Cottingham J, `The Philosophical Writings of Descartes´, *Cambridge, Cambridge University Press 1984, Volumes I-III*

Dennett D (1991) Consciousness Explained, *Boston: Little, Brown and London*

Dilman I (1999) Free Will – A Historical Introduction. *London, Routledge*

Edelman G, Tononi G (2000) The Universe of Consciousness: How Matter becomes Imagination), *New York, Basic Books*

Einstein A (1944) A. Einstein to R.A. Thornton, unpublished letter dated Dec. 7th,1944 (EA 6-574). *Einstein Archive, Hebrew Univesity Jerusalem.* In, Albert Einstein as a Philosopher of Science. DA Howard, *Physics Today, Dec. 2005, p.34f.*

Evans G (1980) Varieties of Reference. *Oxford*, *Oxford University Press*

Fodor JA (1975) The Language of Thought. *Cambridge (Mass.) Harvard University Press*

Fodor JA (1987) Psychosemantics – The Problem of Meaning in the Philosophy of Mind. *Cambridge (Mass.), MIT Press*

Fodor JA (1991) A Theory of Content and Other Essays. *Cambridge (Mass.), MIT Press*

Frege G (1964) [1893] The Basic Laws of Arithmetics. *Berkley (Cal.), Berkley University Press*

Frisby JP (1979) Seeing, Illusion, Brain and Mind. *Oxford, Oxford University Press*

Gazzaniga MS, Ivry R, Mangun GR (2002) Cognitive Neuroscience: The Biology of the Mind. *New York, W.W. Norton, 2nd Edition*

Geach P (1957) Mental Acts. *London, Routledge*

Geyer Ch (2004) Hirnforschung und Willensfreiheit. *Frankfurt a.M., Suhrkamp*

Glock HJ (1997) A Wittgenstein Dictionary. *Oxford, Blackwell Publishers*

Glock HJ (2001) Wittgenstein – A Critical Reader. *Oxford, Blackwell Publishers.*

Hacker PMS (1991) Appearance and Reality. *Oxford, Blackwell Publishers*

Hacker PMS (1991b) Seeing, Representing and Describing. In: Hyman J, (ed.), *Investigating Psychology – Sciences of the Mind after Wittgenstein´ p.122, London, Routledge*

Hacker PMS (1996a) Wittgenstein's Place in 20th Century Analytic Philosophy. *Oxford, Blackwell Publishers*

Hacker PMS (1996b) Mind and Will. (Part I Essays, Part II Exegesis), *Oxford Blackwell Publishers Ltd.*

Hacker PMS (1997a) Insight and Illusion. *Bristol, Thoemmes Press*

Hacker PMS (1997b) Wittgenstein - Meaning and Mind. (*Part I Essays), Vol.3 of an Analytical Commentary on the Philosophical Investigations, Oxford, Blackwell Publishers*

Hacker PMS (1997c) Wittgenstein - Meaning and Mind. (*Part II Exegesis), Vol.3 of an Analytical Commentary on the Philosophical Investigations, Oxford, Blackwell Publishers*

Hacker PMS (2000a) Wittgenstein – Mind and Will. (*Part I Essays, Part II Exegesis),Vol.4 of an Analytical Commentary on the Philosophical Investigations, Oxford, Blackwell Publishers*

Hacker PMS (2000b) Wittgenstein – Mind and Will. (*Part II Exegesis), .4 of an Analytical Commentary on the Philosophical Investigations, Oxford, Blackwell Publishers*

Hacker PMS (2001) Wittgenstein: Connections and Controversies. *Oxford, Clarendon Press*

Harre R, Gillett G (1994) The Discursive Mind. *London, Sage Publications*

Hertz H (1956) [1899] Principles of Mechanics. *London, Dover Publications*

Hobbes T (1951) [1659] Leviathan. *London, Methuen*

Horder TJ (2001) The organizer concept and modern embryology: Anglo-American perspectives. *International Journal of Developmental Biology, 45(1):97-132.*

Hyman J (1991), Investigating Psychology – Sciences of the Mind after Wittgenstein. *London, Routledge*

James W (1983) [1890] The Principles of Psychology. *Cambridge (Mass.), Harvard University Press*

Johnson-Laird PN (1987) How Could Consciousness arise from the Computations of the Brain ? In, *Blakemore C, Greenfield S (eds.), Mindwaves. Blackwell, Oxford*

Johnson-Laird PN (1988) The Computer and the Mind. *Fontana, London.*

Kant I (1996) [1781] Critique of Pure Reason: Unified Edition. *Pluhar WS (Translator), Cambridge, Hackett Publishing.*

Kenny A (1973) Wittgenstein. *London, Penguin Press*

Kenny A (1984) The Legacy of Wittgenstein. *Oxford, Blackwell Publishers*

Kenny A (1989) The Metaphysics of Mind. *Oxford, Blackwell Publishers*

Kröber HL (2004) Die Hirnforschung bleibt hinter dem Begriff strafrechtlicher Verantwortlichkeit zurück. In, *Geyer (2004) Hirnforschung und Willensfreiheit. Frankfurt a.M., Suhrkamp*

Lachman R, Lachman J, Butterfield E (1979) Cognitive Psychology and Information Processing, *Hillsdale New Jersey, Erlbaum*

Libet B, Freeman A, Sutherland K (1999) The Volitional Brain – Towards a Neuroscience of Free Will. *Exeter, Imprint Academic*

Libet B (2004) Haben wir einen freien Willen. In, *Geyer (2004) Hirnforschung und Willensfreiheit. Frankfurt a.M., Suhrkamp*

Lüderssen K (2004) Ändert die Hirnforschung das Strafrecht. In, *Geyer (2004) Hirnforschung und Willensfreiheit. Frankfurt a.M., Suhrkamp*

Phillips CG, Zeki S, Barlow HB (1984) Localization of Function in the Cerebral Cortex, *Brain 107:338-347* p.345

Young JZ (1978) Programs of the Brain. *Oxford, Oxford University Press*

La Metrrie JO (1960) [1749] L'Homme Machine. *Princeton (NJ), Princeton University Press*

Locke John (1961) [1690] An Essay Concerning Human Understanding. *London, Dent&Sons*

Malcolm N (1995) Wittgensteinian Themes: Essay 1978-1989, *Ithaca, NY, Cornell University Press*

Malcolm N (1984) Ludwig Wittgenstein: A Memoir. With a Biographical Sketch by G.H. von Wright and Wittgenstein's letters to Malcolm. *Second edition. London: Oxford University Press.*

Marr D (1982) Vision. *San Francisco, WH Freeman*.

Marcel AJ, Bisiach E (1988) Consciousness in Contemporary Science. *Clarendon Press Oxford 1988*

McGinn C (1989) Mental Content, *Oxford, Blackwell Publishers*

Metzinger T (1995) Conscious Experience. *Mainz, Schöningh*

Moore GE (1962) Some Main Problems of Philosophy. *London, Collier*

Moore GE (1970b) Wittgenstein´s Lectures in 1930-33. In, Philosophical Papers´, *George Allen & Unwin Ltd.*

Moore GE (1970) Philosophical Papers. *London, George Allen & Unwin Ltd.*

Newell A (1980) Physical Symbol Systems. *Cognitive Science 4:135-183*

Newell A, Rosenbloom PS, Laird JE (1989) Symbolic Architectures for Cognition. In, *Posener MI (ed.), Foundations of Cognitive Science; Cambridge (Mass.), MIT Press*

Papineau D (1987) Reality and Representation, *Oxford, Blackwell Publishers*

Pauen M, Roth G (2001) Neurowissenschaft und Philosophie. *Stuttgart, MBI*

Porter R, Teich M (1991) The Renaissance in National Context. *Cambridge, Cambridge University Press*

Preston J (1997) Thought and Language. *Cambridge, Cambridge University Press*

Putnam H (1967) Psychological Predicates. In, *Art, Mind and Religion, Capitan WH and Merrill DD (eds.), Pittsburgh, Penn., University of Pittsburgh Press*

Rhees R (1968) Notes for Lectures. *Philosophical Review 77*

Roth G (2004) Wir sind determiniert. Die Hirnforschung befreit von Illusionen. In, *Geyer (2004) Hirnforschung und Willensfreiheit. Frankfurt a.M., Suhrkamp*

Rowling K (2000) Harry Potter and the Goblet of Fire. *New York, Arthur Levine Books*

Rundle B (1997) Mind in Action. *Oxford, Oxford University Press*

Ryle G (1971a) Critical Essays - Collected Papers, Vol.I. *London, Hutchinson&Co*

Ryle G (1971b) Collected Essays - Collected Papers, Vol.II. *London, Hutchinson&Co*

Ryle G (1979) On Thinking. ***Blackwell Publishers Oxford***

Ryle G (1990) [1951] The Concept of Mind. *London. Penguin Books*

Scruton R (1981) A Short History of Modern Philosophy. *London, ARK Paperbacks*

Seager W (1999) Theories of Consciousness – An Introduction and Assessment´, *London, Routledge.*

Sen A (2001) Development as Freedom. *Oxford, Oxford University Press*

Singer W (2002) Auf dem Weg nach Innen – 50 Jahre Hirnforschung in der Max-Planck-Gesellschaft. In., *Singer W ‚Der Beobachter im Gehirn – Essays zur Hirnforschung'., Frankfurt am Main, Suhrkamp*

Snow CP (1993) The Two Cultures. *Cambridge, Cambridge University Press*

Sterelny K (1990) The Representational Theory of Mind – An Introduction: *Oxford, Blackwell Publishers Ltd.*

Stich S (1992) What is a theory of mental representation?. *Mind, 101, pp. 243-61*

Stillings NA, Wesiler SE, Chase CH, Feinstein MH, Garfield JL, Rissland EL (1995) Cognitive Science – An Introduction. *Cambridge (Mass.), MIT Press*

Strawson PF (1985) Causation and Explanation. *In, Vermazen B & Hintikka MB (eds.), Essays on Davidson, Oxford, Oxford University Press*

Strawson PF (1991) Analysis and Metaphysics – An Introduction to Philosophy. *Oxford, Oxford University Press*

Strawson PF (1999) Individuals*. London, Routledge*

Stufflebaum RS (1999) Representation and Computation. *In: Bechtel W & Graham G, A Companion to Cognitive Science, Blackwell Pubslishers Ltd., Oxford*

Watson RA (1995) Representational Ideas – From Plato to Patricia Churchland. *Dordrecht, Netherlands : Kluwer Academic Publishers*

Weiskrantz L (1999) Neurospychology and the Nature of Consciousness. In, *Blakemore C, Greenfield S (eds.), Mindwaves. Blackwell, Oxford*

von Wright GH (1963) Norm and Action – A Logical Enquiry. *London, Routledge*

Vossenkuhl W (1982) Anatomie des Sprachgebrauchs. *Stuttgart Klett-Cotta*

Vossenkuhl W (1995) Ludwig Wittgenstein. *München, Verlag CH Beck*

Wegner D (2004) The Illusion of Conscious Will. *Cambridge (Mass.), MIT Press*

v. Wright GH (1984) A Biographical Sketch´. In: Malcolm, `Ludwig Wittgenstein – A Memoir´, *Second edition. London: Oxford University Press (1984)*

White A (1989) Methods of Metaphysics, *London, Routledge*

Wolff RP (1973) The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysics of Morals. *New York, Harper & Row*