Mind and Language: Evolution in Contemporary Theories of Cognition

Tanya de Villiers

Dissertation presented for the degree of Doctor of Philosophy at Stellenbosch University

Supervisor: Professor F. P. Cilliers

December 2006

ACKNOWLEDGEMENTS

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the National Research Foundation.

The financial assistance of the Harry Crossley Foundation and the Ernst and Ethel Eriksen Trust respectively is hereby gratefully acknowledged.

DECLARATION

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

_____ _____

T. de Villiers Date

ABSTRAK

Hierdie tesis bied ’n historiese oorsig van sommige van die ooreenstemmende elemente van "philosophy of mind" en taalfilosofie in die twintigste eeu deur veral aandag te gee aan die rol wat kognisietorieë in beide velde speel. Die onderlinge verband tussen die teorieë van Peirce, Chomsky, Derrida en Deacon word bespreek en daar word verder ’n oorsig gegee van die sienings oor "mind" in beide filosofie sowel as in kognitiewe wetenskap. ’n Argument word gevoer dat baie van die klaarblyk onoorbrugbare kwessies wat teorieë van “mind” en kognisie vandag dikwels kelder, teruggespeur kan word na die metafisiese vorm waarin die betrokke filosofiese vrae dikwels gegiet is. Ten spyte van die weerstand wat daar tans vanuit vele filosofiese geledere kom, word die argument gevoer dat taal en "mind" in die lig van ’n neo-Darwinistiese, evolusionêre teorie gesien moet word.

ABSTRACT

This thesis gives an historical overview of some of the isseues connecting philosophy of mind and philosophy of langauge in the twentieth century, especially with regard to the relevance of both disciplines to theories of cognition. Specifically, the interrelation bewteen the theories of Peirce,Chomsky, Derrida, and Deacon are discussed. Furthermore, an overview of twentieth century views on mind in both philosophy and the cognitive sciences is given. The argument is made that many of the apparently insurmountable issues that plague theories on mind and cognition today can be traced back to the metaphysical mould into which the philosophical questions at issue here were cast. Also, despite current resistance to the idea from many philosophical quarters, a case is made for aaproaching language and mind in terms of neo-Darwinist evolutionary theory.

Aan JB

TABLE OF CONTENTS

INTRODUCTION

Philosophy, like all other studies, aims primarily at knowledge. The knowledge it aims at is the kind of knowledge which gives unity and system to the body of the sciences, and the kind which results from a critical examination of the grounds of our conviction, prejudices., and beliefs (Russell 1962:154).

Philosophy, though unable to tell us with certainty what is the true answer to the doubts which it raises, is able to suggest many possibilities which enlarge our thoughts and free them from the tyranny of custom (Russell 1962:157).

Philosophically speaking it would be no great exaggeration to say that the twentieth century was the century of language and mind. Both topics dominated the philosophical scene to a greater or a lesser extent in the analytic and the continental traditions. With the advent of psychology as an independent discipline it was thought that the study of "the mind" as the quintessentially psychological entity, mind-like phenomena would disappear from the philosophical body of work [cf. for example Russell 1962 (1911):155]. Instead, philosophy of mind became a distinctive subject in the late nineteenth and early twentieth century. Ironically, philosophy of mind developed mainly as the result of the founding of scientific psychology. Further contributing factors were the publication of Franz Brentano's 1874 *Psychology from an Empirical Standpoint*, and the publication in 1890 of William James's *Principles of Psychology*. Brentano resurrected Aristotle's and Aquinas' notions of *intentionality*, which held that paradigm-case mental states such as beliefs, desires, hopes, etc. have *intentional objects*, in other words, they are of, or about something (Flanagan 1991:571). For him, intentionality differentiated the mental from the non-mental, and psychology as the study of the mental, would need to give an explanation for it.[1] On the other hand, James compiled a compendium of psychological knowledge, gleaned from the new science of psychology and

[1] Flanagan (1995:571) notes that Freud studied under Brentano when he was at medical school and argues that it may be fruitful to think of psychoanalysis as involving an extension of Brentano's basic insight.

from traditional philosophical thinking about mind, and found that there was a deep tension between a scientific theory of mind and traditional philosophical ways of thinking about the subject. James, inspired by Darwin's theory of evolution by means of natural selection, believed that a naturalistic position with regard to mind should be taken. These beginnings would burgeon and create the foundation for philosophical and psychological speculation for a century to come. While Husserl (a student of Brentano) established phenomenology on the continent, philosophy of mind came to be entwined with the science of psychology in America and England, and was to develop into an independent discipline over the next fifty odd years.

Philosophy of mind established itself as a professional field within philosophy between 1900 and 1950, as a result of the radical different methodologies and hypotheses that developed in the wake of the newly established "science of mind" (Flanagan 1995:571). For a significant part of the century psychology and philosophy would be virtually indistinguishable. Thus, philosophy of mind was concerned with the methodological foundations and epistemic status of psychological subject matter on the one hand, and with "metaphysics of mind" on the other. Hence, Dewey, for example, aimed to establish the mind as a naturalistic entity that should be approached as an adaptive, and fundamentally worldly phenomenon, rather than as a Cartesian incorporeal and virtually mystical substance. On the other hand, Carnap was concerned with the epistemological status of first- and second person psychological reports (572). The aim was to establish psychology on par with the natural sciences, which were thought to be based upon the verifiable observation reports made by independent and impartial scientists. Hence, the psychological observations made by subjective subjects about their own minds needed somehow to be made amenable to a similar objective verifiability. These two approaches were amalgamated and subjected to extensive criticism in Ryle's *The Concept of Mind* [1960 (1949)]. Flanagan (1995:572-573) considers *The Concept of Mind* to be the founding document of contemporary philosophy of mind proper. Ryle criticised the conception of mind as an incorporeal entity as needlessly mystifying, and proposes that "mind" be considered nothing more than certain behavioural dispositions of the human organism. Furthermore, he rejects the view that the first person observer has privileged access to his or her own mental states. Recent trends in philosophy of mind include the analysis of "mental" entities such as sensation, perception, free will, and intentionality. There is also a debate on whether reasons can be causes, whether private language is possible, and whether it is possible to know one's own and other minds. Secondly, there are attempts to develop materialistic alternatives to Cartesian dualism, primarily identity theory, eliminative materialism and functionalism.

After 1950 a new approach to mind was established, stemming from the work done by logicians in the 1930's on formalising thoughts and the application of their ideas in the subsequently developed "thinking machines" or computers. This development led to a computational conception of mind, where the mind was considered as something akin to the mechanical symbol processors that were being developed. Various related disciplines merged into what is today known as cognitive science, which includes disciplines like neurology,

computer science, and psychology. The possibility arose that the mind would become known through scientific experiment in the laboratory, rather than through the *a priori* analyses of the philosophers.

With regard to language, it is virtually impossible to separate theories on mind (whatever form they take) from some sort of theory on language. As one of our primary means to access the mind (the other being introspection), language is inevitably the medium in terms of which mind is studied, and thus invariably becomes part of such theories. Frege set the scene for some of the themes that would dominate the philosophy of language in the twentieth century. He argued that there is nothing "psychologistic" about meaning, but that meaning is the result of the way that the term is employed (Blackburn 1995:456-459). Hence, a systematic theory of meaning would categorise the expressions of a given language and systematically explain how the truth-conditions of sentences are derived from the components of those sentences. Modern analytic philosophy would build on this assumption, and as a result, to quote Blackburn, "[i]n the twentieth century the political and other practical dimensions of meaning have frequently been regarded as a slightly disreputable secondary element, outside the pure theory of representation." In a similar manner, the body would be regarded as a slightly disreputable secondary element to mind. Given the entwined character of theories on mind and language, it can be argued that these two assumptions reinforced and fed off one another. As we shall see, questioning the assumptions of one tradition often leads to the disruption of the other.

Philosophy of language in the English-speaking world was dominated by an "analytic" conception of language, which found expression in logical empiricism and ordinary language philosophy, among others. Both approaches considered the primary role of philosophy to be the clarification of the language used in philosophical speculation to prevent the misuse of natural languages through the uncritical use of language (Ricoeur 1979:244). In this understanding, the problems of philosophy were thought to be primarily linguistic problems (Glendinning 2004:5). Logical empiricism aimed to accomplish this task through constructing artificial languages that do not lend themselves to misuse and sought to exclude metaphysical statements from language use. Ordinary language philosophy sought to identify models that govern the proper use of language within the confines of natural language (Ricoeur 1979:244).

Ricoeur (1979:249-250) notes that as the appeal of ordinary language philosophy seemed to wane in Great Britain, a new form of it began to take root in the United States in the wake of Noam Chomsky's formulation of transformational generative grammar. Chomsky's project held the promise of solving longstanding philosophical problems within the framework of the general linguistics which resulted from Chomsky's analysis of the structure of language. From the publication of *Syntactic Structures* (1957) up until quite recently, the Chomskyan paradigm (in various forms) would dominate the study of language. Chomsky held that the study of language was also the study of a speaker's knowledge of language as it is manifested in the mind of that speaker. Hence, he sees linguistics as a branch of cognitive

psychology in as far as it studies the structures in the mind that are responsible for language competence.

On the continent, the legacy of Husserlian phenomenology made itself felt in the emphasis that was given to the search for the non-linguistic conditions for language (Ricoeur 1979:250). Husserl identified the "paradox of language", namely that language cannot be said to be primary to our apprehension of the world, nor can it be said to be autonomous, and yet any such apprehension is necessarily expressed in terms of language. With regard to the philosophy of language, phenomenology attempted to relate language to "modes of apprehending reality which come to expression in discourse" (250). The focus was on that which allows discourse to refer to things in the world. As such, there was a significant commonality between phenomenology and the linguistics of deep structures and grammar, as formulated by Chomsky.

There were also those continental philosophical schools that attempted to redefine reality in terms of language, such as philosophical structuralism and hermeneutics (Ricoeur 1979:260ff). Philosophical structuralism stemmed from structural linguistics, and extrapolated it to reality. In linguistics Ferdinand de Saussure [1983 (1916)] argued that i) language consists of a system of differences, rather than possessing an "essence" or "substance"; ii) the structure and conventions that govern language are distributed among a community of speakers, rather than originating from individual language users, iii) the fundamental entity of language is the sign, which is arbitrary and consists of the difference between a signifier and a signified. Saussure held that language is a closed system, which need not refer to anything beyond the signifier and the signified. Philosophical structuralism takes this argument as a foundation and formulates a theory of reality in terms of it. It is important to emphasise that the concept of structure has multi- and cross-disciplinary origins (Johnson 1993:2). Hence, linguistic structuralism is just one instance of a theoretical engagement with structuralism.

The movement toward a semiological understanding of the world was not limited to philosophy of language and linguistics. Michel Serres identified a change in paradigm in the philosophy of science dating from the late nineteenth century and gained wide-spread influence across academic disciplines by the mid-twentieth century (Johnson 1993:2-3). Serres calls this new paradigm "neo-Leibnizian, because of its primary methodology which rests on the concept of the code and basic operations of combination and permutation, resulting in the emergence of the new sciences of information theory, molecular biology and cybernetics (3). Hence, whereas philosophy of science in the nineteenth century was dominated by the concept of energy and the three laws of thermodynamics (Prigogine and Strengers 1984), in the twentieth century it is dominated by the concept of information. As Serres (quoted in Johnson 1993:3) sums it up:

> The sciences of today are formalistic, analytical, grammatical, semiological, each of them based on an alphabet of elements… their affinities are so apparent that we are once again beginning to dream of the possibility of a *mathesis universalis*.

One of the first philosophers to take such an approach was C. S. Peirce, who attempted to formulate a semiological (science of signs) model of reality (Ricoeur 1979:261). Peirce viewed the world as a system of codified signs, which can be analysed in terms of a semiological approach. Structure is central to such semiological approaches, where, as with language, the identity or meaning of entities is established in terms of the codes or relationships that govern the structures of which they form part. Freud, for example, conceived of the subconscious in such terms. The structural-semiological school of thought conceives of man and society as products of language, rather than as subjects that constitute language. In Ricoeur's formulation: "the system imposes categories that transform the subject/object dualism which imbues Descartes, Kant and Husserl alike" (263). However, whereas Peirce was motivated by the theory of evolution and incorporated it into his work, structuralism generally embraced an anti-historical stance, and was quite sceptical about the accessibility and reliability of data pertaining to the origin and historical development of phenomena. As with the body as it pertains to mind and the political and practical dimensions of language use as it pertains to the formal structures of language, the historical evolutionary histories of systems were disregarded in favour of the study of the synchronic structural patterns of such systems.

The structuralists viewed signs as being constituted by an internal characteristic of the sign and the sign system. As with analytic philosophy, the "universe of discourse was elevated to the universe of the absolute" (Ricoeur 1979:264). However, for structuralism, language is a world onto itself, not necessarily correlated to objects in the world. Jacques Derrida rang in post-structuralism with the development of a grammatology, or a "science of writing", which supplements and extends the structuralist programme. In this approach, reference is entirely replaced with differentiation within the system. Meaning now derives entirely from the structure of the system of language, leaving no room for referring to objects outside of language, in a world where everything becomes text. It can be argued that Derrida thoroughly merges the semiological and structural projects so that language always precedes, mediates, and structures our knowledge of the world.

In contrast to the structuralist approach, another prominent school of philosophy of language on the continent, hermeneutics, concerned itself with our conception of reality as mediated by language, but also stressed that which went beyond the structure of language itself, such as the intentions of the author of a text for example. In the hermeneutic tradition, phenomenology was considered to be hermeneutic – existence is interpreted as a text would be. The philosophical challenge is in identifying the way in which existence could be accurately interpreted.

What is striking in this brief overview of the major developments in theories of mind and language in the course of the twentieth century is that, despite early efforts by theorists such as Dewey and Peirce, evolutionary theory did not feature very strongly in either discipline. It is only in the latter part of the century that evolutionary theory gained

respectability with regard to theories of mind and theories of language, most noticeably after the publication of Pinker and Bloom's 1990 "Natural Language and Natural Selection". Subsequently there has been an explosion of literature dedicated to the evolutionary origins of mind or language or both.

Yet, philosophies of mind and cognitive science seem to be holding out against these developments, some with indifference and some with outright hostility. Very few, if any, compendiums of the central problems of philosophy of mind make any mention of evolution, even in sections dedicated to contemporary debates on materialism (see, for example, Chalmers 2002a). Theorists like Fodor (e.g. 1998; 2000; 2005) Chomsky (e.g. 1972; 1975; 1994; 2000a; 2000b; 2004), and even Gould[2] (e.g. 2000) are openly and vocally hostile toward the twin ideas of an evolved mind and language. Given the success that evolutionary theory enjoyed in explicating other systemic and biological phenomena, the question arises: Why is there such resistance to an evolutionary approach to the study of language and mind? Mind, as a biological entity seems especially amenable to the evolutionary approach. However, this thesis will show that such an approach goes against the grain of many of the presuppositions and methodologies that have driven philosophies of mind and language over the past century and more. Of the many reasons that could conceivably be construed to contribute to this hostility, the most pertinent one seems to be a residual dualism, however much the research programmes of such critics are touted as "materialistic".

An evolutionary approach to mind would inevitably entail that one treats mind as – to paraphrase Dennett (1995:66) – an effect of certain physical processes, rather than as a First Cause. This was an implication too extraordinary for Descartes' otherwise materialist sensibilities, and it remains a difficult position to accept for those thinkers who wish to view the human mind as a "special" phenomenon in an otherwise mechanistic universe. Of course, mind as a product of evolution need not imply that it is not an extraordinary (and perhaps unique) phenomenon. It does, however, seem to remove the need for mystical, even magical metaphysical explanations of mind. Furthermore, the mind becomes amenable to scientific study, hence undermining any claims of exclusive and unique subject matter that purely "mind" orientated disciplines may have appropriated. This thesis will adopt the view that it is the pervasive and insidious influence of a (common-sensical?) dualistic thinking that primarily informs resistance to an evolutionary/adaptationist approach to mind.

This work will attempt to illustrate some of the advantages that are to be had in adopting an evolutionary perspective on both language and mind. Hence, some of the major assumptions that have shaped thought on mind and language and their interrelation over the course of the twentieth century will be discussed. As we shall see, although the theorists and theoretical developments under discussion represent some of the seminal developments in theories of mind and/or language over the course of the last century, they all invariably ran

[2] Dennett (1995:300) speculates that Gould's incessant and somewhat haphazard salvo's against contemporary Darwinism belies his need to get "personalities – consciousness, intelligence, agency – back in the driver's seat" and to "give the mind some elbow room, so it can *act*, and be *responsible* for its own destiny, instead of being the mere effect of a mindless cascade of mechanical processes".

into philosophical dead ends – philosophical dead-ends that can be addressed by amending them with evolutionary theory, and particularly natural selection. In order to illustrate the argument the co-evolutionary theory of mind and language by Terrence Deacon (1997a) will be examined in detail. The aim of this thesis is not to endorse Deacon's particular co-evolutionary theory (although much of his theory has great merit), but to use it as an illustration of how seemingly intractable theoretical problems can be overcome through the adoption of an evolutionary perspective. As we shall see, these apparently intractable theoretical problems arise out of certain metaphysical loyalties, whether they are a Judeo-Christian-influenced commitment to a "Mind-first"[3] view of the universe (in the case of Peirce – Chapter 1), a commitment to a long-standing research-programme and elegant, "optimal" solutions to design problems (Chomsky – Chapter 2), a "pure" post-structuralist phenomenological methodology (Derrida – Chapter 3), viewing higher-brain functions as abstract computational processes (cognitive science in the form of AI, and functionalism), or an apparent animosity towards Darwinism that seems to be rooted in the string tradition of dualism that underpins and informs philosophy of mind (e.g. Fodor, Gould, Searle).

To state the case more strongly, the position that will be taken here is that in many respects a paradigm shift is taking place in those theories that intersect in the realm of the cognitive, namely theories on the interrelation between mind, consciousness, and language. The position that will be taken here is that approaching these phenomena as products of evolution by means of natural selection, and particularly as co-evolved products of evolution, offers a plausible way out of the apparent philosophical myopia that has characterised theories on mind and language for too long. It will become clear that mind and language are essentially complex phenomena that need to be approached interdisciplinarily, with all of the relevant scientific and philosophical resources at our disposal, if they are ever to be explained adequately. Our key resource, at this junction, is that of the theory of evolution by means of natural selection and its merger with contemporary genetics (what is often referred to as neo-Darwinism). However, as will become clear in the course of the work, all of the positions discussed here have had an important contribution to make to the final, evolution-based attempt at accounting for the very existence and current structure of language and mind.

Although it may seem unlikely at first glance, Deacon (1997) produces a synthesis of some of the main tenets of the positions under discussion. "Unlikely", perhaps, but less unlikely if one were to consider the way in which they are already interrelated – a claim that should become apparent during the course of this work. In many ways the different approaches to language, mind, and their interrelation that will be discussed here are not as radically different as they are often made out to be. Nor are they incommensurable. They all spring from a tradition of mind-body dualism, share the notion that there is something "special" about the human mind, and the hunch that this something special can at least in part be attributed to the anomalous nature of human language. It is in trying to address these

[3] See Dennett's (1995:17-34) succinct and lucid analysis of what such a "Mind-first" view entails.

enormously complex and difficult inklings that they diverge, often hindered by the metaphysical assumptions that have gone before them. It is here that important philosophical work is to be done; work that is not about finding answers as much as it is about establishing what questions can, legitimately, be asked. Hence, rather than assuring ourselves that certain possibilities can be summarily disregarded on the grounds of the certainties within a given theoretical tradition and the authorities that have gone before us, perhaps we will do well to remind ourselves of the origins of tradition and certainties. Russell (1962 [1912]:161) sums this position up as follows:

> Philosophy is to be studied, not for the sake of any definite answers to its questions, since no definite answers can, as a rule, be known to be true, but rather for the sake of the questions themselves; because these questions enlarge our conception of what is possible, enrich our intellectual imagination and diminish the dogmatic assurance which closes the mind against speculation…

With this in mind, this work is in part a historical overview of seminal positions in theories of cognition with an eye on the genealogy of some of the certainties in contemporary theories of mind, cognition, and language. Hopefully this exercise will lead to a broadening of our intellectual imagination, and open the possibility that the mind, language, and cognition are best studied as enormously complex products of evolution by means of natural selection.

The discussion will proceed as follows:

Chapter 1 will focus on Peirce's philosophy, primarily because he set the trend for, or anticipated, many of the developments in philosophy in the twentieth century. As we shall see Peirce's philosophy was an early attempt at incorporating developments in evolutionary theory and in science into philosophical questions, including questions regarding language and mind. Peirce's semiotic theory also anticipates the practise in structuralism and post-structuralism of understanding the word as being constituted by language. Although Peirce was relatively unknown in his own lifetime, he has come to exert quite an extensive influence on contemporary philosophy. As will be discussed, Chomsky and Derrida have Peirce as a common precursor and both attribute some of their most significant insights to his work. Furthermore, Deacon's appropriation of Peirce's semiotics in order to account for the link between language and mind/cognition will be examined. It will become clear that Peirce's contribution to philosophy is still far from exhausted. It will also be argued, however, that Peirce's preoccupation with religion and his desire to allow for "free will" leads him to unnecessary forays into metaphysical speculation.

Chapter 2 pays extensive attention to Chomsky's linguistic theory – which is a reaction to structuralism – given his profound and far-reaching influence on both linguistics and various theories of mind. It will be seen that Chomsky can be considered to be one of the

single greatest influences on contemporary assumptions about the relation between language and the mind, where it manifests. Chomsky's identification of "Plato's problem" and his subsequent conclusion that humans have an innate Universal Grammar are rightly considered to be breakthroughs in terms of conceptions of the relation between language and mind. However, it will also be argued that Chomsky's overly theoretical approach to his subject matter and his skepticism towards the capacity of science in general and evolution in particular leads to a theoretical dead-end. Consequently, Chomsky has repeatedly reformulated his conception of Universal Grammar, without establishing grounds in terms of which his new formulations could be considered to be more accurate than the preceding ones.

Chapter 3 will study the work of Jacques Derrida, as an alternative reaction to structuralism from Chomsky's reaction. Whereas Chomsky focuses on the structure of grammar and on the relation between language and mind, Derrida focuses on the structure of meaning and on the relation between language and the world (and epistemology). We will see that Derrida also appropriates Peircean theory, although to a different aspect than does Chomsky. Derrida's theory can be argued to be fundamentally situated within the structural and phenomenological traditions, in that he focuses on our epistemological apprehension of the world rather than on any structures to do with the human mind, whether hypothetical or empirical. Derrida extends Saussure's insights with regard to language in order to explain how meaning arises within the structure of language. Deacon uses this understanding to explain how symbols come to acquire meaning within a system of symbols.

Chapter 4 considers Darwin's formulation of evolutionary theory by means of natural selection and the delayed impact that it had on theories of mind and language. One of the major reasons for this delay, apart from philosophical prejudice, is the enthusiasm that existed for computational approaches to mind for the better part of the century. These approaches seemed to resonate with the dominant philosophical trends much more than evolutionary theory did. We will examine how the computational approach to mind runs into certain conceptual barriers, which resulted in evolutionary approaches gaining greater acceptance. Evolutionary approaches to language and mind are not homogenous by any means and some of the disputes that have arisen will be considered.

Finally, *Chapter 5* explores Terrence Deacon's theory of the co-evolution of language and mind. He makes use of some aspects of all of the theories that are discussed in the rest of the study, and thus indicates how all of these approaches provide significant insights into the questions under consideration. Whereas Deacon's thesis is not the final word on the subject by a long shot, it indicates the benefits and plausibility of a co-evolutionary, interdisciplinary approach to language and mind.

CHAPTER 1

PEIRCE SETS THE SCENE

In a variety of ingenious ways, Wittgenstein shows what Peirce had said, viz., that our language and consequently our thought is embedded in a highly complex context of habits conventions, and conduct which are social in character (although they may be internalised)
Bernstein (1965:83).

Peirce, almost alone of philosopher-scientists of his age was able to break away from the restricting influences of the Newtonian world-picture
Gallie (1966:234).

1. INTRODUCTION

In many ways Charles Sanders Peirce (1839-1914) heralded several of the main themes that would dominate philosophy in the twentieth century, especially with regard to his semiology. In terms of this study, he will serve as an example of an early attempt at integrating language, mind, science, and evolutionary theory (although not a purely Darwinian evolutionary theory). Understandably, this mammoth task took the better part of Peirce's life and in many respects his work seems speculative and incomplete. As we shall see, despite his innovative approach to select philosophical problems, there were some primary assumptions that Peirce could not overcome and it will be argued that his work suffered for it. It will also be seen that most of the other theorists that are discussed (namely Chomsky in chapter 2, Derrida in chapter 3, and Deacon in chapter 5) were significantly influenced by Peirce's ideas.

Peirce was a philosopher, logician, mathematician, and scientist, and the founder of the doctrine of Pragmatism. As a qualified and practising scientist (he worked for the United States Coast Survey for thirty years)[1], he brought a decidedly scientific and empirical bent to

[1] See Weiss (1965) for a concise bibliography focussing on the important contributions that Peirce made to many of his extensive fields of interest.

philosophy.[2] Although his philosophical work was relatively obscure in his lifetime his original contribution to philosophy has come to be widely acknowledged, even if Hanson (1965:43) declares: "we are still not ready for much of Peirce's philosophy – no more than were our grandparents." Peirce vigorously opposed mechanical philosophy and believed in the reality of absolute chance and the principle of continuity in the universe. He also attempted to account for the origin of laws, which led him to propose a solution for the mind-body problem and to develop his theory of an evolutionary universe (Weiss 1965:7). His work was insightful and original, and in many ways anticipated the concerns of twentieth century philosophy.[3] It is Peirce's semiotic theory that is most pertinent for purposes of our argument[4], but as we shall see, it is impossible to separate Peirce's theory on signs from his ontology, or ultimately from his metaphysics.

In many respects, Peirce could be described as a materialist. However, he was also a logician, with a keen interest in the purely formal procedures of mathematics and deductive logic. His extensive body of work seems to oscillate between these two preoccupations. As we shall see, his later work tends toward the transcendental and the metaphysical. Ultimately, Peirce attempts to develop an encompassing cosmology from his extensive philosophical writings and goes beyond empirical experience, which he held to be the basis of our knowledge.

2. Where Descartes Went Wrong

Peirce was greatly concerned with the first principles on which philosophical arguments are based. His position was that arguments and hypotheses always rest on assumptions that cannot be confirmed. With his background in the natural sciences, Peirce was keenly aware of the constraints that the physical world places upon our experience. He was of the opinion that facts gained from such experience should form the basis of our philosophical endeavours, and that philosophers should be aware of the developments in the natural sciences. As such, he was a harsh critic of the Cartesian project and its implications.

A central tenet of Peirce's philosophy is his concern with the "real" world and its effect on our knowledge. In his work, the world is always "forcing itself" on us, in that it interrupts any possibility of purely formal, disembodied thought. No matter our intellectual preoccupations, there are some aspects of the world that we are compelled to take into account and react to – the danger posed by a predator, for example, or the compelling force of hunger. The world "out there" is a constant, disruptive element that is always already imposing constraints on any epistemology. In this regard, Peirce seems to anticipate the existentialist philosophy of the twentieth century, which reintroduced the body and bodily experience as essential philosophical subject matter.

---

[2] According to Bernstein (1965:72), Peirce is representative of many late-nineteenth century thinkers who reacted against the "intellectualistic" character of Western philosophy.

[3] Cf. Bernstein (1965:66); Goudge (1950:2); Thayer (1996:2); Herbenick (1970:84).

[4] According to Gallie (1966:125), Peirce was the first philosopher to recognise the importance of fact that a sign can only function as an element within a system of signs, and hence only gains its meaning in relation to other signs.

Peirce formulates what he calls the principle of *abduction* to explain the ability of organisms to pay attention to features of the world that are relevant to them, as opposed to a myriad of unimportant phenomena that could be taken into account.[5] Peirce remarks that early scientists (and all scientists) seem to have formulated the "laws of nature" with remarkable ease. His observation leads him to conclude that (Peirce 1965 Vol. V: 106-107 [5.172]):

> However man may have acquired his faculty of divining the ways of Nature, it has certainly not been by a self-controlled critical logic…It appears to me that the clearest statement we can make of the logical situation – the freest from all questionable admixture – is to say that man has a certain Insight, not strong enough to be oftener right than wrong, but strong enough not to be overwhelmingly more often wrong than right, into the Thirdness, the general elements of Nature…This Faculty is at the same time of the general nature of Instinct, resembling the instincts of the animals in its so far surpassing the general powers of our reason and for its directing us as if we were in possession of facts that are entirely beyond the reach of our senses.

He argues that our innate limitations, which are due to our particular cognitive structure, contributes to our everyday theory-construction in that it limits the amount of possible hypotheses that can be formulated with regard to a particular problem or situation. An example of this ability would be the tendency of children to look to where someone is pointing, as opposed to looking at the person's finger as a cat might do. Children are constituted in such a way that they expect people to behave in certain ways, such as pointing to things that they are referring to.

Hence, Peirce adds a further logical category, over and above induction and deduction, namely *abduction*. He defines abduction as: "the process of forming an explanatory hypotheses" (1965 Vol. V:106 [5.172]). *Abduction* characterises the human ability to form a general prediction on the basis of information available to them (1965 Vol. II:153 [2.296]). Peirce argues (1965 Vol. II:47 ff [2.86]) that humans have the remarkable ability to make remarkably accurate observations and guesses when confronted with new or unknown phenomena. This "divinatory power of guessing right" is the result of people being inclined toward certain interpretations rather than others (53-53 [2.96]). He believes that humans and animals inherit ideas through natural selection that adapt them to their environment. Included among these ideas are: notions of force, matter, space, and time, as well as notions of what to expect from their fellow beings.[6] "Our innate mechanical ideas were so nearly correct that

---

[5] Peirce (1965 Vol. V:121 [5.196]) argues that the question of pragmatism is nothing but "the question of the logic of abduction."

[6] As we shall see in Chapter 4, Pinker makes the same claims in terms of his theory of the mind as a product of evolution by means of natural selection.

they needed but slight correction" (1965 Vol. II:476 [2.753]). Abductions are the only origins of new ideas. To quote Peirce (477 [2.754]):

> Side by side, then, with the well established proposition that all knowledge is based on experience, and that science is only advanced by the experimental verification of theories, we have to place this other equally important truth, that all human knowledge, up to the highest flights of science, is but the development of our inborn animal instincts.

Abduction is the only logical operation that introduces new ideas into the logical realm (as opposed to deduction and induction, which only work with existing ideas). Induction only comes into play as a secondary process used to correct hypotheses and based on experiential input, while deduction merely infers the necessary consequences of a hypothesis.

In keeping with his background in the physical sciences, Peirce also rejects the idealist view that meaning corresponds to a “mental image” in the mind. Pierce comprehensively criticises the Cartesian project, which causes Gallie to enthuse that his papers on the subject "foreshadow the most important developments in the theory of knowledge which have been made in the present century" (1966:62).

The western philosophical tradition has long accepted that there are two sorts of knowledge: the immediate and the inferential (direct and indirect knowledge). Inferred knowledge is knowledge that is derived from "direct" knowledge, intuited by the mind. Descartes adopts this traditional view with his theory of intuition. However, Peirce rejects the theory of intuition, and argues that whenever we know something, we know it *as* something – primarily as something standing in relation to other things. And relating or classifying objects in this manner cannot be done without symbols or signs of some sort.[7] Due to abduction Peirce holds (in contrast to Descartes) that our knowledge, even knowledge that is apparently direct and intuitive, in fact takes on the form of a hypothesis. He contends that all our knowledge claims rely on an assumption – the assumption that the object of that knowledge claim can be classified or systematised in a certain way, in other words, that the object is a certain type of thing. The truth or falsity of that assumption is something that is consequently tested through experience. If our underlying assumption is correct, the object of our knowledge should behave as we expect it to, if it is incorrect, it won’t. The relevant necessary consequences of our assumption cannot be realised in a single act of intuition, and is hence incommensurate with the Cartesian doctrine of intuition (Gallie 1966:62-69) (Peirce 1965 Vol V: 135-147 [5.213-5.237]).

This process of inference does not regress *ad infinitum*, however. Peirce holds that all knowledge relies on the ability to manipulate signs and symbols, and it is in the course of learning how to manipulate symbols that the “first assumptions” on which subsequent

[7] See Gallie (1966:67-69) for a succinct summary of Peirce’s position.

assumptions are based, are instilled. When learning how to manipulate symbols, one learns a certain way of classifying and systematising, which is then applied to novel situations. It is this iterability – to use a Derridean term – of the system of classifying and relating that one learns when learning how to use symbols, that allows for the intelligibility of others' thoughts. As Gallie (1966:71) puts it: "Once this is admitted, the great error of Descartes and of all later Cartesians becomes plain: it is the assumption that we cannot know… We must build on what we have; and we quite obviously build better in some cases than we do in others."

Such first assumptions are not static and need not be permanent, but they are a point of departure for the process of acquiring knowledge, which always remains open to questioning and testing. For Peirce there will always be some aspect of our claims to knowledge that is based on untested assumption. We cannot eliminate the prejudices inherent to our system of thought, because "they are things which it does not occur to us *can* be questioned" (Peirce 1965 Vol. V: 156 [5.265]). In light of this Peirce believes that philosophy cannot rely on the apparent "intuitions" of the individual Cartesian thinker, but that it should adopt methodology of the natural sciences and "proceed only from tangible premises which can be subjected to careful scrutiny, and to trust rather to the multitude and variety of its arguments than to the conclusiveness of any one" (156 [5.265]). Such "careful scrutiny" cannot be undertaken by the individual thinker, but is undertaken by a community of thinkers. According to Peirce, our knowledge of our own minds, as well as those of other people, is derived from outward physical events, which communicate our own and other people's behaviour – in other words, from *signs* (Peirce 1965 Vol. V:177-185 [5.294-5.309]). In this understanding, making or reacting to signs is to be "engaged in being a mind" (Gallie 1966:81). We only know what we are thinking insofar as we are able to produce signs that present the conclusion of our thoughts, and insofar as we are able to defend that conclusion through the use of signs, communicating both with ourselves and with others (Peirce 1965 Vol. II:14-17 [2.27]). According to Peirce, "all thinking is dialogic in form" (Peirce 1965 VI: 233 [6.338]). And the dialogue takes place in the form of the words (signs) that we assemble from a community of speakers.

Peirce is not a nominalist, however, and aims to develop a logical doctrine that allows us to distinguish between subjective and objective aspects of the world.[8] As such, it would allow for the existence of an objective world, independent of thought[9], and for the possibility of an object truth that is accessible to us (even if only theoretically). His logical doctrine thus has metaphysical implications.[10] Peirce is not a materialist either.[11] He allows for the possibility

---

[8] See: 1.15–1.26 (Peirce 1965 Vol. I:3-8). As Deledalle (1992:295) puts it, Peirce aimed to replace the Aristotelian logic of predicates with a logic of relations.

[9] Peirce does not equate consciousness and thought. He *does* identify communication and thought as having the same form (Ransdell 1976:01-102).

[10] See: 1.16 (Peirce 1965 Vol. I:4). Also see Parker (1994:51-59) for a comprehensive discussion of the relationship between Peirce's logical realism and metaphysical idealism.

[11] Peirce overtly rejects materialism, which he calls "quite as repugnant to scientific logic as to common sense"; instead he opts for what he calls "objective idealism". See 3.527-3.552 (Peirce 1965 Vol. V:333-

that our access to "reality" is mediated by our perception and interpretation and that our descriptions of laws or logical forms might only be human constructs (Parker 1994:53). He also holds that it would be possible, at least in principle, for us to distinguish between human constructs and objective truths – given enough time and enough people involved in an enquiry.[12] Thus, Peirce's logic attempts to incorporate the nominalist insight that knowledge is a social construct, but he attempts to temper his position with a dose of realism, where truth exists independently of our constructs and serves to constrain our constructions of knowledge, at the very least.[13]

It is in the context of these assumptions that we should approach Peirce's theory of knowledge. Peirce does not comprehensively develop and present his theory of knowledge in any one text. His thoughts on the subject are scattered throughout his extensive writings. Different contexts and different dates result in ideas and assertions that seem contradictory at times, and extensive commentary is devoted to proving the unity or disunity of these different works. What follows will be an attempt to amalgamate and explore the basic tenets of Peirce's theory on knowledge, gathered from his various explorations of the subject.

3. Towards an Epistemology

In his account of knowledge, Peirce makes use of his three categories of being Firstness, Secondness, and Thirdness (Peirce 1965 Vol. I:182-183 [1.355]):[14]

> …I make so much use of threefold divisions in my speculations, that it seems best to commence by making a slight preliminary study of the conceptions upon which all such divisions must rest. I mean no more than the ideas of first, second, third – ideas so broad that they be looked upon rather as moods or as tones of thought, than as definite notions, but which have great significance for all that. Viewed as numerals, to be applied to what objects we like, they are indeed skeletons of thought, if not mere words…

---

345); 6.24 [1965 Vol. VI:20]. Peirce rejects such traditional positions as idealism, materialism, and dualism.

He formulates his own position, synechism, as a combination of idealism and materialism (e.g. 5.509 [1965 Vol. V: 359]; 1.172 [1965 Vol. V:70-71]; also see 5.4 [1965 Vol. V:2-3]

[12] It might not be possible for an individual enquirer to establish *truth*, but knowledge gained from a communal process of enquiry would ensure knowledge that is a closer approximation of the truth (Peirce 1994: 54-56).

[13] See 4.1 (1965 Vol. IV: 3-5).

[14] Peirce used this categorical scheme in many, sometimes even inconsistent, ways. However, throughout his extensive writings he maintains that these categories embody elements in all experience, reality, and being. Furthermore, these categories are both irreducible and sufficient for giving a comprehensive and coherent account of experience, reality, and being. Although Peirce believed that he could prove the necessity and irreducibility of his categories of being, it is not clear that he manages to do so (Bernstein 1965:70).

In terms of cognition, we become aware of things because we recognise their *quale*[15], which can be understood as the distinctive characteristic or quality of a thing that is perceived.[16] Quale fall under the category of Firstness; the realm of possibility not yet actualised. Furthermore, quale account for the *possibility* of experience; they make it possible for consciousness to note "something is there" (Radford 2004:4).[17] When we actually "perceive" an object, we enter the realm of Secondness. Peirce describes Secondness as "occurrence", where we "knock up" against the existence of something and cannot but acknowledge it (1965 Vol. I:184 [1.358]). "It is the compulsion, the absolute constraint upon us to think otherwise than we have been thinking that constitutes experience" (Pierce 1965 Vol. I:170 [1.336]).[18] Some things are forced upon our cognition by the world, by its reaction against us, as it were; the presence of reality, which is not part of our thought. With regard to cognition, Radford refers to Secondness as "the indexical moment of consciousness" (Radford 2004:5). When we become aware of something, it becomes an object of knowledge. Peirce argues that objects (facts) have to be linked to make up knowledge, which leads us to the next level of being, Thirdness. Thirdness requires the use of symbols, in terms of which we can make links between and inductions from the facts at our disposal.[19] We make use of signs to order facts into knowledge and to create new ideas (Peirce 1965 Vol. 1:203 [1.383]):[20]

> …the highest kind of synthesis is what the mind is compelled to make neither by the inward attractions of the feelings or representations themselves, nor by a transcendental force of necessity, but in the interest of intelligibility that is, in the interest of the synthesising "I think" itself; and this it does by introducing an idea not contained in the data, which gives connections which they otherwise would not have had.

Thirdness requires Secondness, and Secondness requires Firstness, but Thirdness cannot be reduced to Secondness or Thirdness (nor can Secondness be reduced to

---

[15] Peirce is often thought to be the first theorist to use the term "quale" (or qualia) (e.g. Livingston 2002:19); a term and concept which would endure to become a contentious issue within the domain of philosophy of mind, and which would invariably be understood as denoting something unique to consciousness. In many ways the "problems" of qualia and consciousness would not move much beyond the Peircean mould for more than a century to come.

[16] "Each *quale* is what it is for itself, without reference to any other" (Peirce 1965 Vol. VI: 150 [6.224]).

[17] Whereas qualities are traditionally thought of as the "whatness" or the basic epistemological unit of something (and hence the primary object of knowledge), Peirce means something precognitive, something that is felt. The knowledge of our awareness of something's qualities is not to be confused with the awareness itself, or the direct experience that we have of qualities (Bernstein 1965:72). Knowledge necessarily involves Thirdness.

[18] See also 2.138 [1965 Vol. II:73] and 5.503 [1965 Vol. V:352-353].

[19] As we shall see in Chapter 5, Deacon (1997a) makes use of Peirce's analysis in order to formulate a theory of cognition in general, and to account for the human mind in particular.

[20] As Smith (1965:97) puts it, Thirdness "involves something more than the ability to react with other things."

Firstness). There is an indeterminacy, or future-orientated aspect to Thirdness, because it embodies generality, or the law-like character of being. As will become clear, Peirce argues that everything in the universe is governed by laws, and these laws can only be understood in terms of Thirdness.

Peirce believes that all living things – things predisposed towards forming habits – and living organisms that are conscious (thus that posses some form of nervous system) in particular, are out of equilibrium with the objects and processes that surround them (Gallie 1966:221). They are still in the process of forming habits and adjusting to their environments. Peirce's conception of this process of adjustment is akin to evolution, however, he conceives of this evolutionary process as being subject to "pure mental processes" that take the form of sign behaviour. In other words, sign behaviour supervenes on organic behaviour and hence causes the adaptation of such organic processes to take place in a reasonable manner (Gallie 1966:221). This aspect of Peirce's theory will be examined in greater detail in a later section. It is important, however, to orientate oneself with regard to the position that Peirce's theory of signs (semiotic) takes up with regard to his overall project. As we shall see, for Peirce, semiotics is central to his ontology and as such it cannot be fully understood in isolation from the rest of his work.

Peirce's theory of signs allows him to overcome the limits of perception in that the possession of, and the ability to manipulate, symbols allows us to make connections between "raw" facts, drawn from immediate perception. Peirce believes that we would not be able to generalise without signs, which in effect means that we would not be able to think, since Peirce equates thought with the ability to generalise (1965 Vol. I:229 [1.420]). Furthermore, Peirce holds that "every thought is a sign" (Perice 1965 Vol1:284 [1.538]; Vol. V:189 [5.314]; Vol. V:323 [5.470]).

As we have seen, Peirce rejects the idealist view that meaning corresponds to a "mental image" in the mind. He emphasises the role that experience plays in the formation of meaning and essentially holds that the meaning of a word – or any other symbol – is formed within a series of experiences, and that meaning develops within a given context. Peirce describes his pragmatic approach to language as follows (quoted in Gallie 1966:14):

> A better rule for avoiding the deceits of language is this: Do things perform the same purpose practically? Then let them be signified by the same word. Do they not? Then let them be distinguished. If I have learnt a formula in gibberish which in any way jogs my memory so as to enable me in each single case to act as though I had a general idea, what possible utility is there in distinguishing between such a gibberish formula and an idea? Why use the term a general idea in such a sense as to separate things which, for all experiential purposes, are the same?

As such, words or symbols gain their meaning from them being experienced and used in a succession of experiences (real or imagined) that serve to constrain our thinking (Peirce 1965 Vol. I:160 [1.321]; 1965 Vol. 5:258-259 [5.402n]). Peirce stresses the general character of language and believes that it “guides and controls the greater part of our thinking” (Gallie 1966:29). In fact, on more than one occasion Peirce equates man with being able to think and manipulate signs.[21] In his understanding, language has to express those aspects of our experience that are general in order to be communicable in the first place. He uses the model of a conversation between two parties, as the primary example of communication to illustrate his conception of language as communication.

Peirce does not completely discard the concept of inference in his epistemology. Inference is an essential feature of the “cognitive mind”, in the sense that all our knowledge is based on intelligent conjecture; to think means to make inferences from certain assumptions that are testable or defendable (Gallie 1966:99).[22] According to Peirce we form habits of inference in our thinking, the legitimacy of which relies on whether or not they usually lead to true conclusions. The test for such derived conclusions relies on the extent to which they correspond to the actual state of affairs. If one were to infer that a moving bus poses no apparent threat, the experience of being hit by a moving bus would expose the falsity of that inference. One’s conception of reality is usually derived from, and tempered by, a community. If one’s inferences more or less correspond to those of a form of community, there is a greater chance of those inferences being true. To stress again, this is not an argument for relativism or nominalism; Peirce believes that there is nothing "to prevent us knowing outward things as they really are" (1965 Vol. V:187-189 [5.311-5.316]). Peirce believes that, given enough time and a large enough community of thinkers, the ideal state of complete information can in principle be reached. Furthermore, it is impossible for an individual to achieve knowledge outside of language and community.

As already mentioned, Peirce repeatedly states that thought consists of signs or that we think only in signs (e.g. 1965 Vol. II:169 [2.302]).[23] It follows that the way in which we think corresponds to the way in which we manipulate signs – if we think correctly (or manipulate signs correctly), we will be able to obtain "true" facts (1965 Vol. V:391 [5.554]). For Peirce, gaining knowledge takes place through a variety of activities which comprise the manipulation of signs. Thus, man is a sign-using animal and it is through signs that human beings gain knowledge, and hence the possibility of "the truth".[24] Given that Peirce defines thought as the

---

[21] In 5.314 (1965:189), for example, Peirce states that "my language is the sum total of myself; for the man is the thought".

[22] A particular inference is the result of a habit of mind. Such a habit can be formulated in a proposition, or a guiding principle. A rational person is not at the mercy of entrenched habits, and can exhibit self-control through critically examining propositions based on them (Bernstein 1965:81-82).

[23] This principle also underlies the work of Jacques Derrida and is echoed in his well-known statement: "There is nothing outside the text", and will be discussed in Chapter 3.

[24] Peirce never explains how this situation comes about, which Gallie (1966:92) attributes to the fact that he lacked the historical imagination necessary for such a project, “like almost every other philosopher of

process where signs develop in accordance with the laws of inference (169-170 [5.283-5.284]), he develops a semiotic (theory of signs) in an attempt to model the mechanisms behind thinking.

4. PEIRCE'S SEMIOTIC[25]

Gallie describes Peirce's semiotic as exhibiting the "first movements of a broad shift in philosophical interests and methods", the repercussions of which is still felt today (1966:111). With his theory of signs, Peirce hoped to develop a conceptual scheme of signs, based on reality rather than any conventional understanding of sign phenomena. His further aim was to develop an ontology based on semiotics. Parker (1994:51) describes Peirce as embracing both logical realism *and* metaphysical idealism. Although these are traditionally taken to be mutually exclusive positions, Peirce aims to derive both positions from his theory on the action of signs.

Peirce was primarily a logician, and as such, he developed a theory on signs and their interrelations that he believed to be equivalent to logic (Melrose 1995:496; Greenlee 1976; Liszka 1990:30).[26] Furthermore, he does not consider logical principles as being regulative principles, but as "truths of being" (Peirce 1965 Vol. 1:259 [1.487]). Hence, he defines semiosis as (Peirce 1965 Vol. II:134 [2.227]):

> …the quasi-necessary, or formal, doctrine of signs. By describing the doctrine as "quasi-necessary", or formal, I mean that we observe the characters of such signs as we know, and from such an observation, by a process which I will not object to naming Abstraction, we lead to statements, eminently fallible, and therefore in one sense by no means necessary, as to what must be the characters of all signs used by a "scientific intelligence", that is to say, by an intelligence capable or learning by experience.

Peirce's conception of signs is extremely broad and includes everything from simple animal communication (and in some cases he extends the ability to communicate to inanimate objects as well) to the more conventional symbols to which the term is usually applied (1956 Vol. II:156-169 [2.274-2.301]).[27] He defines the concept *sign* as follows (1965 Vol. II:135 [2.228]):

---

the western tradition". The following chapters will focus on various attempts to explain how this situation does come about, notably those by Chomsky (chapter 2) and Deacon (chapter 5).

[25] We will replace Peirce's own spelling "semeiotic" with the more contemporary "semiotic".

[26] Peirce does not use the term logic with consistency, but broadly speaking he uses this term to refer to all reasoning (Bernstein 1965:85).

[27] Gallie attributes this state of affairs to the fact that Peirce seems to be attributing properties to all signs, which should only be applicable to those signs used by a scientific intelligence. Peirce fails to indicate the qualifications that need to be made before his analyses are applied to "more elementary forms of sign activity" (1966:112). As we shall see, not all theorists agree with this assessment.

> …something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object*. Its stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen [sign].[28]

As we have seen, Peirce attempts to establish a valid epistemology on the grounds of his theory of signs. As can be deduced from his definition of semiosis, he sees signs as something capable of acting upon, or being interpreted by any "intelligence capable of learning by experience". Hence, his aim is, as Short puts it, to establish a single science of the conventional structures of human speech and natural signs (1981:199).[29]

One of the fundamental aspects of Peirce's work is his rejection of the dualistic conception of the world (1965 Vol. I:26 [1.61]):

> One of the worst effects of the influence of moral and religious reasonings upon science lies in this, that the distinctions upon which both insist as fundamental are dualistic distinctions, and that their tendency is toward ignoring of all distinctions that are not dual and especially of the concept of continuity.

Accordingly, for Peirce, signification is not based on a relation between two phenomena, the sign, and its object – an assumption that characterises the traditional approach within (western) philosophy – but always the result of a *triadic* relation. He believes all sign phenomena to have this general and distinctive property (e.g. 1965 Vol. II:135 [2.229]; 1965 Vol. II:136 [2.230]).

4.1 OBJECT-SIGN-INTERPRETANT

Peirce holds that every sign[30] has an object and an interpretant.[31] A sign is always something that stands for something *to somebody* (1965 Vol. II:135 [2.228]); it is not a sign by virtue of some sort of intrinsic characteristic, it can only function as a sign when it is

---

[28] Peirce sometimes refers to signs in the strict sense (i.e. cognitive signs or logons) as *signs*, and to signs in a more general sense as *representamen* (Colapietro 1987:215). See also Short (1981:198).

[29] The distinction between "conventional" human signs and "natural" signs is problematic and in some instances difficult to uphold, as will be discussed at a later stage.

[30] Peirce does not use "sign" consistently throughout his body of work, sometimes using representemen when referring to signs in general, and sign when referring to "Representanem with a mental Interpretant" (1965 Vol. II:156 [2.274]). This usage is later dropped.

[31] See, for example, Peirce (1965 Vol. III: 210-211 [3.360-3.361]).

interpreted. Hence, the irreducibly triadic character of all sign phenomena, structured as a three-way relation sign-object-interpretant.[32] The sign is a signifier; the *object* is that which the sign refers to, while the *interpretant* is the process that enables one to infer the reference from a sign and its context, whatever that process may be. Hence, the fundamental triad in Peirce's semiotic is that of object-sign-interpretant.[33] As we shall see, the trichotomy *icon-index-symbol* classifies the relation in which the sign can stand to its object.

As already stated, Peirce is by no means a nominalist, even though he places great emphasis on the role that humans play in constructing knowledge. The same holds for the emphasis that he places on the interpretive aspect of sign action. In Peirce's theory, sign interpretation is not and cannot be an arbitrary and solipsistic process. He holds that we always have some access to the way the world is "in itself", in that eventually we "come up against" the world as one of the correlates in the object-sign-interpretant epistemological trichotomy. This aspect of the trichotomy cannot be discounted any more than can interpretation.

The necessary interrelation between a sign, its object, and its interpretant is fundamental to Peirce's semiosis (1965 Vol. II:134-155 [2.227-2.272]; Vol. V:332 [5.484]). For Peirce, the process of signification is one where a sign "creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign" (1965 Vol. II:135 [2.228]).[34] Any process that determines reference is an interpretant, and the kind of interpretive response determines the kind of reference relationship that is invoked. Hence, an essential feature of the interpretant is that it places the sign in the context of other signs, through a process of interpretation. In other words, the sign represents its object to the interpretant, which gains information about that object from its context in a system of signs – thus performing a kind of "translation".[35] Liszka (1994:27) describes the sign as undergoing a kind of evolution when it represents an object, and Peirce states that (1965 Vol. V:173 [5.289]). :

> …no present actual thought (which is a mere feeling) has any meaning, any intellectual value; for this lies not in what is actually thought, but in what this

---

[32] Gallie (1966:117) notes that Peirce became convinced that the sign-relation is involved in all other irreducibly triadic relations, and uses the act of giving as an example.

[33] A further dimension that can be added to this triad is the ground, or the respect in which the object is represented by the sign (Colapietro 1987:208).

[34] Peirce (1965 Vol. V:326-332 [5.475-5.483]) goes on to distinguish between three types of interpretants: the emotional interpretant (the feeling evoked by the sign), the energetic interpretant (a physical action or mental effort), and finally a logical interpretant (a concept). Different interpretive processes may be present simultaneously.

[35] Liszka (1994:33) defines translation as "essentially the substitution of one sign for another, through which the latter develops the former".

> thought may be connected with in representation by subsequent thoughts; so that the meaning of a thought is altogether something virtual.

And (379 [5.569]):

> A sign is only a sign *in actu* by virtue of its receiving an interpretation, that is, by virtue of its determining another sign of the same object. This is as true of mental objects as it is of external signs. To say that a proposition is true is to say that every interpretation of it is true. Two propositions are equivalent when either might have been an interpretant of the other. This equivalence, like others, is by an act of abstraction (in the sense in which forming an abstract noun is abstraction) conceived as identity.

This is *not* to say that misinterpretation is impossible, that all interpretations are equivalent, or that false propositions do not exist (379 [5.569]):

> …a false proposition is a proposition of which some interpretant represents that, on an occasion which it indicates, a precept will have a certain character, while the immediate perceptual judgement on that occasion is that the precept has not that character. A true proposition is a proposition belief in which would never lead to such disappointment so long as the propositions not understood otherwise than it was intended.

The triadic nature of semiosis is irreducible, and something cannot be a sign unless it is mediated by means of an interpretant, which relates it to an object. In other words, meaning comes about within the object-sign-interpretant interrelation. Furthermore, this process of translation is continuous, where a sign stands in relation to both previous and future translations, forming a system of signs. Hence, reference and meaning are irreducibly subject to the context of the interpretant.

It is important to emphasise that interpretants should not be equated with rules of interpretation.[36] Short (1996:511) believes this misunderstanding to be common and "a serious error". He speculates that the misunderstanding arises because Peirce's semiotic is often conflated with Saussure's (1983 [1916]) "semiology" and he goes on to point out important differences between the two theories, and it is worth discussing them in brief:

Whereas Sausurre's sign is a dyad (consisting of the signifier and the signified), Peirce's sign is defined triadically. Furthermore, whereas Sausurre's signified is always a concept, Peirce's objects of signification "include all that may in any way be thought of, spoken of, defined, described, judged to be, judged not to be, postulated, imagined, pictured, depicted, mapped, outlined, diagrammed, felt, pointed to, inferred, or, among other

[36] See for example Liszka (1994:34).

possibilities, shown". In other words, the objects of signification may be anything (e.g. physical things, imaginary things, impossible things, times, thoughts). Furthermore, actions and feelings can be interpretants and can occur within the same semiotic structures that "give thoughts their intentionality" (Short 1996:516).[37]

A further difference is that, in Saussure's theory, the signifier/signified relationship always relies on convention.[38] Peirce's interpretant is not a rule and hence not merely a matter of convention. He includes "natural signs" and non-human interpreters into his semiotic, which he sees as incorporating a "natural" or "necessary" dimension of the interpretation of a sign into his theory. Furthermore, Short puts the explanatory power of Peirce's semiotic down, in part, to the possibility to link interpretation by rule to interpretation on grounds such as compulsion and resemblance: "One thing we can learn from Peirce is that conventional language works only in conjunction with natural signs, including our instinctive tendency to point and to look where others point" (Short 1996:512).[39] It is important to reiterate that a sign and an interpretant are *not* in a determinate relationship – misinterpretation always remains a possibility. This is because the relationship between the sign and its object *constrains* the type of interpretant that may justifiably be applicable to a given sign. Furthermore, there can always be instances where a sign cannot be applied to an object with any meaningful result. In this manner, Peirce attempts to preserve both the constructive aspect of knowledge (the interpretative element that "translates" the sign into some form of mind), and the possibility of true knowledge (ensured by the constraints placed upon possible interpretations by an external "reality" (Pierce 1965 Vol. V:324-325 [5.473].

An interpretant can always potentially function as a further sign, in a potentially endless series of interpretations. Peirce emphasises that every sign is essentially incomplete and our competent understanding or use of a given sign is always a matter of degree (Short 1966:127). As such, the status of an interpretation of a sign mirrors that of any scientific hypothesis: always open to further questioning and testing. The accuracy of any given interpretation is determined by the context in which it takes place (Peirce 1965 Vol. V:154-154 [5.262-5.263]; 170-174 [5.286-5.290]); 183-184 [5.305-5.307] 300-302 [5.448n], among others). Thus, with his triadic conception of reference, Peirce adds a vital third element to the signification process: interpretation. A sign can only be a sign in as far as it is interpreted to be one, or in as far as it possesses the potential to be interpreted. This seemingly obvious insight allows for a fundamental shift in the theories of signs, language, and mind. However, as we shall see, Peirce does not fully avail himself of this fundamental conceptual shift.

---

[37] Cf. Peirce (1965 Vol. V: 324-325 [5.473]).

[38] Sausurre (1983:67-68).

[39] Cf. Peirce (1965 Vol. V:324-343 [5.473 – 5.493]). This observation links with Peirce's theory of *abduction*. As will become apparent in Chapter 2, Peirce's theory of abduction inspired Chomsky's postulation of the existence of a Universal Grammar.

4.2 ICON, INDEX, SYMBOL

As part of his larger scheme of semiotic relationships[40], Peirce distinguished between three categories of referential associations – those of *icon*, *index*, and *symbol*. He uses these categories to describe the formal relationship between signs and the objects that they represent. According to Deacon (1997a:70), Peirce's three forms of reference reflect the classic trichotomy of possible modes of reference in philosophy, which are a) similarity, b) correlation, and c) law, convention, or causality. What appeals to Deacon in Peirce's analysis, as opposed to other prominent philosophers of mind (Locke, Hume, Kant, etc.) is that he "rephrased the problem of mind in terms of communication, arguing that all forms of thought (ideas) are essentially communication (transmission of signs), organised by an underlying logic…that is not fundamentally different for communication processes inside or outside of brains" (70). Similarly, Deacon applauds Peirce's notion that it might be possible to investigate the logic of thought processes by studying the sign production processes in forms of communication accessible to us (see above).

Peirce comprehensively defines and discusses each of the three possible referential associations that he attributes to signs:[41]

For an object or symbol to be an icon of something else, it needs to resemble that entity in some way.[42] Peirce sees the resembling characteristic as something that "naturally" belongs to the iconic object, which it "would posses just the same though its object did not exist" (1965 Vol. V:50 [5.73]). Religious icons are the obvious example that one can think of in this regard.

For something to be an index, it needs to be either causally related to, or spacio-temporally associated with, the thing of which it is an index.[43] An index is relational by virtue of a characteristic that it would not posses if its object did not exist, but which it would continue to posses, whether it is interpreted or not. Peirce includes all natural signs and physical systems in this category (1965 Vol. III:211 [3.361]). A thermometer would be a classical example of an index – it indicates the temperature of its surroundings, and usually

---

[40] Icon, index, and symbol are closely linked to Firstness, Secondness, and Thirdness respectively (See, for example, Melrose 1995:496)

[41] cf. Peirce (1965 Vol. V:50-52 [5.73-5.76]) and (1965 Vol. III:211-214 [3.359-3.365]).

[42] Peirce calls signs that are iconic in the sense that they represent their objects by their similarity hypoicons (Peirce 1965 Vol. II:157 [2.275]). He further subdivides hypoicons into those that share qualities with their object (*images*), those that mirror the relationships between the parts of their object in their own make-up (*diagrams*), and those that represent a parallelism in something else (*metaphors*) (Peirce 1965 Vol. II:157 [2.277]).

[43] As a useful summary of Peirce's definition, Melrose (1995:496) quotes the following extract from an article written by Peirce in Baldwin's *Dictionary of Philosophy and Psychology* (1902):

> Indices may be distinguished from other signs […] by three characteristic marks: first, that they have no significant resemblance to their objects; second, that they refer to individuals, single units, single collections of units, or single continua; third, that they direct the attention of their objects by blind compulsion […] Psychologically, the action of indices depends upon association by contiguity, and not upon association by resemblance or upon intellectual operations.

works in an analogue manner. It might cease conveying information if no-one were around to interpret its reading, but the thermometer would continue to function al the same.

For something to be a symbol of something else, the link between them needs to be established by means of convention, agreement, or code.[44] Peirce describes a symbol as something "whose representative character consists precisely in its being a rule that will determine its interpretant. There need not be any similarity or factual connection with its object, and it fulfils its function only when interpreted (1965 Vol. III:211 [3.361]). For Peirce, "[a]ll words, sentences, and other conventional signs are Symbols" (1965 Vol. II:165 [2.292]). Hence, musical notation is an example of a system of symbols that represents the music it codifies.

It is important to note that there is nothing necessary to the system we are accustomed to using; it came into being by means of convention. Symbols develop out of other signs (especially icons) and Peirce sees "the symbol part of signs" as akin to what is usually meant by *concept* (1965 Vol. II:169 [2.302]. We create new symbols by means of the concepts in terms of which we think, "[s]o it is only out of symbols that a new symbol can grow" (169 [2.302]). Furthermore, the meanings of symbols are constituted, and grow and change, within "use and "experience". Hence, the meaning of a symbol is not constant – it changes and adapts over time.

Peirce gives the following concise and useful summary of his analysis (1965 Vol. II: 168-169 [2.299]):

> The icon has no dynamical connection with the object it represents; it simply happens that its qualities resemble those of that object, and excite analogous sensations in the mind for which it is a likeness. But it really stands unconnected with them. The index is physically connected with its objects; they make an organic pair, but the interpreting mind has nothing to do with this connection, except remarking it after it is established. The symbol is connected with its object by virtue of the idea of the symbol-using mind, without which no such connection would exist.

In keeping with his metaphysical idealism and his conception of a semiotic as having truth as its regulative ideal, Peirce's signs are teleological. Signs share the unifying purpose of expressing reality as accurately as possible, and reality is intelligible because Peirce conceives of it as operating according to semiotic principles (Peirce 1965 Vol. I:1.259 [1.487]; Parker 1994:60).[45]

Gallie (1966:46) sees the realisation that "every symbol – be it a word, a sentence, or a scientific formula – is essentially something *to be developed*, something that requires or

---

[44] Peirce further divides symbols into a term-proposition-argument trichotomy (1965:52 [5.76]).

[45] In the words of Parker (1994:62): "The real is clearly to be conceived, at least in part, as a semiotic phenomenon and hence as metaphysically ideal."

calls for development if it is to fulfil its proper function of expressing and communicating intelligent thought" as Peirce's most fundamental philosophical insight. As we have seen, in Peirce's understanding all signs need to be interpreted, and such interpretation always happens in terms of further signs – the meaning of a sign is enhanced, or developed, by means of other signs (1965 Vol. V:169-171 [5.284-5.285]). Furthermore, a sign must be repeatable and be applicable and meaningful in a world that is continually changing, and where an exact set of circumstances would in all probability not be repeated. Hence, the circumstances in which a sign refers to an object, and in which a sign is interpreted are in flux. Since Peirce conceives of knowledge as the communication of signs, a viable epistemology requires a theory of signs (semiotic) that is adequate to these circumstances. There can be no such thing as an absolutely simple, self-sufficient, and universal conception – every conception is necessarily already related to other conceptions (Gallie 1966:48).[46]

To understand a sign or a hypothesis means knowing how to produce evidence for or against it; understanding what practical consequences might result from the truth of a given conception, in other words, by working out its implications. According to Peirce, the sum of the possible practical consequences that can necessarily result from the truth of that conception constitute the entire meaning of the conception (1965 Vol. V: 6 [5.9]). Elsewhere Peirce defines meaning as the "entire intended general interpretant" of a given term, proposition, or argument (110-111 [5.179]).

Language is inherently vague for Peirce, which leads him to conclude, "Every utterance naturally leaves the right of further exposition to the utterer" (1965 Vol. V:299 [5.447]). The vagueness of language is counteracted through conventions, which limit and define possible interpretations (Gallie 1966:175). Discrepancies in interpretation are inevitable, given the different contexts and different experiences of the utterer and the interpreter. The "vagueness" of language is due to its development, and the changes that particular languages undergo over time and in space. This leads Peirce to conclude (1965 Vol. II:130 [2.222]):

> [E]very symbol is a living thing, in a very strict sense that is no mere figure of speech. The body of the symbol changes slowly, but its meaning inevitably grows, incorporates new elements and throws off old ones. But the effort of all should be to keep *essence* of every scientific term unchanged and exact; although absolute exactitude is not so much as conceivable. Every symbol is, in its origin, either an image of the idea signified, or a reminiscence of some

[46] Such a relational understanding of meaning may give rise to objections that, if all signs and conceptions are defined in terms of further signs and conceptions, we are trapped in a self-referential circularity that renders definition meaningless. Gallie (1966:48) maintains that Peirce's experience as a symbolic logician causes him to hold that all definitions are circular, to some degree. He goes on to distinguish between *viscously* circular definitions – where a definition refers to the very term that it is supposed to define – and inevitable, "innocent" circularity, where concepts are defined in terms of relational, correlated terms (e.g. greater and less). Because *all* conceptions are essentially relational and because there are no genuine conceptions, they can only be defined in terms of "their mutual implications."

> individual occurrence, person or thing, connected with its meaning, or is a metaphor.

As Peirce rightly points out, this understanding of language is present in our common-senseconception of how language works and manifests itself in how we employ language: "…in ordinary life all our statements, it is well understood, are, in the main rough approximations of what we mean to convey (1965 Vol. V:396 [5.568]). This insight would explain, for example, the need in spoken language to emphasise and clarify meaning through gestures, facial expressions, inflection, and the like. In written language, various formal and informal ways have been devised in an attempt to clarify meaning (think of the use of icons and "smileys" in text and msn messaging, for example).

For Peirce, because a sign is partly determined by previous thoughts of its object and is subject to the law of association and continuous development in the mind, it is not identical with the thing signified. He concludes that a sign must have "some characters which belong to it in itself, and have nothing to do with its representative function"; these are the sign's material qualities [1965 Vol. V:169-171 (5.283-5.287)]. Furthermore, he concludes that signs must be capable of being connected with other signs of the same object or the object itself. Without the ability to be connected, signs would be without value. Peirce gives the example of the weathercock as an example of a sign the usefulness of which relies entirely in its being connected to what it signifies through the power of association. He calls the connection of the sign with its object, which he believes can be direct or through the mediation of a further sign, the *pure demonstrative application* of the sign. From this analysis, he concludes that the representative function of a sign does not lie in its material qualities or in its pure demonstrative application. The reason for this conclusion is that the representative function of a sign lies in its relation to *thought*.

For Peirce, material qualities and demonstrative applications belong to the sign, independently of any thought. Hence his conclusion: "…if I take all the things which have certain qualities and physically connect them with another series of things, they become fit to be signs" [1965 Vol. V:169-171 (5.283-5.287)]. For Peirce, representation (meaning) is always a function that is relational to thought. However, no single thought has meaning in itself. Meaning only arises when a thought may be connected with subsequent thoughts, through representation, so that "the meaning of a thought is altogether something virtual" (173 [5.289]).[47] Cognition and representation arise in the relation of states of mind at different instants.[48] Peirce calls these kinds of thoughts "complex thoughts". He also allows for non-

---

[47] Peirce later formulates this point as follows: "…the meaning of a word is the conception it conveys" (1965 Vol. V:186 [5.310]). A consequence of this understanding of meaning is that the absolutely incognisable has no meaning, because no conception can be attached to it.

[48] In a footnote to this statement Peirce makes the following observation, which clarifies his point (1965 Vol. V:173 [5.289n]):

> Accordingly, just as we say that a body is in motion, and not that motion is in a body we ought to say that we are in thought and not that thought is in us.

complex thoughts, but such thoughts cannot be anything other than sensations or emotions (Peirce 1965 Vol. V:177 [5.293]).

Peirce comes to the insight that all determination occurs through negation (1965 Vol. V:178 [5.294]).[49] We recognise a characteristic of a given object by comparing the object to another object that does not possess that particular characteristic. It is from this insight that he concludes that it is impossible to have a conception that is universal in every respect (173 [5.289]). In his own words (178 [5.295]):

> We do not obtain the conception of Being…by observing that all the things which we can think of have something in common, for there is no such thing to be observed. We get it by reflecting upon signs – words or thoughts; we observe that different predicates may be attached to the same subject, and that each makes some conception applicable to the subject; then we imagine that a subject has something true of it merely because a predicate (no matter what) is attached to it – and that we call Being. The conception of being is therefore a conception about a sign – a thought, or word; and since it is not applicable to every sign, it is not primarily universal, although it is so in its mediate application to things.

Together with sensation, thought consists of the power of abstraction or attention (Peirce 1965 Vol. V:178 [5.295]). Attention is not an object of immediate consciousness, but focuses on "one of the objective elements of consciousness", that of cognition. Attention consists in the capacity that the cognition it emphasises has for affecting memory or influencing subsequent thought (179 [5.295]). One cannot speak of subconscious attention, in other words. Peirce argues that the greater the attention that is paid, the stronger and more enduring the associated memory will be, and the more accurate the subsequent logical sequence of thought will be. Attention may also recover a thought that has been forgotten. He sees this analysis of the action of attention as proof that it is the power by which thoughts are connected and related in time.

Attention arises when the same phenomenon or predicate repeatedly presents itself to consciousness. Through an act of induction, we come to the conclusion that such phenomena have something in common. Hence, attention is an act of induction (Peirce 1965 Vol. V:179 [5.296]). Attention has certain effects on the nervous system – it creates habits or "nervous associations" (179 [5.297]. When we recognise that a certain act has the same consequence on a number of occasions, a habit is formed. Hence, the formation of a habit is

[49] This is one of the many parallels that can be drawn between Peirce and Saussure, who gave rise to twentieth century structuralism with his argument that language consists of a system of differences.

an induction, and is necessarily connected with attention (abstraction). Our voluntary actions arise from such habits.[50]

Peirce is led to conclude that we do not have something like an image in our imagination of objects in the world. What we have is the consciousness that we will recognise something when we see it (1965 Vol.V:181 [5.300]). We have no intuitive power of distinguishing between one subjective mode of cognition and another. Peirce compares cognition to dreaming, where we might think that something is presented to consciousness in the form of a picture, while "it is really constructed from slight data by understanding" (182 [5.303]). In trying to reconstruct and relate images that seemed to make sense while we were dreaming, we often need to add something that might not have been present in the dream in order to obtain some consistency which can aid us in relating the dream in some intelligible way. But he goes even further than cognition and argues that we have no images, even in perception (182 [5.303]). One of the reasons that he gives for this assertion is that we are unaware of the blind spot near the middle of the retina. Peirce argues that the blind-spot-less picture that we "see" must be a construction of the mind, "at the suggestion of previous sensations" (182). He proposes that these sensations are signs and argues that understanding could attain all the knowledge that we derive from outward sight, by reasoning from them. The image or representation that is thus formed cannot be absolutely determinate, however, given its "composite" character.

This line of argumentation leads Peirce to question whether we have any conceptions that do not take the form of judgements (1965 Vol. V:184 [5.307]). Predictably, his answer is "no".[51] His argument is that signs denote what they do on the grounds of the three principles of resemblance, contiguity, and causality. In other words, signs are associated with what they denote on one or more of these three grounds. He further argues that the association of ideas consists in a judgement occasioning another judgement, of which it is the sign. Hence, all our general conceptions take on the form of judgement, more specifically of inference.[52] All association is by signs (184-185 [5.307–5.309]).[53] This means that all our cognitions are derived logically from earlier cognitions. But, barring the possibility that this regression continues ad infinitum, where do our first cognitions, our first concepts come from?

Peirce's argues for his solution to this problem as follows (186 [5.311]):

---

[50] As we shall see, Peirce does not limit the formation of habits to human beings, or even to intelligent/conscious entities.

[51] See 5.545 (1965 Vol. V:385) where Peirce explicitly states, "Every new concept first comes to the mind in a judgement."

[52] See also 5.350-5.351 (1965 Vol. V:218-219), where Peirce explains that the universality of inference from induction is "only the analogue of true universality." As we shall see, true universality ultimately lies in the agreement of "all men", given a sufficiently long succession of inferences.

[53] See 5.480 (1965 Vol. V:330) where Peirce argues that both animals and humans have similar first concepts or abductions (compare footnote 36). Human concepts take the form of conjecture or *first logical interpretants* of the phenomena that suggest them.

> At any moment we are in possession of certain information, that is, of cognitions which have been logically derived by induction and hypothesis from previous cognitions which are less general, less distinct, and of which we have a less lively consciousness. These in their turn have been from others still less general, less distinct, and less vivid; and so on back to the ideal first, which is quite singular and quite out of consciousness. This ideal first is the particular thing in itself. It does not exist *as such*. That is, there is no thing which is in itself in the sense of not being relative to the mind, though things which are relative to the mind doubtless are, apart from that relation.

Peirce reiterates in a footnote that his understanding of the ideal is "the limit which the possible cannot attain" (1965 Vol. V:186 [5.311]). In this conception, cognitions reach us by an infinite series of inductions and hypotheses. Such cognitions can be true or untrue, i.e. their objects can be real or unreal. Peirce believes that our ability to distinguish between the real and unreal arises from the fact that we can correct ourselves, from finding a certain conception contradicted by experience.[54] The real is that which would be confirmed independently of personal vagaries or idiosyncrasies in conception. From this, Peirce concludes that the conception of reality involves the notion of a community in its essence. Such a community does not have definite limits, and is capable of a definite increase in knowledge (186 [5.311]). The real will continue to be re-affirmed by the community, while the unreal would continue to be denied. From this, Peirce concludes, "there is nothing…to prevent us from knowing outward things as they really are, and it is most likely that we do thus know them in numberless cases, although we can never be absolutely certain of doing so in any special case" (186 [5.311]). Every truth will be known in time, and once known will stay known.

Peirce argues that, since no individual cognition of ours can be absolutely determinate, generals must exist (1965 Vol. V:187 [5.312]).[55] Far from being a "metaphysical fiction", Peirce presents the belief in existent generals as a realist position, where there is no "more recondite reality than that which is represented in a true representation." If a word is true of something, what it means – its general – is real. His argument is that a reality which has no representation has no relation or quality.

From the argument that generals exist, Peirce moves on to describing the reality of mind. He reiterates his claim that the content of consciousness, or the "entire phenomenal manifestation of mind", is a sign resulting from inference (188 [5.313]). Because of his

---

[54] See also 5.405-5.410 (1965 Vol. V: 265-271) and 5.581 (404).

[55] Peirce argues that there are two kinds of indeterminacy: indefiniteness and generality. Indefiniteness consists in one not being able to make an indubitable, determinate interpretation of a sign, because it does not express itself "sufficiently". Generality gives the interpreter the right to complete the determination "as he pleases" (hence, supply part of the meaning of the sign) (1965 Vol. V:300-302 [5.448n]). Peirce argues that the sign is indeterminate in the latter sense. And the reason he gives for his claim is that the entire universe (and not merely the universe of existents, "thus the universe that we are all accustomed to refer to as 'the truth'") is perfused with signs, if not completely composed of signs.

contention that the incognisable does not exist and its implication that the phenomenal manifestation of a substance "is the substance", he concludes that the mind is a sign developing according to the laws of inference. This, then, is what Peirce means with his numerous statements that man is a sign.[56] Peirce goes on to argue that man can be distinguished from a word in terms of material qualities, but insists that these two entities have much in common. However, because man can only think "by means of words or other external symbols", we cannot mean anything other than what we have learned from symbols. For the same reason, words cannot mean anything other than what man wants them to mean (Peirce 1965 Vol. V:188 [5.313]):

> In fact, therefore, men and words reciprocally educate each other; each increase of a man's information involves and is involved by, a corresponding increase of a word's information.

In Peirce's understanding of the mind, there is no element of consciousness that does not have something corresponding to it in a word or sign. Language is the sum total of the self (189 [5.314]). The body (organism) is only an instrument of thought. Peirce argues that human identity lies in the consistency of what he thinks and does, while identity is equated with "the intellectual character of a thing" (189 [5.315]).

Ultimately, the reality of anything comes down to Peirce's conception of the ideal state of complete information. All of reality, for Peirce, depends on the ultimate consensus of the community at some ideal future date when thought is fully developed and all is known. The individual, apart from a community of thought can only be a negation, or in that famous formulation, a "glassy essence" (Peirce 1965 Vol. V:189 [5.316-5.317]). Reality does, however, make its presence felt and hence influences the interpretation of meaning in the compulsions that it imposes on us. These "compulsions" are perhaps better described as constraints. We have very little control over the constraints that reality imposes on our interpretations and the meanings we ascribe to symbols. Closely linked to the insurmountable immediacy of reality is the "intentionality" of signs.

We have seen that the intended effect, or proper significate outcome of a sign is its interpretant. The interpretant need not be mental, but can consist of actions and feelings, as well (see 1965 Vol. V:324-325 [5.473] and 326-327 [5.475-5.476]). Peirce makes "the triadic production of an interpretant" essential to a sign, and defines the interpretant as all that is explicit in the sign, apart from its context. Peirce divides interpretants into three general classes, namely emotional interpretants, energetic interpretants, and logical interpretants. The proper significate effect of an interpretant could thus be to cause a feeling, cause a response, or create a mental effect. One or more of these outcomes would thus be the intended effect of a sign.

---

[56] E.g. 5.283 (1965 Vol. V:169), 5.314 (1965 Vol. V:189).

The essential mental effect that the logical interpretant creates is that of a *habit* (Peirce 1965 Vol. V:334 [5.487]). Peirce argues that it is only muscular effort of some sort or another that can modify our perceptions of the real (or external) world, whereas the fancies of the inner world can be modified at will (non-muscularly). In other words, those things that only exist in our imaginations are not subject to any constraints, while those perceptions that are deemed to be facts, need to be modifiable through experience. Pierce argues that we can be affected by both our precepts and fancies, and the nature and the extent of such effects depends on both our dispositions and our habits (334 [5.487]). While habits have been acquired through repetitive behaviour, dispositions are inborn. In other words, if similar combinations of precepts and fancies have resulted in similar behaviours on more than one occasion, the tendency or habit to behave in a similar way under similar circumstances develops. Such reiterations of situations can take place in both the inner and the outer worlds, thus in terms of both fancies and percepts. As we have seen, these habits have the power to influence further behaviour. Thus, a person develops the habit of acting in a certain way when he desires a given result. Peirce argues that the *final* logical interpretant cannot be a concept, proposition, or argument. Only the deliberately formed, self-analysing habit is "the veritable and final logical interpretant" (342 [5.491]).[57] "Consequently, the most perfect account of a concept that words can convey will consist in a description of the habit which that concept is calculated to produce" (342 [5.491]) (See also 330 [5.480]).[58]

Peirce does not limit the formation of habits to the mental, however (1965 Vol. V:342 [5.492]).[59] He holds that a stream forms a habit when it carves out a bed for itself, while plants take habits in growing in a particular way.[60] Habits are not limited to the mental, because they are completely unconscious, and Peirce takes consciousness as being the only distinctive characteristic of the mind. Peirce goes on to assert that nothing but feeling is exclusively mental (342 [5.492]). He insists, however, that he does not mean that the mental is merely epiphenomenal. The mind is both influenced by the external world, and capable of affecting the outer world through the operation of habits (342-343 [5.493]).

In Peirce's teleological conception of the actions of signs, it becomes the purpose of every sign to "express fact" and, by being joined with other signs, to determine an interpretant's meaning as nearly as possible to its absolutely True meaning, which can, in principle, be determined by a community of speakers (1965 Vol. II: 134–136 [2.227–2.229]). "Nature's signs and human thinking go hand in hand" (Radford 2004:7). Whereas physical

---

[57] See 5.491 (1965 Vol. V:342).

[58] Habits can also be formed on the grounds of repetitive imagining of acting in a certain way in ascertain situation (1965 Vol. V: 377 [5.539]).

[59] "What we call a Thing is a cluster or habit of reactions, or to use a more familiar phrase, is a centre of forces" (Peirce 1965 Vol. IV:135 [4.157]).

[60] See 5.538 (1965 Vol. V:376) where Peirce's definition of habit is given as something that "denotes such a specialisation, original or acquired, of the nature of a man, or an animal, or a vine, or a crystallisable substance, or anything else, that he or it will behave or always tend to behave, in a way describable in general terms upon every occasion (or upon a considerable proportion of the occasions) that may present itself of a generally describable character."

processes would exhibit dyadic action, "mental" processes (keeping in mind that "mental" is not understood as being limited to species with minds/brains) "involves a triad of symbols" (Peirce 1965 Vol. II:169 [2.300]). Thus, the modes of relation brought about by action between signs become fundamental to our understanding of the universe.[61] As Ransdell (1976:104) puts it, the ultimate utterer of all signs – all interpretable phenomena – is reality itself.

5. A MISINTERPRETATION

Meyers (1992:505) questions the philosophical significance of Peirce's theory of signs and questions whether it is relevant to modern discussions of language and mind. Subsequently he comes to the conclusion that Peirce's theory is a failure, primarily because Peirce fails to explain "the basic mystery of reference" (518-519). Meyer's failure in comprehending Peirce's relevance might be attributable to the fact that he restricts his discussion of Peirce's work to "cognitive" or "intellectual" signs (symbols), and completely disregards icons and indices. Furthermore, he declares "the interesting philosophical issues" to be centred on "intentional signs", which he defines as signs "referring or being used to refer to something other than themselves" (505). According to Meyers, Peirce attempts to reduce "intentional signs" to "natural signs" (506), and fails in his attempt to "incorporate the notion of semantic content within a naturalistic conception of the world", based on the rejection of dualism (515). Meyers believes that Peirce's rejection of dualism commits him to viewing sign-use as consistent with the rest of nature, and furthermore to denying that "non-physical laws" come into play with sign-use. Meyers further describes Peirce as "insisting" on the intentionality of sign behaviour, and hence trying to incorporate intentionality within a "naturalistic, non-dualistic scheme" (515). Meyers's position does not seem to be that Peirce is an "extreme" non-dualist *a la* Quine. He sees Peirce as reducing the mental to an "enriched concept of the physical", one where intentional notions are always already presumed (516). In Meyers's formulation, nature is "intentionalised" (516).

Meyers further rejects Peirce's suggestion that the relation between a sign and its interpretant is based on habit that associates the sign with its customary interpretants, that this habit is law-like, and that these laws are teleological (Meyers 1992:517; 518). His rejection of this understanding of meaning is based on his criticism that it "does not explain how any one sign can be *about* a non-sign" (518). According to him, the explanation is always given in terms of another sign and this does not explain the "notion of the representative function" itself; it does not explain the "aboutness" of signs, and he rejects Peirce's theory of signs as a failure. Meyers goes on to assert that no theory has been able to explain "the

[61] See Parker (1994:61-64) for a discussion on how dyadic and triadic action in the universe corresponds to real things and existent things (as well as to the type-token distinction).

essential intentionality of signs" (519). Meyers takes the view of Brentano and Chisholm that intentionality (as Brentano understood it) cannot be explained by means of science.[62]

In reply to Meyers, Short (1996:524) argues that, while Peirce did not provide a full explanation of reference, or of the intentionality of signs, his theory indicates that the answer lies in the "goal-directed process of semiosis rather than in any class of interpretant". Short argues that Peirce does *not* try to explain significance (meaning) in terms of the intentionality of thought, but conceived of thought as "a type of sign". In this instance, intentionality becomes a "special case" of significance or meaning (525). Short believes that Meyers errs in assuming that Peirce takes the meaning of the sign as a fixed rule for interpreting it. In effect, the "rule" at stake here is transitory, in that it comes into being within a specific context, and hence is only applicable to that particular context. He agrees with Meyers that habit or convention becomes the ultimate logical interpretant of intellectual signs in Peirce's later work, but emphasises that Peirce identified meaning with its interpretant and *not* with the rules by which signs are interpreted, and that the interpretant is not a rule of interpretation (526).

Short's second objection to Meyers is that the "habits" that Meyers refers to are not linguistic habits that govern the association of *verbal formulations* (my emphasis) (1992:526). Meyers describes them as habits that associate verbal formulations with one another. But, Short argues, they in fact regulate extra-verbal conduct and function outside the realm of words. Their semiotic function is not to connect words among themselves, but to "bring the process by which words are related to words to an end". Short believes that the importance of holding habit to be the ultimate logical interpretant is that it allows for the fact that the meaning of concepts and propositions depends on practical considerations or experience. Habits indicate how concepts and propositions can be brought to the test and open the possibility that one can distinguish between a correct and incorrect interpretation. It becomes the process of interpretation that confers "intentionality" to both the sign and the interpretant (527).[63]

Short (1996:527) rightly argues that Meyers got off to a wrong start when he divided signs into causal ("natural") and conventional ("intentional") signs, and then limits the "interesting philosophical issues" to intentional signs (Meyers 1992:505). All signs (including natural symbols) signify, and therefore possess intentionality. There is no separate interpretant that bestows intentionality, and Short describes "thought-interpretants" as "merely additional signs" (1996:527). These and other interpretants such as habits, actions,[64] and

---

[62] See the introduction and our discussion in Chapter 4 of Brentano's conception of intentionality and its influence on twentieth century philosophy.

[63] "It is the teleological structure of semeiosis that explains the intentionality of its parts" (Short 1996:527).

[64] Short argues that "inexistence" is explicable in naturalistic terms, and is therefore not limited exclusively to human consciousness (cf. Short 1996:528-529). See also Liszka (1994:30).

feelings all derive intentionality from their respective roles in semiosis.[65] The major difference between thoughts and other interpretants is that thoughts are also signs (529). Were Meyers to include the categories of icons and indices, which he had excluded from the outset of his analysis, he might have come to a different conclusion regarding the success of Peirce's semiotic.

5.1 The Problem with Objects and Minds

Although we have dismissed Meyers's criticism, two fundamental questions arise with regard to the triadic structure of signs that are not as easily dismissed. The first has to do with the implication that every sign has an object. We have seen that Peirce's conception of what constitutes a sign is very broad, and that he definitely does not limit his conception of signs to cognitive signs.[66] Peirce argues, for example that the order of an officer to "Ground arms!" is a sign (1965 Vol. V:324 [5.473]), as is the performance of a piece of music (326 [5.475]). In the first case the object is the will of the officer that his troops lower their muskets to the ground, while in the latter it is the composer's musical ideas (even though Peirce sees them as "merely" a "series of feelings").[67] Furthermore, he does not regard the formation of habits, which as we have seen are the result of continually associating a particular interpretant with a particular sign, as limited to human minds (Peirce 1965 Vol. V:342 [5.492]):

> …habit is by no means exclusively a mental fact. Empirically we find that some plants take habits. The stream of water that wears a bed for itself is forming a habit…habits in themselves are entirely unconscious, though feelings may be symptoms of them, and…consciousness alone – i.e. feeling – is the only distinctive attribute of mind.

And (1965 Vol. V: 325 [5.473]):

> whether the interpretant be necessarily a triadic result is a question of words, that is, of how we limit the extension of the term "sign"; but it seems to me convenient to make the triadic production of the interpretant essential to a

---

[65] It is important to reiterate here that Short sees aspects of his own theory as finishing what Peirce had "left incomplete" in his later writings (1996:527). What is at issue in terms of this dissertation is to ascertain whether Deacon's use of Peirce's theory of sings is tenable. In this regard, Short's application seems to mirror much of Deacon's, and therefore provides some support to Deacon's theories.

See Gallie (1966:43-45) where he accuses the editors of the Collected Papers of laying Peirce open to charges of inconsistency and of being "a philosopher of threads and patches", through their selection and treatment of Peirce's extensive writings.

[66] Cognitive signs are defined as "those signs whose essence is that in knowing them we know something more" (Colapietro 1987:214).

[67] Pierce divides interpretants in accordance with the "significate" effects of signs into emotional, energetic, and logical interpretants (1965 Vol. V:326-327 [5.475-5.476]).

> "sign,"…On these terms it is very easy…to see what the interpretant of a sign is: it is all that is explicit in the sign itself, apart from its context and circumstances of utterance.

Can “signs” such as musical performances and imperatives be said to have an “object” in any meaningful sense of the word (cf. Colapietro 1987:207-112)? Short (1981:217) argues that signs such as “and” do not appear to have a signifying function as they “have no object”, and as such, should not be classified as *bona fide* signs. Is it possible that some words are not signs? In answer to this question Colapietro (1987:216) quotes Peirce as referring to “sufficiently complete” signs in *On the Foundations of Mathematics*. He takes this to indicate that Peirce thinks of signs as complexes of parts, some of which may be absent from an entity without depriving it of its status as a sign. In this understanding “and” might not be a complete sign, but it would still be a sign.

The case of imperatives and other “incomplete” signs raises the issue of context and the role that it has to play in semiotics. Peirce emphasises the importance of context for sign interpretation. A word such as “and” usually has an indexical function and is used to “conjoin X and Y” (Colapietro 1987:216) within a given context. As such, it can be understood to be referring to something, that something being an action of choosing, or adding, or linking, or elaborating – in other words, an object. Colapietro uses the distinction between sufficiently complete and insufficiently complete signs to argue that all words should be included within the scope of Peirce’s semiotic with the proviso that one is aware that “not all words fully realise the essential function(s) of those teleological processes which are sufficiently complete signs” (217). Colapietro does, however, question the accuracy of the term “representation” for characterising the relationship between a sign and its object in Peirce’s semiosis (218). Essentially he argues that the sign is *passive* in relation to its object, in that the object has an effect upon the sign, but remains unaffected by the relation. Furthermore, he sees the sign as active in relation to its interpretant, determining it without being effected by it (218-219). However, there are passages in Peirce, where the sign is described as being active in relation to its object, even creating it (e.g. Peirce 1965 Vol. IV:422-423 [4.536]).

Peirce addresses this issue by distinguishing between an immediate object (the object as the sign represents it) and the dynamical object (“reality”, or the object “in itself”) (Peirce 1965 Vol. IV:422-423 [4.536]). Colapietro (1987:219) explains the difference between the two as follows: “The sign determines its immediate object, while the dynamic object determines its sign.” This characterisation of both the object-sign and sign-interpretant slips a teleological element into the semiological structure. Colapietro (221) describes both the sign and the object as being something in themselves, apart from their semiotic relationship. The sign forms part of a system of signs, while the system of signs is embodied in a specific context. It is only within such a context that an object can determine its sign (222). The object that is signified need not be real, in the sense that a sign can refer to an object that doesn’t exist, such as phlogiston, or unicorns. Against this background is seems justified to say that,

in Peirce's theory, everything is potentially a sign, and possibly even a sign of many other things.[68] What is not adequately explained is *how* the object determines its sign. How does the context constrain the application of possibilities of the sign? Peirce seems to be content with the explanation that every sign acts in virtue of a final cause (1965 Vol. II: 46-49 [2.86]). All that Peirce regards as necessary to a final (or teleological) cause is that, given certain initial conditions, a certain characteristic will result (Gallie 1966:135). Gallie questions the aptness of this explanation, given that action in accordance with a final cause is usually understood to be an action towards an envisaged end. This is one of the many instances in which Peirce seems to conceive of sign-action as something purposive in itself, quite independent from any mind in which the actual creation and interpretation of signs takes place. Pierce seems to believe that sign-action is the prototype of all non-mechanical forms of activity (Gallie 1966:136).

Appropriately, a further question that arises from the irreducibly triadic character of sign-action is whether the interpretant must be some form of mind, or not. From Peirce's definition of a sign as an entity that stands for something to someone, and his contention that a sign cannot function without being interpreted, it would seem logical to conclude that the interpretant must necessarily be a person, or at least a mind of some sort. But Peirce does not consistently use interpretant in this manner. In fact, in some usages he seems to deliberately shun the idea that some form of mind must be linked to the interpretant (Peirce 1965 Vol. IV:438 [4.551]):

> Thought is not necessarily connected with a brain. It appears in the work of bees, of crystals, and throughout the purely physical world; and one can no more deny that it is really there, than that the colours, the shapes, etc. of objects that are really there.

In his definition of interpretant Peirce states that (1965 Vol. V:110-111 [5.179]):

> It [the sign] addresses somebody, that is, creates in the mind of that person an equivalent sign. That sign which it creates I call the interpretant of the first sign.

This leads to an understanding of the sign as something transcendental, unrelated to the entity for which it is a sign.[69] Gallie (1966:118) takes this to mean that when Peirce refers to the interpretant of a sign, he generally refers to a further sign of the object of the given sign.

---

[68] For an opposing point of view, see Greenlee (1976:137) who describes such readings of Peirce as "monstrously mistaken".

[69] According to Colapietro (1987:224), Peirce saw two sides to semiosis: generating a sign on the one hand (*utterance*), and grasping a sign on the other (*interpretation*). Utterance and interpretation are explicitly linked to "Quasi-minds": "a Quasi-utterer and a Quasi-interpreter" (1965 Vol. IV:438-439 [4.550-4.551]). And Colapietro (1987:224) is quick to point out that a quasi-utterer "simply means a source from which a sign springs", and a quasi-interpreter "signifies a form into which a sign grows". He goes on to explicitly deny that there need to be anything "mental" about either of these entities.

Gallie emphasises that in Peirce's general notion of the sign and in his formal theory of sign-action, the interpretant does not necessarily refer to a person or (human) mind, or even to that of some kind of organism, but rather to some further sign (118-125). While Gallie is right in maintaining that a sign can only function as an element in a working system of signs, he fails to adequately explain how the interpretation of a sign as an element within a working system of signs can occur without the mediation of some sort of mind, understood in the broadest sense possible. Here, metaphysics is based on a semiotic, which means that the cosmos can be understood in semiotic terms. In other words, logic becomes a "normative semiotic", and *truth* becomes the regulative ideal of semiotics (Parker 1994:59). The problem is compounded in Peirce's later work when he defines intelligent mind as "a sign developing in accordance with the laws of inference" (1965 Vol. V:188 [5.313]). The considerations that lead him to this definition are, among others, the observation that children seem to communicate before they reason, that animals of various species communicate, but that he is loathe to attribute thought and reasoning to them, and his contention that man thinks by communicating with himself (Gallie 1966:131-132).

As we have seen, Peirce argues that if reality does exist independently of our thoughts, and if it is accessible in principle (as he believes his analysis shows) then Peirce is led to believe that general laws and types (i.e. universals) exist independently of human enquiry (e.g. 1965 Vol. I:8 [1.26]); 1965 Vol. V:287-290 [5.430-5.434]).[70] In other words, there is an irreducibly general element in every (reliable) sign that corresponds to an irreducibly general element in the world (Gallie 1966:148). Realism, for Peirce, means accepting that the natural laws are real and this is as important a fact as any other fact of existence (153). As explained by Parker (1994:56):

> The modern nominalist…values "facts" which are often spoken of in isolation from the system of thought in which they occur. The realist, on the other hand, considers the whole system of ideas itself, including the facts, as the principle object of knowledge. This epistemological priority of generals translates into a metaphysical priority, as well. The nominalist values knowledge of particulars, and sees the general or universal part of thought as having *no* metaphysical significance. This difference in value motivates the tendency of modern nominalism toward a materialistic metaphysics, with its identification of "the real" as particular bits of matter in motion. The realist, however, sees the general as primary, and the general must be conceived in terms of *idea* rather than in terms of *matter*. Peirce's affirmation of metaphysical idealism is thus rooted in his logical realism.

[70] See Bernstein (1965:79) where he describes meaning as consisting of habits or conduct: "these are the generals or universals that are inescapable aspects of reality and experience."

6. Natural Laws, Evolution, and Laws of Habit

We have seen that Peirce conflates the nature of natural laws with the nature of signs.[71] With regard to natural laws, Peirce argues that the only way to account for the laws of nature – and for uniformity in general – is to regard them as the result of evolution (1965 Vol. VI:15 [6.13]).[72] He defines evolution as growth "in the widest sense of the word" (1965 Vol. I:71 [1.174]), and argues that growth does not entail mere increase, but diversification. Reproduction is one incidence of growth among others, and hence Pierce casts his conception of the mechanisms of evolution wider than that of Darwin.[73] Furthermore, he argues that the world does bear evidence of diversification or an increase in variety, in that the nebula from which the solar system originated was far simpler than the solar system it spawned, for example. Hence, he concludes that mechanical laws which science views as the only agency of nature, cannot produce diversification – the very nature of such a law presupposes that "like precedents can only produce like consequents" (72 [1.174]).[74] Peirce argues that the world shows evidence of real growth, and concludes that the world (universe) is subject to evolutionary, rather than mechanical law.

---

[71] An apt example of his reasoning reads as follows(1965 Vol. I:159 [1.316]) :

> [E]very scientific explanation of a natural phenomenon is a hypothesis that there is something in nature to which human reason is analogous; and that it really is so that all the successes of science in its application to human convenience are witnesses. They proclaim that truth over the length and breadth of the modern world. In the light of the successes of science to my mind there is a degree of baseness in denying our birthright as children of God and in shamefacedly slinking away from anthropomorphic conceptions of the universe.

[72] See also: Peirce (1965 Vol. I:42-45 [1.103-1.109]), and Peirce (1965 Vol. I:71-72 [1.173-1.175).

[73] Peirce agreed with the Darwinian theory, in that he believed evolution to be an essentially adaptive process. As such, it could not fit neatly into the "Ockam-Descatres-Hobbes-Newton world-picture". Peirce conceded that natural selection is the main cause of evolution, but he argues that biological evidence pointed to other causes as well (Gallie 1966:235). As Chomsky points out (2000b:163), Darwin denied that he believed the modification of the species to be due to natural selection only. His position (which is explicitly stated) is that natural selection is the main, but not exclusive, means of modification. He also noted a range of other possible factors, such as non-adaptive modifications, and functions that were not selected for, but determined by structure.

[74] Hence Peirce's claim that the "mechanical philosophy of the universe" is false (1965 Vol. VI:372 [6.553]):

> The endless variety of the universe has not been created by law. It is not of the nature of uniformity to originate variation, nor of law to beget circumstance. When we gaze upon the multifariousness of nature we are looking straight into the face of living spontaneity.

He goes on to argue that the law of evolution is not mechanical and cannot be explained as a consequence of the mechanical principle of the conservation of energy (an endeavour he accuses Herbert Spencer of) (see also 1965 Vol. VI:15 [6.13]]). He states his reason for dismissing this attempt as follows (373 [6.554]):

> The principle of the conservation of energy may be stated in this form: whatever changes can be brought about by forces can equally happen in the reverse order (all the movements taking place with the same velocities, but in the reverse directions), under the government of the same forces. Now, the essential of growth is that it takes place in one determinate direction, which is not reversed… It is thus the immediate corollary from the doctrine of the conservation of energy that growth is not the effect of force alone.

Again we see that Peirce foreshadows the sciences of complexity, which would make a similar argument.

Peirce goes on to argue that there is no reason to think that natural laws are absolute.[75] If the universe exhibits continuous growth – movement from non-existence to existence – might the forces of nature not be subject to growth as well (Peirce 1965 Vol. I:72 [1.175])?[76] This leads him to conclude that, just as the meaning of signs grows with experience and time, so general laws operating in physical systems will grow or change.[77] In this manner, Pierce introduces the law of evolution into his cosmology.

Peirce again recalls the three logical principles of First, Second, and Third which he believes are the materials from which a philosophy that most accurately reflects the state of knowledge at the end of the nineteenth century could be built. The metaphysics that results (what he calls a "Cosmogenic Philosophy") works with the assumption that the universe began with “a chaos of unpersonalised feeling, which being without connection or regularity would properly be without existence” (Peirce 1965 Vol. VI:26 [6.33]). Eventually, a “generalising tendency” would have arisen in this chaos, through Darwinian accident perhaps, which was the beginning of the tendency to *habit*. It is from the tendency to habit that all the other principles of evolution and the regularities of the universe would have evolved. An element of pure chance would have survived, though, and remains present until “the world becomes an absolutely perfect, rational, and symmetrical system, in which mind is at last crystallised in the infinitely distant future”. Eventually this process results in "creative evolution".[78]

If the laws of nature are subject to the laws of evolution, and thus liable to change, it means that they are not absolute and that there is an element of indeterminacy, spontaneity, even chance to them (Peirce 1965 Vol. VI: 15 [6.13]):

> Just as, when we attempt to verify any physical law, we find our observations cannot be precisely satisfied by it, and rightly attribute the discrepancy to errors

---

[75] Elsewhere he states, for example: "…I do not believe that anybody, not in a state of case-hardened ignorance respecting the logic of science, can maintain that the precise and universal conformity of facts to law is clearly proved, or even rendered particularly probable, by any observations hitherto made" (Peirce 1965 Vol. VI: 37 [6.48]).

[76] See also Peirce (1965 Vol. VI:41-45 [6.58-6.65]). Peirce’s contention is that the variety and irregularity of the universe cannot be the result of the laws of mechanical necessity.

[77] "As to the common aversion to recognising thought as an active factor in the real world, some of its causes are easily traced. In the first place, people are persuaded that everything that happens in the material universe is a motion completely determined by the inviolable laws of dynamics; and that, they think, leaves no room for any other influence. But the laws of physics stand on quite a different footing from the laws of gravitation, elasticity, electricity, and the like… Now who will deliberately say that our knowledge of these laws is sufficient to make us reasonably confident that they are absolutely eternal and immutable, and that they escape the great law of evolution? Each hereditary character is a law, but it is subject to development and to decay. Each habit of an individual is a law; but these laws are modified so easily by the operation of self-control, that it is one of the most patent of facts that ideals and thought generally have a very great influence on human conduct” (Peirce 1965 Vol. I: 178 [1.348]).

[78] The parallels with Hegel are obvious and although Peirce recognised close affinities between Hegelianism and pragmaticism, he insisted that Hegel’s philosophy was incomplete in that it did not add to our understanding of man in the universe in the way that pragmaticism could (Bernstein 1965: 68). (Bernstein refers to Hegel’s philosophy as "the spectre that haunts most of philosophy in the latter part of the nineteenth century" (68)).

> of observation, so we must suppose far more minute discrepancies to exist owing to the imperfect cogency of the law itself, to a certain swerving of facts from any definite formula.

Peirce goes on to distinguish between three viable variations of evolutionary theory (having dismissed Spencer's attempt at explaining evolution in terms of mechanical principles).[79] The first is that of Darwin, where evolution is brought about by heredity that leaves room for accidental variations that might enable its inheritor to flourish and hence to produce enough viable offspring to ensure that the accidental characteristic is multiplied throughout the species. Peirce sees this as a viable theory of evolution (16 [6.15]).

The second is the Lamarckian theory that changes that take place within an individual of a species through "effort and exercise" can be transferred to its offspring as a hereditary characteristic. Peirce sees Lamarck's theory as viable as well and conceives of the possibility that it might operate alongside Darwinian evolution. He characterises Darwinian evolution as "evolution by the operation of chance" and Lamarckian evolution as "evolution by the effect of habit and effort" (Peirce 1965 Vol. VI:16-17 [6.16]).

Thirdly, he discusses "Catastrophism" as put forth by Clarence King. The argument here is something akin to the Kuhnian (1970) conception of the development of a new scientific paradigm, where the species seem to be little changed under ordinary circumstances, but undergo great changes under novel, or catastrophic, circumstances. Peirce is taken with this argument and argues that changes in the environment could break up habitual ways of life, lead to changes in the kind and amount of food that individual has access to, and lead to changes in the effects of the elements which the individual is exposed to. He argues that these factors could lead to changes in both acquired characteristics and to accidental hereditary changes, hence catastrophes could induce both Darwinian and Lamarckian evolution (Peirce 1965 Vol. VI:17 [6.17]). He also argues that such a theory of evolution would imply that evolutionary changes are neither haphazard (Darwinian), nor determined by "an inward striving" (Lamarckian). This results in an evolution that is sparked by changes in the environment and has a "positive general tendency" to adapt organisms to their environments. With regard to catastrophic evolution, Peirce concludes that (17 [6.17]):

> This mode of evolution, by external forces *and the breaking up of habits*, seems to be called for by the some of the broadest and most important facts of biology and palaeontology; while it certainly has been the chief factor in the historical evolution of institutions as in that of ideas; and cannot possibly be refused a very prominent place in the process of the universe in general [my emphasis].

---

[79] See also Peirce (1965 Vol. VI:197-204 [6.296-6.305]).

7. OBJECTIVE IDEALISM AND THE "LAW OF MIND"

Pierce extends the implications that evolution has for mechanical processes to the mind.[80] He argues that a mechanical conception of the universe cannot but understand the mind as part if the physical universe (Peirce 1965 Vol. VI:42 [6.61]). In such an understanding, free will and consciousness would be reduced to "illusory aspects of the material system". Any choice we make would, in principle, be the result of the mechanical laws that govern matter, and as such would have been predetermined by mechanical events long before we even came into existence.[81]

Peirce argues that the elementary phenomena of the mind fall into three categories: feelings (all that is immediately present to the mind), sensations (the reaction when more than one feeling is present in the mind and there is a relation between them, or when muscular effort is called for), and general conceptions (when the connection between two feelings is governed by a rule)[82] (1965 Vol. VI:17-20 [6.18-6.23]). In an exclusively material and mechanistic universe, "true" subjective feeling would not be able to arise. At most, the illusion of feeling, caused by the distribution of qualities and matter in space and time, would be possible in such a universe.[83] Peirce finds this possibility intolerable. Hence, even though he rejects dualism, he does not endorse materialism in its stead. He sees materialism as repugnant to both common-senseand scientific logic, because it requires that some kind of mechanism (whatever process it is that governs the mind) should feel, seeing that feeling is one of the elementary phenomena of mind.[84] He also rejects "neutral" monism as an alternative to dualism, because he believes that Ockham's razor prohibits the possibility of a monistic universe: Whereas in terms of this maxim one should "not suppose more independent elements than necessary", Peirce argues that placing the "inward and outward substance" on par, would render both primordial, and hence one would be supposing more substances than are necessary (Peirce 1965 Vol. VI:20 [6.24]). Hence, Peirce believes that

---

[80] It is important to emphasise that Peirce does not conceive of mind as a "ghostly substance", nor as something that resides within brains or nervous systems. Ultimately, he understands "mind" to be the reality "which is manifest in the occurrence of triadic action in phenomena" (Parker 1994:59). Peirce formulates a variation on idealism, which he calls "semiotic idealism" or "interpretational idealism" (65). In this understanding, whatever exists, exists by virtue of the representational and interpretational activities of "mind". Hence, whatever exists has the nature of a sign.

[81] According to Peterson (1991:224) Peirce often calls his category of Thirdness the "category of law." For Peirce a law is a general fact which is logically prior to its instances. For him, laws are not just regularities in nature; they precede and explain peculiarities in nature. Hence, wherever there is law, there is Mind, as a manifestation of the evolutionary process moving towards such laws.

See also Aronson (1969).

[82] Peirce explains that general conceptions arise in the mind because of the formation of habits in the nerve cells. Habits in the nerve cells of the brain are molecular changes that occur because of activity in the brain and are "probably connected with its nutrition" (1965 Vol. VI:19 [6.22]).

[83] Chomsky (1966:6-8) describes Descartes' conclusion that a world based on "mechanistic principle" could not allow for the creative and innovative aspects of human behaviour, especially linguistic behaviour, and hence declared that a "thinking substance" must be postulated to account for human behaviour.

[84] cf. Peirce (1965 VI:30 [6.37-6.38]).

he has sufficient reason to insist that mind is not reducible to matter. In fact, he goes further than placing mind and matter on a par (1965 Vol. VI: 42-43 [6.61]):

> …by supposing the rigid exactitude of causation to yield, I care not how little – be it but by a strictly infinitesimal amount – we gain room to insert mind into our scheme, and to put it into the place where it is needed, into the position, as the sole self-intelligible thing, it is entitled to occupy, that of the fountain of existence; and in doing so we resolve the problem of the connection between soul and body.

Peirce, in fact, concludes that the only intelligible theory of the universe is that of *objective idealism*, where "matter is effete mind" (Peirce 1965 Vol. VI:20 [6.25]).[85] He comes to this startling conclusion because he believes that objective idealism is capable of "explaining the tri-dimensionality of space, the laws of motion, and the general characteristics of the universe, with mathematical clearness and precision…" (21 [6.25]). In terms of laws, given that matter is effete mind, we cannot suppose the laws of mind to be limited to the mechanical, any more than we can suppose physical laws to be thus limited (184 [6.277]):

> …we ought to suppose a continuity between the characters of mind and matter, so that matter would be nothing but mind that had such indurated habits as to cause it to act with a peculiarly high degree of mechanical regularity, or routine…This hypothesis might be called materialistic, since it attributes to mind one of the recognized properties of matter, extension, and attributes to all matter a certain excessively low degree of feeling, together with a certain power of taking habits. But it differs essentially from materialism, in that, instead of supposing mind to be governed by blind mechanical law, it *supposes the one original law to be the recognised law of mind, the law of association, of which the laws of matter are regarded as mere special results* [my emphasis].

Pierce contends that psychical causation exists, which is different from physical causation. He understands “the law of mind” to be the association[86] of ideas. In terms of this law ideas tend to spread continuously, and affect other ideas that stand in some sort of relation to them (1965 Vol. VI: 87 [6.104]). As they spread, ideas lose their intensity and their

---

[85] A useful formulation of the theory of objective idealism is given by Smith (1965:102):

> …reality is independent of what is thought by this, that, or the other individual, and even of any finite collection of individuals, but…it is not independent of thought in general.

[86] Peirce describes the notion that "associationalism" is restricted to materialistic phenomena as "that most widely spread of philosophical blunders (1965 Vol. VI:29 [6.36]).

power to affect other ideas, but become more general as they become connected with other ideas" (87 [6.104]).[87]

The law of mind is different from physical laws in that it does not require exact relation of cause and effect; it only makes a given feeling *more likely* to arise. It thus resembles what Peirce calls the “non-conservative” forces of physics, such as viscosity, "which are due to statistical uniformities in the chance encounters of trillions of molecules” (Peirce 1998 [1891]:168-169). He argues that an idea is an event in an individual consciousness, and once it has passed, it is gone forever. However, ideas recur, both in the same consciousness and in different consciousnesses. Peirce believes that such recurrence is only possible within consciousnesses, because "an occult power from the depths of the soul forces us to connect them in our thoughts after they are both no more" (88 [6.106]). If an idea is in consciousness and then passes, it would seem impossible for one to experience it again. However, past ideas are experienced again in the present, which leads Peirce to conclude that the present is connected to the past in a series of infinitesimal steps (89 [6.109]). He argues that consciousness is, and must be, aware of the passing of an interval of time, otherwise we would not have any conception of time. We must be aware of a past experience in a series of changes in order to be able to understand that a present experience is of the same thing, but that it has undergone change.[88] Thus, consciousness and its objects are continuous, and we can relate present and past ideas through "comparative perception": "(l)et there be, not merely indefinite succession, but a continuous flow of inference in a finite time, and the result will be a mediate objective consciousness of the whole in the last moment" (90 [6.111]).[89]

---

[87] “The one primary and fundamental law of mental action consists in a tendency to generalisation. Feeling tends to spread; connections between feelings awaken feelings; neighbouring feelings become assimilated; ideas are apt to reproduce themselves. There are so many formulations of the one law of the growth of mind. When a disturbance of feeling takes place, we have a consciousness of gain, the gain of experience; and a new disturbance will be apt to assimilate itself to the one that preceded it. Feelings, by being excited, become more easily excited, especially in the ways that they have previously been excited. The consciousness of such a habit constitutes a general conception” (Peirce 1998 [1878]:168).

[88] Peirce describes the present as "half past and half to come" (1965 Vol. VI:99 [6.126]).

[89] Gallie (1966:200) sums this principle up as follows:

> Every experience, he claims, contains elements of memory and anticipation, and, more generally, elements of interpretation, however *un*prominent these may be; and the kind of interpretation which memory or anticipation or explicit inferences provide requires that mental action shall be continuous in character. Unless, for example, our every experience contained, as an essential feature of itself, the power to continue acting on or actively to overlap with some of the experiences which succeed it, memory – which involves that what h*as been* still in some sense *is* – would be utterly impossible. More generally, just as every experience which we would ordinarily count as a thought or a sign *means* something in virtue of its capacity to determine a theoretically endless series of interpretants, and is in itself in some degree determined by logically prior thoughts or signs, so *every* experience may be said to have meaning, in a very extended sense, in so far as it serves to mediate continuously between some of the experiences that went before it and some that will succeed it. An experience counts as First in virtue of what it immediately is; as second in virtue of what it immediately or actually does to other things; as Third in virtue of its ‘would-be’ and ‘would-do’ – its capacity to influence *any* of a describable set of later experiences by developing, with reference to such later experiences, the ‘would-be’s’ and ‘would-do’s’ of certain experiences which preceded it. In sum, Peirce maintains that our every experience possesses in some degree what every sign possesses in eminent degree, a *virtual* character, and that the possession of this

In terms of the law of mind, time is irreversible, in that it flows from past to future. The relation of the past to the future is essentially different from that of the future to the past. Pierce sees this aspect of the law of mind as fundamentally different from "the law of physical force", which is essentially reversible (Peirce 1965 Vol. VI:100 [6.127]).[90] Due to the flow of time, every present state of feeling (idea) is affected by earlier states (100 [6.128]). Furthermore, Peirce insists, time cannot exist unless there is something to undergo change. Feelings are not to be limited to human beings, nor should they be understood in terms of sentiment. As we have seen, feelings fall under Peirce's category of Firstness, and should be understood as *quale*, or the possibility of experience. Hence, amoeba and slime-moulds "feel" when they are in a condition of excitation (101 [6.132]). This, for Peirce, means that feeling is not limited to some form of mind, but has spatial extension; he refers to feeling as a phenomenon that "essentially involves externality" (102 [6.133]).

8. Ideas, Feeling, and Protoplasm

An idea consists of three elements: its intrinsic quality as feeling, the energy with which it affects other ideas and the tendency of an idea to bring other ideas with it[91] (Peirce 1965 Vol. VI:103 [6.135]). As an idea spreads, the power that it has to affect other ideas diminishes, but its intrinsic quality (feeling) remains more or less the same. A general idea arises when a series of feelings are "welded together" because of *association* in a finite period of time (103 [6.135–6.137]). Peirce describes an idea that results from such a process as being "living feeling", which is immediately present and consists of a continuous field of quality.

Peirce believes that his analysis of consciousness (and the mental) has dealt a blow to nominalism, because it shows that ideas affect one another, and we can perceive these affects.[92] The continuum of quality also allows for ideas to resemble one another without complication. As we have seen, feeling is not limited to a mind, but Peirce argues that feelings (ideas) can have an effect or be affected even before passing into immediate consciousness. Habit ensures that an idea is brought to consciousness through its bond with an idea already present to consciousness (Peirce 1965 Vol. VI:103-105 [6.138-6.142]):

> We can now see what affection of one idea by another consists in. It is that the affected idea is attached as a logical predicate to the affecting idea as subject.

---

character requires that some elements in our experience are continuous, or are developed continuously from phase to phase.

[90] Peirce was one of the first philosophers of science to suggest that the fundamental division in physics is between the laws of reversible and irreversible processes (Gallie 1966:232).

[91] The tendency to suggest other ideas is, of course, a form of habit. In the light of the law of mind Peirce defines habit as "that specialisation of the law of mind whereby a general idea gains the power of exciting reactions" (1965 Vol. VI:106 [6.145]).

[92] See also Peirce (1965 Vol. VI:109 [6.150]).

> So when a feeling emerges in immediate consciousness, it always appears as a modification of a more or less general object already in the mind. The word suggestion is well adapted to expressing this relation. The future is suggested by, or rather is influenced by the suggestions of, the past.

A central aspect of Peirce's argument here is that ideas cannot be connected without continuity in the form of a continuum of feeling. He comes to this conclusion because he believes that a universe of diversity and chance cannot guarantee that related ideas can come into contact with one another to form general ideas. It is "the law of continuous spreading" that produces mental associations. It is the continuous connection between ideas that are associated in a "living, feeling, and perceiving general idea" (Peirce Vol. VI:105 [6.143]). Associations become habitual and general ideas gain the ability to excite reactions. This mental process is employed when learning to speak a language. Mental actions are not necessary or fixed, by mental law, which allows for flexibility and variations in language and meaning. Pierce sees the uncertainty of the mental law as its essence in that the mind is not subject to law in the same rigid manner that matter is, which allows for spontaneity in mental action (108 [6.147–6.149]).

Peirce insists that general ideas are not "mere words", but living realities. Thus, he reformulates his law of mind as follows (Peirce Vol. VI:109 [6.152]):

> And to say that mental phenomena are governed by law does not mean merely that they are describable by a general formula; but that there is a living idea, a conscious continuum of feeling, which pervades them, and to which they are docile.[93]

In terms of this doctrine, ideas can only be affected by other ideas that are in a continuous connection with it, and nothing else. Ideas can only be conveyed from one mind to the other by means of a continuum, which consists of the diverse elements of nature. Hence, Peirce argues, in order to be able to convey ideas, these elements of nature must, in fact, be embodied ideas, and mechanical energy and mechanical necessity cannot be applied to them. As far as human beings are concerned, feelings are communicated to nerves by continuity, which means that there must be something like nerves in the objects that stimulate them. Again, Peirce comes to the conclusion that what we call matter is, in fact, mind "hidebound with habits" (Peirce 1965 Vol. VI:111 [6.158]). He believes that there is no other way to explain the fact that we have sensations. In a later paper Peirce redefines continuity, because he believes that he did not explain the concept accurately in "The Law of Mind" (1965 Vol. VI:118 [6.174]). Here he argues for a conception of our universe as an offshoot

---

[93] Elsewhere Peirce refers to a quality as consciousness. He refers to this type of consciousness as a quale–consciousness – not a "waking" consciousness, but a possibility of consciousness (1965 Vol. VI:149 [6.221]). The commonalities between qualia (including the synthetical unities that Kant attributes to the mind) originate in the quale-consciousness on which the mind operates (151 [6.225}). As far as human consciousness goes Peirce argues as follows:

> The brain shows no central cell. The unity of consciousness is therefore not of physiological origin. It can only be metaphysical. So far as feelings have any continuity, it is the metaphysical nature of feeling to have a *unity* (1965 Vol. VI:152 [6.229]).

from a Platonic world of ideas. The process of evolution does not only apply to our universe, but also to the process by means of which the Platonic forms are developing (132-135 [6.185-6.194]).

Peirce's metaphysics seems to be narrowly linked with the then prevalent idea of matter consisting of protoplasm (or "life-slimes" as Peirce also refers to it) – a complex and homogenous substance.[94] Two pertinent qualities of protoplasm that Peirce discusses and that explicitly links the theory of protoplasm with his metaphysics are: i) that protoplasm takes on habits[95] and ii) that protoplasm feels. Peirce holds that protoplasm not only feels, but also exercises "all the functions of mind" (1965 Vol. VI:167 [6.254-6.255]). In the case of protoplasm, Peirce believes that mechanical laws are insufficient to account for its tendency to take on habits. Peirce argues that in order for protoplasm to form habits it is necessary for trillions of molecules to aggregate in the same condition and in the same vicinity. Peirce does not believe that mechanical laws are able to account for this "evidence of primordial habit-taking tendency", where "like things [act] in like ways because they are alike" (172 [6.262]). Furthermore, because protoplasm feels, it has consciousness. But, the slime is a chemical compound and can in principle be manufactured in a laboratory. One must then conclude that synthetic protoplasm would also feel. Peirce argues that the laws of mechanics cannot account for the ability of protoplasm to feel, and again solves the dilemma by proposing that physical events are underdeveloped forms of psychical events (172-173 [6.264]). Hence, the diversification of the universe can be explained in terms of the vestiges of chance spontaneity that belongs to the psychical or mental aspect of matter, and mechanical laws are nothing but acquired habits that stem from the mind's tendency to generalise, or spread feelings.

Peirce's objection to materialism is that, if mind is an aspect of matter and matter is governed by mechanical laws, it follows that mind is also governed by such laws (Peirce 1965 Vol. VI:181 [6.275]). This assumption goes against the grain of our common-sensical experience of free will, and it seems to contradict Peirce's contention (as discussed above) that we experience time as flowing in one direction only, while mechanical laws are reversible in principle. Furthermore, it seems impossible to account for the possibility of the interchange of energy between mind and matter and uphold the principle of the conservation of energy (see Peirce 1965 Vol. VI:181-184 [6.275-6.276]). According to Peirce, observation shows that "naked protoplasm" has continuity in time and probably has continuous extension in space. This means that feelings in space act according to the mental law of association, directly acting upon feelings close to it in the protoplasm.

The principle of continuity states that we should as far as possible assume that things are continuous, hence we should assume that there is a continuity between the characters of mind and matter "so that matter would be nothing but mind that had such indurated habits as to cause it to act with a peculiarly high degree of mechanical regularity or routine (Peirce 1965

---

[94] See Peirce (1965 Vol. VI:164-169 [6.246-6.258]) for a summary of the supposed characteristics of protoplasm. Peirce argues that there are different protoplasms (e.g. 1965 Vol. VI:185 [6.278].

[95] See also Peirce (1965 Vol. VI:185 [6.280])

Vol. VI:184 [6.277]). In this instance, the interaction between mind and matter falls under the law of mental association. In order to construct his metaphysical system, Peirce attributes the traditionally mental property of feeling to matter, and the traditional material property of extension to mind. This theory is not materialistic given that mind is not governed by mechanical law, in fact, the "original law" becomes the law of mind – the law of association – and the laws of matter become special instances of this law. Peirce is aware of possible objections to his theory, but insists that it should be judged through comparing its consequences with observation, which he of course believes is a favourable comparison (184 [6.277]).

Peirce believed that the focus in the nineteenth century on economic questions has lead to a philosophy that exaggerates the benefits of individual greed, seeing it as a personal characteristic that will elevate the human race, and also as a major factor driving evolution in the universe. Similarly, sentiment was disparaged and marginalised, both in everyday life and in science (Peirce 1965 Vol. VI:192 [6.290]). This is one of his main objections to Darwin's evolutionary theory. He believes that Darwin merely extends a politico-economic conception of progress to the world of nature, and suggests that its favourable reception was as much due to the fact that the age was well disposed to "greed-philosophy" as to its scientific merit (196 [6.293]; 199 [6.297])[96]. Peirce, in turn, argues for what he calls an agapastic theory of evolution – or evolution by creative love (203 [6.302]).[97] Agapastic evolution embodies a teleology similar to that of Hegel (204 [6.305]). Peirce's agapastic teleology envisions a theory of Ideas akin to that of Plato – ideas are not merely representation or formulas for things in the world, but have a life of their own (Tiercelin 1998:3). Such ideas have causal power and seem to embody the character of a subject or agent. Peirce defines agapism as (1965 Vol. VI: 205 [6.307]):

> The adoption of certain mental tendencies … by an immediate attraction of the idea itself, whose nature is divined before the mind possesses it, by the power of sympathy, that is, by virtue of the continuity of mind.

Such a "sympathy of ideas" develops through evolution, although not a Darwinian evolution, but a Lamarckian one (Tiercelin 1998:3-4). This "evolutionary love" allows Peirce to introduce (a developmental) teleology into evolution, an aspect which he finds lacking in Darwin (Hausman 1998:629-630). Agape allows for spontaneity in the form of an agency to enter into evolution. And such spontaneity allows for the very laws of evolution themselves to evolve. Peirce's evolution is generalised beyond the temporal, so that evolution in time

---

[96] As will be discussed in Chapter 4, Darwin did not believe that evolutionary principles should be extrapolated to human affairs or characteristics. Hence, although his theory was appropriated by many into what has become known as "Social Darwinism", this development was by no means a necessary consequence of Darwin's ideas.

[97] It is interesting to note that for all his "physicistic prejudices" (Peirce 1965 Vol. VI:221 [6.322]), Peirce often calls upon Christian principles and the teachings of the gospels to confirm his metaphysical claims.

becomes only a special case of general evolution (Wells 1965:22). Hegel's system fails to allow for an element of freedom in the universe, which makes for a deterministic system. Peirce believes that his tychism[98] allows for that freedom and hence overcomes a mechanically deterministic conception of evolution (Wells 1965:23).[99] Agapism relies on the continuity of mind, and its purpose is the development of ideas.

As far as signs go, Peirce still holds that everything to do with Thirdness is essentially a sign, but he qualifies this statement by claiming that it would be more accurate to describe everything as being the soul of a sign (1965 Vol. VI:312 [6.455]):

> Such is everything which is essentially a Sign – not the mere body of a Sign, which is not essentially such, but, so to speak, the Sign's Soul, which has its being in its power of serving as intermediary between Object and Mind. Such too is the living consciousness, and such the life, the power of growth of a plant. Such is a living constitution – a daily newspaper, a great fortune, a social "movement".

As Gallie (1966:221) argues, it seems that Peirce's contention is that all living things, especially those with some form of intelligence or mind, are out of equilibrium with the universal processes around them – they have not yet "crystallised". They have not yet fulfilled their evolutionary capacity. "Pure" mental processes exhibit sign-like or semiotic behaviour which co-ordinates and controls phases of organic behaviour and renders such behaviour more continuous, or more accurately, more reasonable. All organic behaviour has the teleological character of developing towards Thirdness, or "concrete reasonableness" (221).

Thus, in order to overcome his objections to nominalism, retain the possibility of an existent and accessible objective truth, and resolve the matter of mind-matter dualism, without denying the existence of mental phenomena, Peirce develops what he calls an "objective idealism", where matter becomes effete mind, and entrenched habits become physical laws (Gallie 1966:170).[100] What started off as a philosophical doctrine dedicated to the hard-boiled empiricism of the natural sciences seems to have taken a decidedly esoteric metaphysical turn. And as his work progresses, we see that Peirce's cosmology takes on a more and more religious undertone and seems to become even further removed from Peirce's avowed pragmatism.

---

[98] Tychism is the doctrine, central to Peirce's evolutionary cosmology, that the fundamental laws of nature are statistical and mental. Furthermore, the laws governing matter are limiting cases of general mental laws where probability has almost reached certainty (Burks 1980:286).

[99] Peirce held that Hegel failed to allow for the irreducibility of the three categories, which Peirce believes is necessary to be able to account for individuality (Bernstein 1965:90).

[100] See also Liszka (1990:18).

9. PEIRCE'S (PAN-PSYCHIST) METAPHYSICS

It would seem especially true of sign interpretation that it cannot be thought of as anything other than a process that necessarily takes place within a mind/brain/nervous system. However, Colapietro (1987:226) argues that it is because of Peirce's desire to provide an alternative to nominalism that he calls upon "extra-mental relations of a higher order" where sign-activity takes place. To include the mind and interpretation as essential features in our understanding of the world, leaves open the possibility that the empirical world is our own construct. Allowing the world to possess a sort of Platonic, intrinsic meaningfulness, allows us the possibility at least, to objectively read the signs of nature.

As we have seen, Peirce maintains that reality is governed by the "law of mind". His commitment to an objective idealism leads to an ontology where everything that exists has the nature of a sign, and signs consist of the object-sign-interpretant trichotomy. Peirce's theory of signs *necessarily* involves a mental component.[101] We have also seen that Peirce does not restrict mental phenomena (where the interpretant occurs) to the brains/nervous systems of specific organisms. It seems to be something that permeates the universe, independent of matter. Peirce's ontology relies on a disembodied mind that exhibits sign-behaviour, supervenes on material or organic behaviour, and intervenes in its process of development, by rendering it "more reasonable" (Gallie 1966:221). This overtly anthropomorphic cosmology seems to contradict Peirce's explicit attempt to escape nominalism and adhere to his own Pragmatic principles. As Gallie (1966:179) states:

> [W]e cannot admit the suggestion that any of Peirce's metaphysical statements function as genuine hypotheses, possessing consequences that admit of empirical verification.

Peirce's metaphysics rests on phenomenology, and more specifically his three absolutely necessary and universal categories: Fist, Second, and Third.[102] Because of the phenomenological character of these categories, it is unclear how we could go about verifying them. One of the main aims of Peirce's metaphysics was to account for the very existence of natural laws, and to attempt to account for their nature. Peirce rejects the practise of attributing the laws of nature to a divine creator (e.g. Peirce 1965 Vol. VI:178-180 [6.273]), but it is difficult not to come to the conclusion that his own religious beliefs contribute to his cosmology quite significantly, especially when it comes to his conception of evolutionary love. Gallie (1966:240) points out that Weiner quotes Peirce as claiming that a contributing factor to

---

[101] Every genuine triadic relation involves thought or meaning (Peirce 1965 Vol. I:175 [1.345]).

[102] With regard to Peirce's three Universal Categories Gallie (1966:186) states:

> …whatever its obscurities and how ever dubious its claims to completeness, Peirce's categorical doctrine is an impressive statement of certain tendencies of thought which were certainly dominant in Peirce himself, and which, if he predicted rightly, points us to the 'general line of growth' of nineteenth-century science.

his cosmology is that it "satisfies his religious instinct" better than does the Newtonian world-view. While Gallie is of the opinion that Peirce had managed to separate his religion and his philosophy, Peirce's writing on metaphysics in particular suggests that his Christian faith did, in fact, influence his philosophy significantly.[103] His reference to a pragmatist conception of "God" as an "analogue" of a mind (Peirce 1965 Vol. VI:346 [6.502]) could very well account for his conception of his categories of being as being universal and teleological, even thought it seems to be in complete contradiction to his empirical bent. An example is the following extract from Peirce's "Reality of God" (1965 Vol. VI:329 [6.483]):

> It is that course of meditation upon the three universes that gives birth to the hypothesis and ultimately to the belief that they, or at any rate two of the three, have a Creator independent of them, that I have throughout this article called the N. A. [Neglected Argument], because I think the theologians ought to have recognized it as a line of thought reasonably productive of belief…In the mind of the metaphysician it will have a metaphysical tinge; but that seems to me rather to detract from its force than to add anything to it. [104]

And (Peirce 1965 Vol. VI:333 [6.487]):

> Were the theologians able to perceive the force of this argument, they would make it such a presentation of universal human nature as to show that a latent tendency to believe in God is a fundamental ingredient of the soul, and that, far from being a superstitious ingredient, it is simply the natural precipitate of meditation upon the origin of the Three Universes.

Gallie (1966:227) puts Peirce's apparent contradiction of his own project down to the fact that he was extremely taken with the use made of the concept of chance in Darwin's theory of natural selection, and the kinetic theory of gasses.[105] Peirce constructed his metaphysics in terms of his firm belief that evolution "must eventually restore the rejected idea of *reasonableness* mechanising the world…" (234). [106] Peirce attempted to incorporate the

---

[103] See Melvil (1966) who argues that Peirce was a religious man "until the end of his days", which caused a constant tension or duality between his scientific and his religious affiliations to run through his work.

[104] See "A Neglected Argument for the Reality of God" (Peirce 1965 Vol. VI:311-339 [6.452-6.493]) and "Answers to Questions Concerning my Belief in God" (340-355 [6.494-6.521]).

> For Peirce, a conception of God is a natural result of the scientific project (See Delaney 1992), and his religious thinking, although it has often been ignored or considered peripheral is central to his doctrine as a whole (see Raposa 1991).

[105] Cf. Peirce (1965 Vol. V:226 [5.364]). These two scientific developments would have a profound influence over the thought of most theorists over the next century, and more.

[106] Cohen (1998), minimises the influence that Darwin had on Peirce (1998:XXXIV), but as we have seen Peirce incorporated Darwinism into a broader evolutionary theory, and the influence that this broader conception of evolution had on his thinking is crucial to understanding Peirce's metaphysics.

ideas of chance and probability into his later work, with the very ambitious intention of applying these principles to the universe as a system, and deducing a single evolutionary law that governs all regular processes, and he attempts to explain how relatively closed systems, develop from previously relatively open systems. As such, Gallie sees Peirce's Cosmogenic principle as "a pertinent, if very imperfectly formulated, criticism" of scientists' "linear conception of the development of the physical world" (231). Peirce holds that science made an indefensible assumption in holding reversible, mechanical processes to be the paradigms of lawfulness and intelligibility. As a result, Peirce does not see the material universe as completely determined by the laws of dynamics and presents his own cosmology as an alternative to the Newtonian world-view. Peirce identifies other possible laws at work in the universe; laws that may have as much, if not more, influence on how the universe unfolds.[107]

In his attempt to ensure the possibility of objectivity and absolute truth, Peirce tries to remove any form of finite mind from his process of semiosis.[108] But he seems unable to identify any plausible entity to replace finite minds. In terms of his religious philosophy Peirce overcomes this difficulty by suggesting that an Absolute Mind could perhaps be "a Deity relative to us" (Peirce 1965 Vol. V:69 [5.107]). But Peirce remains tentative and unconvinced in his own attempts to equate Mind with God (Raposa 1991:344). According to Raposa (344), Peirce did regard the universal Mind as "a symbol of God's thought and purpose".[109] With this metaphysic, Peirce undermines the very structure necessary to support his semiotic and hence his ontology. Unless, of course, we concede the existence of a disembodied, metaphysical mind, which, seems an absurd and unnecessary concession.

10. CONCLUSION

Peirce is not wrong in wanting to emphasise that signs are not restricted to the human realm. There are many species with the competency (however basic) to make use of signs. The argument here, however, is that signs are not inherently meaningful and can only be understood in terms of *interpretation*. As was discussed, Peirce came to the same conclusion in his analysis of the possible referential modes of signs. In his attempt to formulate an account of sign behaviour, without recourse to an essentialism that attributes a "meaning essence" to individual signs, Peirce developed a highly plausible theory of sign meaning; one which foreshadows and surpasses that of Saussure. He also anticipates the

---

[107] Peirce's cosmology is an attempt in his later years to unify his philosophical projects into a coherent philosophical system. Gallie (1966:216) poetically describes Peirce's cosmology as an attempt at "totally reorientating our ordinary conceptions of the origin and the destiny of the universe, of the nature of general laws and their relation to things or events which conform to them, and at drastically revising – if not obliterating – the familiar distinction between living and non-living, mental and non-mental processes".

[108] See also Greenlee (1976:140), where he describes Peirce's position as "epistemological realism" and explains that, in this view, "the object of perception and of knowledge is directly known, not known indirectly by means of an intervening mental entity…"

[109] Raposa (1991:347) holds that it is religious considerations that fuelled Peirce's criticism of Darwinism.

twentieth century preoccupation with meaning and language as it related to epistemology. Language/signs are of primary importance, in that they mediate our knowledge of the world. Peircean semiosis has influenced many theorists, and some instances of his influence will be examined in the coming chapters.

However, as has also become apparent Peirce undermines much of the "scientificity" of his project by his adherence to certain philosophical assumptions and in some areas his project takes on the character of highly speculative metaphysical conjecture. In spite of his discomfort with Cartesian dualism, Peirce seems unable to overcome the mind-body distinction as framed by Descartes. His common-sense conception of the physical constitution of the universe – which he sees as being governed by mechanical laws – and his common-sense conception of the mind – which cannot be subject to mechanical laws if it is to exhibit free-will – are incompatible. This incompatibility leads him to postulate a universe that literally consists of extended mind. And even though this move enables him to overcome some of the tensions in his common-sense conception of the difference between body and mind, his conjecture is unverifiable and hence unscientific. Peirce resorts to the very metaphysics which he sought to overcome.

Peirce is not alone in his philosophical allegiances. Many philosophical projects run up against their proponents' common-sensical presuppositions about aspects of the world. Such presuppositions are fundamentally tied to a given theoretical paradigm and are very difficult to perceive, let alone overcome. In the coming chapters we will discuss attempts at overcoming some of the inherited philosophical certainties with regard to body and mind in order to surmount certain obstacles that these certainties entail. As will become clear, this is no easy task and, as Kuhn (1970) shows, the necessity of a given paradigm with given assumptions can never be overcome; a sentiment apparently shared by Peirce (1965 Vol. V:374 [5.535]):

> I hold…that man is so completely hemmed in by the bounds of his possible practical experience, his mind is so restricted to being the instrument of his needs that he cannot, in the least, mean anything that transcends those limits.

Peirce was to assert a substantial, if low-key, influence on theory in the twentieth century, especially with regard to its movement towards a semiotic understanding of various phenomena. Whereas theory in the nineteenth century was dominated by the concepts of energy and thermodynamics, theory in the twentieth century would be dominated by the concepts of information and code. Hence, while Peirce could be considered idiosyncratic and even bafflingly metaphysical in the context of his own milieu, his work can be seen as heralding a new paradigm. One of the most influential theorists in this paradigm with regard to language and mind was to be Noam Chomsky. As the semiological turn took root in the first half of the twentieth century it would become clear that language and mind would be two of the fundamental objects of study in this new paradigm. With the publication of *Syntactic*

Structures in 1957, it would become clear that no new work could be produced within the paradigm without taking Chomsky's work into account. And, as will be seen in the following chapter, one of his fundamental influences was the work of Charles Sanders Peirce.

CHAPTER 2

CHOMSKY: INNOVATOR AND SCEPTIC

The proper way to exorcise the ghost in the machine is to determine the structure of the mind and its products (Chomsky 1975:23).

Work of the past few years suggests further, and if successful, possibly quite far-reaching revisions of the general picture of language (Chomsky: 1994:161).

[L]anguage has, for the better or the worse, been central to philosophy this century (Otero 1994a:7).

1. INTRODUCTION

Few theorists have had the far-reaching and comprehensive influence on a discipline (or, more accurately, a range of disciplines) that Noam Chomsky had (and still has). And whether one profoundly agrees or disagrees with him, one cannot ignore his contributions to the subject.[1] In fact, Carlos Otero (1994a:1) describes him as "the philosopher who is to the period initiated by the cognitive revolution of the mid-1950's what Descartes was to the first phase of the age of modern philosophy." Those familiar with Chomsky's work will know that the parallel between Chomsky and Descartes is not arbitrary or undue, given that Chomsky did much to revive some of Descartes' and his progeny's ideas on language and the mind in his quest to bring the mind into the sphere of objects amenable to scientific investigation.

Chomsky's work spans the length of an extraordinarily productive 40-year career, in which he has revised and rewritten many aspects of his theories. As such, it is all but impossible to present a comprehensive and coherent account of his theories in the space of one chapter. This chapter will concentrate on those fundamental aspects of his theories that have remained relatively consistent throughout the permutations of his revisions. Generally, the elements in Chomsky's work that have remained constant have to do with his conception of the role that language can play in studying the mind, and, notably, on what he believes

[1] Otero (1994a:2) notes that Chomsky made important, and perhaps unsurpassed, contributions to four areas of philosophy, which represent the most important focus areas in contemporary philosophy, namely: 1) philosophy of language, and philosophy of mind and cognition; 2) epistemology and philosophy of science; 3) ethical theory and social theory; 4) history of philosophy.

language can tell us about the structure of the mind/brain. Chomsky's theories represent a turning point in the 20th century's approach to language and mind, which left almost no aspect of disciplines associated with these phenomena untouched.[2] As he himself points out on numerous occasions, he does not view his contribution to be revolutionary, but rather a revival of a tradition which holds that some of the classical questions of philosophy should be addressed through science.[3] Thus, the Chomskyan and Peircean projects have much in common. The question is whether Chomsky manages to succeed where Peirce fails.

2. THE NATURALISTIC FRAMEWORK

Chomsky (1994:153) holds that the study of mind is an enquiry into certain aspects of the natural world. For him, facets of such an enquiry that focus on those phenomena that are traditionally considered to be "mental" events, processes, or states are amenable to a "naturalistic approach" or "methodological naturalism" and should therefore be studied in a way similar to any other object that falls within the scope of scientific study.[4] Accordingly, when he refers to the study of mind, he uses the term to refer to a domain of enquiry with "no metaphysical connotation" (155), where "mind" means something like "the appropriate level of abstraction about properties of some physical systems of the brain" (Chomsky 2004:60).[5]

---

[2] Various theorists have referred to the "Chomskyan revolution" in linguistics (e.g. Searle 1994:68-94 and Newmeyer 1994:918). As we shall see, Chomsky's own view is that it is only with his revisions in the 1980's that his work can be seen as breaking away from a well-established linguistic tradition.

[3] As Otero (1994a:7ff.) points out, this scientific approach to philosophical questions was temporarily overthrown by Frege's influence on epistemology. He, and other positivists, held up a "philosophy of language", as we have come to know it in the twentieth century, in the place of an earlier philosophical realism. Frege's methodological stance inspired the move in philosophy towards defining epistemological issues in terms of questions of logic and conceptual analysis as a generalisation of his analysis of mathematical proofs. For the positivists, the study of linguistic meaning was the proper starting point of philosophy (9). A subsequent philosophical approach to language was to give priority to ordinary judgements and practices over philosophical principles, so-called ordinary language philosophy, as characterised by Moore, Austin, and the later Wittgenstein. Otero notes that this approach can properly be regarded as language-studies, because it looks to language for answers to specific questions (9). (See Katz and Fodor (1994:35) for a comprehensive discussion of their view that both positivism and ordinary-language philosophy are mistaken in their basic assumptions about language, and the appropriate techniques for analysing language).

Elsewhere (1994a:13) Otero describes Chomsky's work as " the successful merging of two distinct lines of enquiry: the venerable tradition of the study of language, which goes back over 2,000 years; and the century-old tradition…initiated by the founder of modern logic, Gottlob Frege".

[4] Chomsky (2000b:115) defines a "naturalistic inquiry" as: "a particular human enterprise that seeks a special kind of understanding, attainable for humans in some few domains when problems can be simplified enough. Meanwhile, we live our lives, facing as best we can problems of radically different kinds, far too rich in character for us to hope to be able to discern explanatory principles of any depth, if these even exist". Furthermore, he argues that almost all approaches to studying language and mind involve unjustified non-naturalist assumptions. He argues that this is due to our profoundly dualistic commonsense understanding of the world, which makes these ideas so compelling (163).

[5] Elsewhere, he adopts a formulation of Anthony Kenny's and holds that "the mind of O" is "the innate capacity of O to construct cognitive structures, that is, to learn" (Chomsky 1975:22). For him mind is the capacity to form cognitive structures, rather than first-order capacities to act. In terms of modern analytic philosophy, Chomsky argues that there has been a tendency to use terminology such as "disposition" or "capacity" where he believes "cognitive structure" would be a more appropriate term.

Chomsky (2004:60) argues that "mental talk" is not inherently dualistic, but he does not summarily discount the possibility of dualism. He does not explicitly commit to dualism, or monism. He argues that,

Chomsky sees language as a "species property", which is central to human thought and understanding. He believes language as a system is relatively well circumscribed and open to inquiry and study – features which may allow us to make some significant inroads into studying human cognition, *if* we can formulate a convincing theory of language. He does not, however, equate the study of language with the study of mind, as many philosophers do (Chomsky 1977:36; 1980:4). His is a nativist approach to language, in that he holds that human beings are born with an innate capacity for natural language, which only needs to be triggered and allowed to develop. Chomsky (2000a:11) argues that there is no coherent mind-body question and that questions of representation or intentionality are radically misconceived. Chomsky (2000b:109-110) argues that contemporary theories that cast certain philosophical questions in mind-body terms have failed to take account of the collapse of this Cartesian distinction with Newton's postulation (and science's acceptance) of gravity. This led to the collapse of a mechanistic understanding of the universe. In a milieu where science invokes "fields of force, curved space, infinite one-dimensional strings in ten-dimensional space" etc., it becomes all but impossible to equate the "physical" or the "material" to that which is subject to "mechanical causes" . In other words, "the discussion presupposes some antecedent understanding of what is physical or material, what are the physical entities", which does not make sense outside of a mechanical philosophy (109). Hence, the mind-body question can only be raised in its Cartesian form on the grounds of unjustified dualistic assumptions.

Throughout his extensive body of work, Chomsky (e.g. 1966:ix; 1972:ix, 66, 100) holds that since the everyday use of language always involves (human) intellectual abilities of the highest order, it is only natural to suppose that the study of language contributes significantly to our understanding of the nature of the human mind and its functioning. By studying language, we may discover abstract principles that govern the structure and use of language in general. Such principles would not be accidental, but would be universal by biological necessity in that they derive from the mental characteristics of the species (1975:4). Essentially Chomsky's (1975:5; 1977:18) is an attempt to address the classical epistemological question: How do human beings know as much as they do, given their limited contact with the world? He often refers to this issue as "the poverty of stimulus" problem or alternatively "Plato's problem" (cf. Chomsky 1986:xxvff). "The problem is to account for the specificity and richness of the cognitive systems that arise in the individual on the basis of the limited information available."[6] As we shall see, his methodology will be to shift the main

---

contrary to what is often assumed, the concept of the body is not understood in any great depth and the notion of "the body" is continually changing and expanding. He holds open the possibility that the apparent problems of mind might be solved by reducing mind to brain, or by "expanding brain to mind". In order to avoid some of the "entanglements" inherent in dualism, it might be necessary to change our conception of the body.

[6] He characterises the cognitive systems as resulting from the interaction of experience and the organism's method of constructing and dealing with it, which would include analytic mechanisms and determinants of maturation and cognitive growth (Chomsky 1986:xxv). Chomsky (2000b:7) often likens the development of a child's language to *growth*: "it is something that happens to a child, not that the child does."

burden of explanation from the world to the mind, much in the manner of Kant's categories (Chomsky 1975:6-7).

Because human language is so complex, Chomsky argues that it would be "an extraordinary intellectual achievement" for a creature to learn language, if it were not specifically designed to be able to do so. Yet, practically all human children accomplish this task with seemingly minimal effort. All children seem to acquire a language (or languages) within equivalent (and relatively short) periods of time, proceeding through various stages in the same order and at a comparable rate. Chomsky concludes that there is no non-innate theory of language that can account for this phenomenon, and believes that empirical theories of learning or training are at a particular disadvantage in accounting for this aspect of language-learning. He suggests (1972:ix-x) that language is a special case of learning, in that the acquisition of it might be largely predetermined, and as such, he suspects that it might reflect the characteristics of human intellectual capacities.[7] In studying the structure, organisation and use of language, he hopes to gain some insight into specific characteristics of human intelligence; into the distinctive human cognitive capacity that enables the species to acquire and use language, and raise the possibility that it might even be possible to extend these insights to other aspects of human cognition.

Chomsky (1972:ix) further believes that the problem of accounting for the normal, everyday use of language is beyond the scope of scientific enquiry, although such enquiry might be able to address this problem in the future. He is extremely critical of the behavioural sciences (and of modern linguistics), accusing them of employing arbitrary methodological restrictions that yield scientific knowledge of a trivial character. Furthermore, he believes that analytic philosophy could benefit from incorporating certain aspects of linguistic theory (161). He qualifies this view by stating that we should not expect too much from the interchange. One of the reasons he gives for his cautionary stance is the fact that "neither field makes use of research techniques of a sophisticated or specialized nature", which limits the probability that the results of one field can be built on by the other. The achievements of both fields owe very little to science and modern technology; the gathering of data is informal, very little experimental work is done (if any) and theory is most often based on "intellectual constructions". Chomsky does not see these characteristics as necessarily negative, he views the intellectual constructions concerned as illuminating and nontrivial, but he questions whether either field can lay claim to the designation "scientific" (Chomsky 1972:165). He thinks linguistics can contribute to traditional philosophical questions in producing theories relevant to the problem of how knowledge is acquired, and how the character of knowledge is determined by certain properties of the mind (168).

In Chomsky's work, *language* is always understood to mean *human* language, and is characterised by the belief that it is impossible to generalise between the many possible

[7] According to Chomsky (2000a:51), to say language is innate is "to express the belief that some crucial and relevant internal nature differentiates my granddaughter from rocks, bees, cats and chimpanzees."

manifestations of language (e.g. animal, formal, etc.).[8] Furthermore, the defining characteristic of language is not communication (there are other, non-language systems of communication), but its ability to express human thought (e.g. Chomsky 1966:21).

Chomsky argues that all organisms have a certain *Umwelt* or way of perceiving the world. Such organism-specific cognitive spaces are largely determined by the specific nature of particular organisms, biological properties in general, and by specific interactions with the environment. Given an organism's particular cognitive system (and environmental niche), it becomes possible to identify a category of "problem situations" particular to that organism, which could be deduced from the organism's nature and prior history (Chomsky 1994:155-156). Some such problem situations fall within the organism's cognitive capacities and others do not. In other words, some problem situations will remain completely impenetrable to a particular organism, because of the limits in its cognitive capacities. Chomsky holds (1972:67, 100) that human language is a unique phenomenon that depends on unique cognitive capacities that are without an analogue in the animal world, and not something that has evolved out of systems of animal communication.[9]

He extends this argument to humans, arguing that there are very likely to be problem situations that will by their very nature remain forever beyond our cognitive abilities. In this view, where we do happen to solve some of the mysteries of the world, by means of the natural sciences, for example, it is probably due to a chance "convergence" between properties of the natural world and properties of the human mind (Chomsky 1994:156; 1975:137-228). Hence, we are able to solve such problems because they lend themselves to being understood in terms of our cognitive orientation and, we very probably solve them in a manner unique to human cognitive tendencies. As will be discussed, Chomsky draws on Peirce's theory of abduction to formulate his idea of limited cognitive capacities.[10] However, he explicitly differs from Peirce's proposal that evolutionary factors will ensure that we ultimately will discover the Truth about the world (Chomsky 1994:156; 1972:17, 90).

Chomsky rejects metaphysical dualism, which argues that the "mechanistic" character of the physical world would not allow for the free and creative aspects of language use, and hence postulates that there must exist, a wholly different substance, which accounts for the "mental"[11] properties that are found in the world.[12] Similarly, he rejects "metaphysical

---

[8] See Chomsky's *Cartesian Linguistics* (1966:1-30) for the similarities between this view and that of Descartes, among others.

[9] See Chomsky (1972:67) where he criticises Karl Popper's (1966) contention that language evolution passed though several stages, from primitive systems of communication to a sophisticated tool that can be used for description and critical argument. Chomsky argues that human languages are not simply more complicated versions of animal communication systems, but is subject to an entirely different principle of organisation, supported by a specific type of mental organisation (70). In this sense, language is an instance of "emergence", defined as "the appearance of a qualitatively different phenomenon at a specific stage of complexity of organisation" (70).

[10] See Chapter 1.

[11] Chomsky uses the concept of "mentalistic terms" fairly extensively and quite loosely. He does not hold that "mental" phenomena are by definition precluded from being realised physically (see, for example, 1972:98-99). When he speaks about mental phenomena he is usually referring to the mechanisms that underlie behaviour or the acquisition of knowledge. In his words (Chomsky 1972:98):

naturalism".[13] He argues that the mind/body distinction cannot be formulated in any manner other than as a terminological device to "distinguish various aspects of the world" (1994:157).[14] If "mental" properties, unique to humans, do exist, he argues, it is a fact about the world that should be understood, as with any other fact about the word, in naturalistic terms.[15] Chomsky holds, however, that although mind and matter are not different categories of things, they pose very different challenges to our cognitive orientation. Hence, it might be that there are some aspects of mental phenomena that are cognitively inaccessible to us, because they require a cognitive orientation that is entirely out of our reach due to our natural mental configuration. He names the creative aspect of language use as an example (157). He deems it entirely plausible that we may never understand what it is that makes it possible for human beings to use language “as an instrument for the free expression of thought and feeling”, or what qualities of mind make the creative use of language, which an essential feature of human existence, possible (Chomsky 1972:101).

In light of his criticism of contemporary empiricist methodology employed by behavioural scientists, language philosophers, and linguists in studying language, Chomsky

---

> It is an interesting question whether the functioning and evolution of human mentality can be accommodated within the framework of physical explanation, as presently conceived, or whether there are new principles, now unknown, that must be invoked, perhaps principles that emerge only at higher levels of organisation than can now be submitted to physical investigation. We can, however, be fairly sure that there will be a physical explanation for the phenomena in question, if they can be explained at all, for an uninteresting terminological reason, namely that the concept of "physical explanation" will no doubt be extended to incorporate whatever is discovered in this domain, exactly as it was extended to accommodate gravitational and electromagnetic force, massless particles, and numerous other entities and processes that would have offended the common-senseof earlier generations. But it seems clear that this issue need not delay the study of the topics that are now open to investigation, and it seems futile to speculate about matters so remote from present understanding.

[12] The reader will recall that the problem of free will scuttled Peirce’s attempt to overcome mind-body dualism.

[13] Chomsky (1994:157) defines this position as holding that "the study of mind must be ‘continuous’ with the physical".

[14] See Chomsky’s "Naturalism and Dualism" (2000b) for an extensive discussion on various dualistic elements in theories of language and mind, where he argues that such dualistic assumptions are adopted far too casually and should either be justified or abandoned.

[15] It is interesting to note that Chomsky does not treat "Reason" as such a cognitive or mental capacity. For him, "reason" is not a mental mechanism and it does not have fixed operations (McGilvray 1999:36, 88). Here, reason is a human attribute that is enabled by our common-sense understanding and our science-forming capacity. His position is akin to that which Toulmin describes as an Aristotelian conception of reason, where reason is context-dependent and somewhat variable (Toulmin 1990).

As we shall see Chomsky (1995a:43) approaches "reference" in a similar manner. He does not believe reference in natural language to be amenable to scientific enquiry. He argues that philosophers often treat reference as being simpler than it is; understanding it as a dyadic relationship between a word and its object. He also disagrees with Peirce’s conception of triadic reference, which he calls an invented, technical notion of reference. Chomsky believes that reference is a tetradic relationship between a word, person, circumstances and an object. In his words (Chomsky 2000b:150):

> More generally, person X uses expression E with its intrinsic semantic properties to talk about the world from certain intricate perspectives, focussing attention on particular aspects of it, under circumstances C, with the "locality of content" they induce".

Reason and reference are normatively governed in a way that visual processing or linguistic processes are not, which is why science cannot deal with them (McGilvray 1999:38).

revisits earlier work done in the field. For him, the idea that the study of language can contribute to our understanding of human nature threads its way through modern western thought (Chomsky 1972:1). Since the Enlightenment, classical problems of the relation between language and mind have reappeared in various guises and within various disciplines; the most recent appearance having been in the cybernetic and communication sciences[16], although, such contemporary disciplines actively disassociate themselves from earlier (modern) movements within linguistic theory (Chomsky 1966:1). Chomsky (1975:12) argues that the grip of the empiricist doctrine on the modern period (outside of the natural sciences) can be explained on historical and sociological grounds. He believes that the attraction of the empiricist doctrine has long been the perception that it is "progressive", with undetermined human nature leaving open the possibility of social change, progress, and enlightenment, for example (Chomsky 1975:129ff).[17] He feels strongly, however, that adopting a "rationalist" or naturalistic approach to mind does not condemn human nature to determinism, and even argues that in many respects his approach is less restrictive than a "blank slate" conception of mind (134).

3. A CARTESIAN[18] – AND NEWTONIAN – LEGACY

Chomsky (1972:5) notes that in many respects, contemporary intellectual preoccupations mirror those of the seventeenth century. Not least of these is a fascination with automata, one of the most famous examples being the work done by Descartes. Descartes could conceive of human beings (and animals) as being nothing more than automata, operating according to a lever-and-pulley system similar to that of the automata then *en vogue* and exhibited in the French Royal Gardens (Flanagan 1991:1). However, he could not conceive of a way to account for understanding will, or the diversity of human behaviour and its appropriateness to novel situations in such mechanical terms (Chomsky 1966:6; 1972:10-11).[19] Chomsky argues that at the heart of Cartesian philosophy there lies the realisation that there exists a significant gap between those things that we can know with relative certainty, and those that seem unfathomable to human capabilities. This realisation is

---

[16] More on this in Chapter 4.

[17] See Steven Pinker's *The Blank Slate* (2002) for a comprehensive (and mostly convincing) discussion of the historical and socio-biological factors that many feel have contributed to the hold that the doctrine of the human mind as a blank slate has had on the social sciences. Like Chomsky, he argues that the main attraction of the doctrine seems to be its association with a progressive political stance, rather than its scientific validity.

[18] Chomsky (1966:2-3) applies this term to the prominent linguistic movements within the seventeenth, eighteenth, and early nineteenth centuries. This coherent body of ideas developed out of the Cartesian revolution. He contrasts the Cartesian approach to a doctrine of assumptions which he terms "empiricist linguistics" (76, footnote 3).

[19] Descartes addresses the problem of other minds by proposing that having a mind entails having a generative (or creative) ability in the use of language, which goes beyond the bounds of mechanical explanation (Chomsky 1972:11).

what led Descartes to conclude that the mind is something altogether unlike matter (Chomsky 1972:6).

Chomsky goes on to trace theories of mind and language that developed out of Descartes' findings (e.g. Chomsky 1966; 1972:6-23) and it becomes clear that a common element among these theories was the conviction that hypotheses appropriate to human physiology are incapable of accounting for "mental" phenomena.[20] Phenomena which are obvious to us upon introspection (from a first-person point of view) or obvious in our observation of other people seem to be resistant to objective study.

Descartes worked within a scientific paradigm which envisaged the world as operating in terms of a kind of "contact mechanics" (Chomsky 1994:156ff). The normal properties of language use, however, could not be explained within this paradigm. Language is unbounded, it is not determined by external stimuli or an internal state, language use is usually coherent and appropriate to a given situation, but cannot be said to be caused by those situations in a mechanical way. The Cartesian linguists referred to these properties of language as the "creative aspect of language use" (156).[21] The creative aspect of language use led Descartes to draw the conclusion that a "creative substance" must exist alongside the mechanical phenomena in the universe (Chomsky 1966:8). From here a new principle was devised, one which would allow for the second, "thinking" substance that was qualitatively different from matter and had thought as its essence, and extension and motion as properties – the principle of the mental. Of course, if "thought" is not subject to the mechanistic natural laws of the universe, it becomes necessary to explain the nature of the mental. Furthermore, this tactical move resulted in the "unification problem", the problem of explaining how mind and body interact. Such questions, raised by postulating the existence of two distinct substances in the universe, would dominate much of philosophy for the following three centuries.[22]

A generation after Descartes, Newton argued that contact mechanics was fundamentally inadequate in that it could not account for the basic properties of the motion of bodies (Chomsky 1994:156). He postulated the existence of an attracting force to account for the motion of bodies, namely gravity (Chomsky 1972:7). Newton's speculation that the terrestrial and planetary motions are not subject to "mechanical laws", as they were understood to be in mechanical philosophy, caused the assumption that the mental is the only

---

[20] The period of Cartesian Linguistics, as defined by Chomsky (1966:29), is characterised by identifying linguistic and mental processes.

[21] In a similar vein, Rorty (1996) holds that the Enlightenment cleared the way for "the exultation in human beings' ability to use language in new, surprising, ecstatic ways", a position which he terms "romanticism". The Cartesian fascination with the creative use of language should not be understood in a restrictive, aesthetic sense, however. Every sentence that a person utters is a new construction, in that a person is unlikely to have heard or uttered that exact sentence, in the same context before. "Creativity" here refers to the ability to construct a new, meaningful sentence which is appropriate to relatively novel circumstances, virtually all the time and seemingly effortlessly.

[22] Chomsky (1994:156; 2000b:108) argues that this metaphysical dualism was "basically naturalistic" in that it used empirical evidence to formulate factual theses about the world, even if these theses turned out to be wrong.

substance not subject to mechanical laws to collapse (Chomsky 2000b:108). However, Newton's debunking of the mechanical view of the universe was not generally accepted. Chomsky notes that Newton's concept of gravity was severely criticised by the scientists of his day, including Leibniz and Huygens, as relying on "spiritual forces" and relapsing into medieval ways of thought, because his theory allowed for interaction "without direct contact". The idea of a force acting at a distance clashed with common-senseideas about how the world works, and Chomsky speculates that the orthodox Cartesian would have viewed such a force as an occult quality such as those envisioned by the scholastics (Chomsky 1972:7-8). It is interesting to note that an attractive force did not sit well with Newton's own sensibilities either and he tried to find an explanation for the cause of gravity on mechanical grounds (Chomsky 1972:8; 2000b:109). Of course, despite his qualms, Newton's "anti-materialist" theory became scientific common sense.[23]

Chomsky (2000b:109) points out that contemporary theorists of mind commonly assume that Newton "remained within a materialist world picture," and then argue that contemporary attempts at "reducing" mind to matter ("naturalism", materialism", "physicalism") oversteps these bounds.[24] What these objections fail to take into account is that is not a foregone conclusion what "a materialist world picture" would look like. Newton and his contemporaries did not view his world picture as materialist, and we can only understand him as being "materialist" if we assume that materialism is "whatever science constructs" (109). Chomsky argues that the distinction between the physical and material, and the mental rests on an antecedent assumption of what the physical is (2000b:110):

> These terms had some sense within the mechanical philosophy, but what do they mean in a world based on Newton's mysterious force, or still more mysterious notions of fields of force, curved space, infinite one-dimensional strings in ten-dimensional space or whatever science concocts tomorrow? Lacking a concept of "matter" or "body" or "the physical," we have no coherent way to formulate issues related to the "mind-body" problem." These were real problems of science in the days of mechanical philosophy. Since its demise the sciences postulate whatever finds a place in intelligible explanatory theory, however that may be to common-sense. Only on unjustified dualistic assumptions can such qualms be raised specifically about the domain of the mental, not other aspects of the world.

---

[23] This leads Chomsky (2000b:112) to argue that, theoretically, there is much to be gained from abandoning common-sensefoundations and seeking "best theories". As the history of physics and chemistry has shown, "best theories" often do not have independent standards for evaluation other than their contribution to understanding. It should of course be kept in mind that many of our scientific hypotheses are based on "best-theory" constructs and, as such, remain open to revision and refutation.

[24] Thomas Nagel (1982), for example, holds this view, on the grounds that our knowledge of mental phenomena is necessarily based on a first-person perspective, and hence cannot be objectively verified in the way that "physical" knowledge can.

The post-Newtonian "anti-materialist" world-view rejected mechanical philosophy and with it materialism as it is envisioned in a common-sense understanding of the world. Some theorists concluded that nature would remain unknowable, while others argued that theories about the world should only be given an operationalist interpretation (Chomsky 2000b:110). Chomsky argues that the post-Newtonian world-view allowed for a conception of science as an attempt to postulate "best-theories", which are evaluated in terms of their explanatory possibilities, and not in terms of whether they appear counterintuitive or not. He argues that it has become commonplace in physics to accept counterintuitive and "weird" principles because of their ability to explain certain aspects of the world.[25] He contrasts this state of affairs in physics with that in some disciplines within philosophy where recourse to "right-thinking" and "common sense" can still be used to foil a philosophical theory (112). Chomsky's point is that this has been the case in much of the theoretical work done on language and mind, where recourse to "common sense" has inhibited theoretical progress.

Newton's demolition of the common-sense understanding of the body led many subsequent theorists to conclude that both human thought and action were properties of organised matter, because, if matter is compatible with forces of attraction, it could just as well be compatible with "sensation and thought" (Chomsky 1994:157; 2000b:113), and this assumption is not unique to twentieth century theory.[26] This move became possible because the post-Newtonian world view allowed for hypotheses that appear counterintuitive to the common-senseworld-view. Newton has made it possible to alter our conception of matter to include both the "mental" and the "physical".[27]

The question remains as to what the naturalistic properties of things in the world are and how they arise. Chomsky argues that a new version of the unification problem arose,

---

[25] To quote Chomsky:

> The domain of the "physical" is what we come more or less to understand, and hope to assimilate to the core natural sciences in some way, perhaps modifying them as enquiry proceeds. Ideas that yield understanding and insight are judged legitimate, part of the presumed truth about the world; our criteria of rationality and intelligibility may also change and develop, as understanding grows. If humans have "ghostly properties" apart from those common to all of matter, that's a fact about the world which we must try to comprehend in naturalistic terms, there being no other (Chomsky 1994:157).

[26] Chomsky disagrees with Nagel's (1993) assessment that John Searle's (1992) claim – assuming that consciousness is a physical property of the brain – is "radical". He (2000b:86) argues that Searle's position is (and was previously) a natural reaction to Newton's "demolition" of mechanical philosophy, and "hence of the mind-body problem". He shows that eighteenth-century figures already contended that "thought and language are properties of organised matter – in this case, mostly the brain, not the kidney or the foot" (115 ff). Chomsky sees such contentions as "close to truism" and just as uninformative, since the brain sciences are far from accounting for the "mental" aspects of the brain. He argues that we currently have "good and improving" theories of language and mind, but only "rudimentary ideas" of how they can be related to the brain. A naturalistic approach to language and mind will attempt to improve both theoretical domains in the hope of unifying them in a meaningful way.

[27] Chomsky (2000b:113-114) argues that common-sense conceptions of phenomena are responsible for many pseudo-problems within philosophy, such attributing intentionality to parts of people and to other objects, because we can generally (and correctly) attribute intentionality to a person as a whole. He sees this move as underlying debates about whether machines can think or whether an artefact or algorithm can translate Chinese. He believes, in accordance with Turing (1950a) that such questions are meaningless; they are not questions of fact, but matters of decision, based on metaphorical usage of terminology.

namely, "How can organised matter have properties?" He also believes that very little progress has been made on this problem (Chomsky 1994:157). Even though mind and matter are not different categories of things, Chomsky believes that they pose different challenges to human intelligence. He speculates that what we broadly term the "mental" may be cognitively inaccessible to us. He argues that in the past generation, other aspects of language have been successfully studied and many of the traditional questions have been recast, which has led to new understanding of some of the aspects of the processes of the mind. Advances in the sciences provided the concept of generative (recursive) procedure, which provided new formal insights which made the pursuit of such traditional problems possible.[28] However, there are many aspects of language and the mental which we do not understand any more than the Cartesians did, which leads Chomsky to another contemporary preoccupation which mirrors that of the Cartesian Linguists, namely the creative aspect of language use.

As we have seen, the Cartesian linguistic tradition holds that any study of human language must account for the "creative aspect" of language use (Chomsky 1994:156; 1972:6, 11; 1966:8-12). By this is meant the ability to use an apparently finite means in an innovative manner to generate a potentially infinite range of new expressions. The creative aspect of language use cannot be controlled by stimuli, whether internal or external. Human language is taken by the Cartesian Linguists to serve as a medium for human thought and self-expression, over and above any communicative function it may have, which means that it cannot be determined in any way (Chomsky 1966:29). A further characteristic of the creative use of language is that it is appropriate to a situation, even if the situation is entirely novel (which is usually the case) and coherent. Chomsky argues that these aspects of human language seem to go beyond the bounds of physical explanation. It is not his contention, however, that the Cartesians managed to propose satisfactory answers to these problems by attributing them to a principle called "mind" (Chomsky 1972:13). [29]

Furthermore, the linguist should be able to discover how structures that underlie the language ability came into existence in the human mind. Descartes argued that the study of mind raises the problem of the quality of its complexity, rather than its degree of complexity. "Mind" seems to be something very different from, and unaccountable for in terms of, the physical world as we understand it. This enabled Descartes to argue that the presence of a human mind is indicated by an entity's ability to use language creatively, which in turn is the result of a second substance which has a "creative aspect" as one of its basic properties (Chomsky 1972:6-7).[30] The laws and principles of language seem to be partially unique to

---

[28] Recursive procedures will be discussed in greater detail in Chapter 4.

[29] As we shall see, Chomsky approaches the problem of the creative use of language in a manner somewhat different from that of Descartes. He does not hold the human mind to be completely free and equipped to deal with universal problems, but argues for a kind of "biological rationalism", where the human mind is constrained by its biological make-up, but still creative and free to a very large extent (McGilvray 1999:85).

[30] Notice the similarities between Descartes' conditions for the presence of mind and Alan Turing's famous Turing Test, where a computer that can fool a human interlocutor into thinking that he/she is speaking to another human being can be said to "think" (see Chapter 4).

language and partially unique to mind – they cannot be formulated in terms of concepts proper to the analysis of behaviour and the interaction of physical bodies (6). For Chomsky, this argument is to be taken seriously, despite its deficiencies (1972:7):

> There is nothing at all absurd in the conclusion. It seems to me quite possible that at that particular moment in the development of Western thought there was the possibility for the birth of a science of psychology of the sort that still does not exist, a psychology that begins with the problem characterising various systems of human knowledge and belief, the concepts in terms of which they are organised and the principles that underlie them, and that only then turns to the study of how these systems might have developed through some combination of innate structure and organism-environment interaction.

The rationalist approach to philosophy of language in the seventeenth century developed a general theory of linguistic structure, which came to be known as "philosophical grammar", the particulars of which are relatively unknown today (Chomsky 1972:14). What is surprising is that much of the work that had been done on philosophical grammar mirror current linguistic trends. The theory, however, virtually disappeared with the advent of romanticism and the development of modern linguistics in the nineteenth century (76). Modern structural and descriptive linguistics were restricted to the analysis of the surface structure[31] of language, analysing formal properties explicit in the signal. When De Saussure suggested that the only methods that were proper to linguistic analysis were segmentation and classification, his view was considered to be as an advance in the field. It was thought that his methodology would render the structure of language clear. Not much attention was paid to syntax in the theories that followed. The idea that there are universal properties in the form of language was rejected, and so too was the assumption that we can learn something about the general properties of human intelligence from the study of language (14-20).

Chomsky (1972:65-66) contrasts the Cartesian approach with the practise of postulating, *a priori*, mechanisms that must underlie the acquisition of knowledge and belief, as he believes has happened in more contemporary linguistics. He accuses the behavioural sciences of a militant anti-psychologicalism and anti-mentalism, because of the shift of emphasis from the principles underlying mental structures toward the evidence that mental structures exist, i.e. behaviour. He sees this shift as decidedly unscientific, akin to natural sciences restricting themselves to organising meter readings instead of studying these readings as evidence of underlying structures and principles. This shift in orientation can also be seen in the "anti-mentalism" of philosophy of language and linguistics. According to Chomsky, the most fundamental contribution made by structuralism to linguistics is to make this anti-mentalism apparent and to lay the groundwork for demonstrating that this approach

[31] The “surface structure” of a language is the representation of the phrases that constitute linguistic expression and the categories to which these phrases belong (Chomsky 1972:105).

is inadequate to the task of addressing the problems of language and mind. Chomsky (22) maintains that the successes of the structural approach, such as broadening the scope of available linguistic information and improving on its reliability, should be augmented with the recognition of its limitations and ultimate inadequacy compared to the philosophical grammar tradition.

Chomsky (1994:160) holds that the Cartesian linguists were essentially correct in their formulation of the project of linguistics, but that their ideas had little impact.[32] Instead, the more restricted structuralist and descriptive approaches were far more influential. He criticises the optimism within these developing sciences of the early to mid twentieth century, which seemed to agree that B.F. Skinner's behaviouristic (stimulus-response) framework would, with a few minor modifications, be applicable to all aspects of language use (Chomsky 1972:2).[33] Skinner's model was furthermore applied to a range of human abilities beyond those associated with language use.

The technological advances of the 1940's, especially the development of the computer, strengthened such convictions.[34] In Chomsky's words, the assumption arose that "any apparent complexity would be disentangled by the electronic marvel" (3). With the notable exception of theorists such as John von Neumann, it seems that the consensus was that mathematics, technology, behaviouristic psychology and linguistics would converge on a coherent and clear theory of human abilities, fully adequate to addressing the classic questions of mind and language. However, by the 1960's, Chomsky detects a general tendency in all of these disciplines to find the concepts and principles on which they were founded to be "fundamentally lacking" (4). He puts this realisation down to the fact that the kinds of structures that are realisable in terms of these theories are inadequate in that they cannot be made to correspond with the apparent structures underlying natural language. He argues that these approaches were not merely inadequate, but fundamentally misguided (Chomsky 1972:4):

> It has, I believe, become quite clear that if we are ever to understand how language is used or acquired, then we must abstract for separate and independent study in a cognitive system, a system of knowledge and belief, that develops in early childhood and that interacts with many other factors to determine the kinds of behaviour that we observe.[35]

---

[32] Cf. Bracken (1994:876ff) who argues that linguistics lost its forward momentum once it was seen as a "science whose a priori character would not be as satisfactorily established as geometry's. He sees this realisation as signalling an end to the attempt at providing an "ontology of mind" (884).

[33] Among the optimists he lists in this regard are: Leonard Bloomfield, Bertrand Russell, and positivist linguists, psychologists, and philosophers in general (Chomsky 1972:2). Theorists critical to this turn in psychology and related fields were largely ignored (3).

[34] These developments will be examined in more detail in Chapter 4.

[35] Ironically, one attempt at overcoming the inadequacy of the computational approach to mind, namely an evolutionary approach, is rejected by Chomsky and computational theorists alike.

Where structuralism, in turn, is severely lacking, is in its inability to account for the mechanisms that underlie the *creative use aspect* of language and semantic content. What remains to be done, in Chomsky's view at least, is to specify the mechanisms that underlie and produce knowledge of language – which he terms, *linguistic competence* (22). A system of linguistic competence would be qualitatively different from any structures that can be described in terms of stimulus-response psychology, theories of simple automata, mathematical theories of communication, etc., because these theories are aimed at describing simple, and immediately given, or superficial phenomena. Chomsky argues that the reality of linguistic competence cannot be accounted for by extrapolating from simple descriptions of surface features, because "mental structures are not simply 'more of the same' but are qualitatively different from the complex networks and structures that can be developed by elaboration of the concepts that seemed so promising to many scientists just a few years ago" (Chomsky 1972:4, 72).[36] He defines linguistic competence as "an abstract system underlying behaviour, a system constituted by rules that interact to determine the form and intrinsic meaning of a potentially infinite number of sentences" (71).[37]

A distinction is made between linguistic *competence* and *performance*. Performance involves many factors over and above the linguistic principles that determine the phonetic and the semantic content of a sentence. Such factors would include the extra-linguistic beliefs concerning the speaker and the situation, principles of cognitive structures, as well as linguistic competence. A general theory of linguistic structure would be concerned, among other things, with the highly restrictive conditions that appear to be essential properties of human language. When Chomsky (1972:116 ff) discusses the semantic and phonetic interpretations of sentences, he is referring to the idealised representations that are determined by the grammars underlying language. He proposes to unite these two approaches to linguistics and synthesise them.

Chomsky's (1972:25) greatest criticism against both the classical rationalist and empiricist philosophies of mind is the unquestioned assumption within these theories that the properties and contents of mind are available to introspection. He observes that this assumption has rarely been questioned, even in the wake of the Freudian revolution. With regard to Cartesian rationalism, Chomsky believes that it failed to appreciate the abstractness of the structures underlying language and the complexities involved in the operations that link semantic content of an utterance to its physical realisation (25). With regard to structuralist and behaviourist theories of language, Chomsky accuses them of severely underestimating

---

[36] We saw in Chapter 1 that Peirce too believed that mental structures are qualitatively different from other structures. So much so that he could not conceive of any possible way in which mental structures could interact with other kinds of structures, which caused him to speculate that the entire cosmos is a "mental structure". In Chapter 5 it will become clear that Terrence Deacon (1997a) also believes that there is something qualitatively different to human cognition. He attributes this qualitative difference to the human capacity for symbolic interpretation.

[37] A person, who has internalised the system of rules that determine the phonetic shape of a sentence in a language and its intrinsic semantic content, has developed a specific linguistic competence (Chomsky 1972:115). It is not the degree of complexity, but the *quality* of complexity that is relevant here.

the complexity of mind and language and their structures.[38] These miscalculations allowed such theories to propose, without sufficient evidence, that language is a "habit structure", or a "network of associative connections", or a matter of "knowing how" (26). Such approaches completely overlooked that the principles underlying mental organisation "may be as inaccessible to introspection as the mechanisms of digestion or coordinated movement" (26*)*. Chomsky attributes the “fatal inadequacies” of the empiricist tradition in language studies (from which much of twentieth century linguistics and psychology took their cue) to their unwillingness to study the abstract aspects of language competence. He declares: “Had the physical sciences limited themselves by similar methodological strictures, we would still be in the era of Babylonian astronomy” (Chomsky 1972:112). Chomsky’s aim in his various linguistic theories would be to identify and study the abstract aspects of language and hence to establish linguistics as a scientific discipline.

4. PEIRCE’S INFLUENCE

The influence that Peirce has had on Chomsky’s conception of innate linguistic abilities is interesting for the purposes of our discussion (e.g. Chomsky 1972:90; 1980:136; 2000b:80).[39] As we have seen, Peirce argues that human intellectual capabilities are limited by the principle of *abduction*. All knowledge takes on the form of hypotheses, and is provisional, subject to confirmation or refutation by further experiences. Peirce argues in *Pragmatism and Pragmaticism* (1965 Vol. V) that our innate limitations, which are due to our particular cognitive structure, contributes to theory-construction in that it limits the amount of possible hypotheses that can be formulated with regard to a particular problem or situation. What appeals to Chomsky in Peirce’s theory, is his insight that evidence from the history of science seems to suggest that early scientific theorists hit upon approximately correct hypotheses with relative ease, considering the vast amount of possible explanations that could have been formulated, and the knowledge and evidence available to them (Chomsky 1972:93; 2004:160).

As we have seen, with the principle of abduction, Peirce added a further logical category to those of induction and deduction, namely *abduction*.[40] Abduction is the only logical operation that introduces new ideas into the logical realm (as opposed to deduction and induction, which only work with existing ideas). Induction only comes into play as a secondary process used to correct hypotheses and based on experiential input, while deduction merely infers the necessary consequences of a hypothesis. As Chomsky (2004:160) explains it:

---

[38] As will be discussed in Chapter 3, Derrida makes a similar claim.

[39] According to Otero (1994a:13) Chomsky describes Peirce’s theory of abduction as a development "of Kantian ideas to which recent Anglo-American philosophy has not been very receptive."

[40] Peirce defines abduction as: "the process of forming an explanatory hypotheses" (1965 Vol. V:106 [5.172]).

> [Peirce] argued that the mind has to have an essentially genetically-determined class of admissible hypotheses. So the mind is just structured so that certain hypotheses are admissible. Other possible ones are not part of our cognitive capacity and are not theories that are intelligible to us. So there are admissible hypotheses.

In terms of this pre-adaptation, we acquire knowledge by means of abduction. There are some similarities between abduction and animal instinct in that "innate ideas" that are crucial to survival have developed in the human mind due to evolution.[41] Peirce believed that it could be possible to study and determine the principles of abduction and thus create a comprehensive theory of what makes for permissible hypotheses.[42]

Chomsky adapts Peirce's theory and argues that in order to acquire language in particular, some limitations in terms of permissible and impermissible forms of grammar need to be pre-adaptively present in the organism. This is necessary in order to account for the relative ease and consistency of the form of language that is eventually acquired by a great many users, especially given the endless permutations of grammar that is possible without such restrictions (1972:91). Such innate restrictions are a precondition for the acquisition of language to be at all possible (Chomsky 1972:91):

> The child cannot know at birth which language he is to learn, but he must know that its grammar must be of a predetermined form that excludes many imaginable languages. Having selected a permissible hypothesis, he can use inductive evidence for corrective action, confirming or disconfirming his choice. Once the language is sufficiently well confirmed, the child knows the language defined by this hypothesis; consequently, his knowledge extends enormously beyond his experience and, in fact, leads him to characterise much of the data of experience as defective and deviant.

Chomsky argues that Peirce is "rather vague" about the origin of innate ideas. We have seen that Peirce believes that three different evolutionary processes are at work in the universe, and that the natural laws themselves are subject to evolution. We have also seen

---

[41] The reader will remember that Peirce distinguished between different types of evolution, and maintained that something akin to Lamarckian evolution had to be possible, in order to allow for acquired knowledge to be passed on to one's offspring (cf. Peirce 1965 Vol. VI:16-17 [6.16]).

[42] The reader will recall that abduction plays an important role in Peirce's *law of mind*. The idea is that the structure of mind that allows for abduction (which applies in the case of Firstness or the icon) is natural. He explains this postulation as follows:

> Our minds have been formed under the influence of phenomena guided by the laws of mechanics, certain conceptions entering into those laws become implanted in our minds, so that we readily guess at what laws they are (1965: Vol. VI:13 [6.10]).

See Melrose (1995:500) where she points out the similarities between abduction and Hofstadter's argument that the brain has "representational structures".

that he postulates the existence of a sort of Hegelian teleology in natural processes, where natural processes "aim" to manifest Thirdness in the universe. The mind, as subject to and a product of the natural laws is then, by logical necessity, able to manifest these laws by being a primary instance of Thirdness. The difficulty of accounting for how the mind comes to encompass knowledge of the material world (especially since Peirce concurred with Descartes that the mental is a different kind of substance) eventually leads him to propose a "universal mind"; an argument which is difficult to accept if one does not share Peirce's metaphysical preoccupations. Not surprisingly, Chomsky (1972:97) rejects Peirce's explanation for how the mind comes to posses innate restrictive ideas, and argues that the process by which the human mind develops its innate structure remains a complete mystery. He argues that attributing the development of the mind to "natural selection" amounts to nothing more than saying that one believes there to be a naturalistic explanation for it (Chomsky 1972:97):[43]

> The laws that determine possible successful mutation and the nature of complex organisms are as unknown as the laws that determine the choice of hypothesis. With no knowledge of the laws that determine the organisation and structure of complex biological systems, it is just as senseless to ask what the "probability" is for the human mind to have reached its present state as it is to inquire into the "probability" that a particular physical theory will be devised.[44]

As we shall see, Chomsky appropriates and adapts Peirce's insights with regard to abduction and creates an analogue to abduction in linguistic theory, namely Universal Grammar (Chomsky 2004:160) to create a class of "admissible hypotheses", in that it "stipulates" the allowed format for all possible grammars. In this manner, he attempts to address one of the quandaries of language identified by the Cartesian Linguists, and which he has dubbed *Plato's Problem*.

5. UNIVERSAL GRAMMAR

Throughout his linguistic project Chomsky is concerned with providing both descriptive and explanatory adequate theories of language. Chomsky (2000b:7) characterises these two concepts as follows:

---

[43] As will become apparent in Chapters 4 and 5, attributing the development of mind to evolution through natural selection amounts to rather more than "believing there to be a naturalistic account for it". See Dennett (1995:390-393) where he speculates that Chomsky is antagonistic towards Darwinism. As such, he (Chomsky) allows that physics may one day explain mind, but not biology.

[44] Elsewhere Chomsky (2000:3) refers to evolutionary theories of language in general as "fairy tales". In following chapters the argument will be made that Chomsky overstates his case in this regard.

> The grammar of a particular language satisfies the condition of descriptive adequacy insofar as it gives a full and accurate account of the properties of the language, of what the speaker of the language knows. To satisfy the condition of explanatory adequacy, a theory of language must show how each particular language can be derived from a uniform initial state under the "boundary conditions" set by experience.

Much of the permutations throughout his work have been the result of trying to reduce the tension between these two theoretical requirements. Chomsky (1994:160; 1995b:4; 2004:x) claims that traditional grammars are descriptively inadequate in that they do not convey the "facts of language". He argues that they implicitly assume that the language-user must somehow already have the prerequisite notion of the structure and general conceptual resources of a language, which enables the reader to decode the "hints" provided by the grammar. This process is akin to what happens in the case of dictionaries – they can do no more than hint at the meaning of words through giving examples that "stimulate the conceptual resources of the mind" (Chomsky 1994:160). However, it seems impossible to formulate the generative procedures that a language user employs to determine sound and meaning in terms of traditional approaches to grammar. Chomsky (2000:12) speculates that the only way to give a descriptively adequate account of a particular language would be to formulate complex, language-specific grammatical constructions with different properties that occur internally to particular languages and across languages.

However, this conclusion has to be rejected on the grounds of the second empirical requirement that is posed on a theory of language, namely explanatory adequacy. A theory of grammar that is explanatorily adequate needs to account for Plato's Problem, namely how it is possible that almost all children acquire natural language under conditions where there is a poverty of stimulus in that the sentences that they are exposed to during language acquisition do not cover even a fraction of the possible sentences that a language-user will be exposed to. The question of how children know how to use language arises, given that it is extremely unlikely that a child is born with the ability to speak a particular language. It follows that language must somehow be simple and exhibit a universal structure that underlies the specific variations of particular languages. With these observations, two theoretical goals with regard to grammar arise for Chomsky: one is to attain descriptive adequacy of specific grammars, the other is to obtain explanatory adequacy for universal grammar (1994:160; 1994:159).[45] Chomsky proceeds to pose questions about language that have been avoided or denigrated by modern behavioural science and structural approaches to language (1994:160) in an attempt to address the tension between the goals of descriptive and explanatory adequacy.

[45] In other words, descriptive adequacy requires the construction of detailed and intricate rule systems to account for the structure of grammar, which would be language specific. However, as we have seen Chomsky believes that language acquisitions has to be predetermined, which would mean that explanatory adequacy requires that the rules of grammar be universal and general (1994:160).

Chomsky develops a concept of language that is internalist, individual, and intensional, in that it is concerned with the "internal state of the brain of a particular individual" (1994:158). He argues that the person who has acquired a language (has knowledge of a language), has in some way internalised a system of rules on how to relate sound and meaning, and the linguist constructing a grammar is, in fact, constructing an hypothesis on what this system looks like. The internalised system of rules, or grammar, is one factor among several that determines how an utterance will be used or understood, which complicates the problem of confirming grammars on the basis of empirical evidence (1972:26-27). He argues that the native speaker constructs a grammar on the basis of very little evidence, and that the grammar has empirical consequences that reach far beyond the evidence that the learner has been exposed to. Hence, the person acquiring a language is constructing a particular grammar for himself (26). In accordance with Peirce's theory of abduction and the possibility that it offers for addressing Plato's Problem, Chomsky assumes that the ability to construct a grammar must be innate to human minds, just like the ability of Peirce's scientists to hit upon accurate hypotheses about the world.[46]

Chomsky defines a theory of grammar as being concerned with the question, "What is the nature of a person's knowledge of his language, the knowledge that enables him to make use of language in the normal, creative fashion?" (1972:103).[47] Chomsky bases his approach to the study of language on the observation that languages differ systematically in phonology and basic syntactic structures, but not in meaning (McGilvray 1999:73), and he holds that any adequate theory needs to account for this. To account for this observation, he argues that abstract, and in part universal, principles that govern human mental faculties must be postulated in order to account for language phenomena. The greatest challenge for theoretical linguistics is to discover the principles of such a universal grammar (UG), which in conjunction with the rules of a particular grammar, provides an explanation for language phenomena, which otherwise appears to be chaotic (1972:48, 71, 120-121; 1995b:170).[48]

In terms of his rationalist approach,[49] Chomsky holds that all members of the human species have a faculty of mind in common, namely the faculty of language. Hence, Chomsky argues that there is an innate component of the mind/brain, which yields knowledge of language when it is presented with linguistic experience. The language faculty is a "biological

---

[46] Chomsky (1975:13) avoids using the term "innate hypothesis" in reference to his theory, because "it can only mislead".

[47] The language-user has no consciousness of having learned these rules, or of using them. Furthermore, Chomsky (1972:104) doesn't see any reason to suppose that such rules could be brought to consciousness through introspection. What is needed is a scientific approach where a collection of data about sound-meaning correspondence in various languages is interpreted and the theorist tries to develop a system of rules that will account for such data, and furthermore to establish the principles that underlie the formation of such rules.

[48] Chomsky argues that the most persuasive arguments illustrating that such a challenge is surmountable comes from work done in phonology (e.g. 1972:38-48).

[49] Chomsky (1975:39) characterises a rationalist theory as one where "Learning is primarily a matter of filling in the detail within a structure that is innate". He deviates from the rationalist tradition in that in this case the "a priori system" is biologically determined.

endowment" that converts experience into a system of knowledge or grammar (1977:63; 1986:xxvi).[50] Furthermore, the language faculty would be unique to humans, as would be the creative use of language as an expression of thought, which is made possible by this faculty – it does not have a significant analogue anywhere else (1975:40). Since universal grammar manifests itself in the language faculty, Chomsky (1972:27) sees the study of universal grammar is the study of human intellectual capabilities. Such a study tries to formulate the necessary and sufficient conditions that a system must have in order to be a human language. These conditions are not accidentally true; they are rooted in the organisation of the human language capacity.

When Chomsky speaks of a person knowing a language, it is analogous to saying that that person's language faculty has attained a certain state – language is a state of the language faculty (1980:48; 1994:157; 2000:8). The theory of this state can then be called the "grammar" of the language, and the theory of the initial state of the language faculty is called "universal grammar" (UG). "Universal grammar", then, is the result of the principles that determine the form of a particular grammar by selecting the appropriate grammatical form based on certain innate data (Chomsky 1972:26). The theory of universal grammar is not limited to a particular grammar, although this also forms part of the linguist's subject matter. In terms of a particular grammar, the linguist is concerned with a particular cognitive system of a particular speaker-hearer; at the level of universal grammar, the linguist is concerned with general properties of human intelligence (28).[51]

UG is defined as "the system of principles, conditions, and rules that are elements or properties of all human languages not merely by accident but by necessity – …biological, not logical necessity" (Chomsky 1975:29), and can be taken to be expressing the essence of human language. Chomsky thinks of UG as a kind of "language acquisition device" within the brain, situated within the (idealised) language faculty. The language faculty, as a "mental organ", is assumed to be autonomous and to actually exist in the brain as a modular subsystem of the brain (Chomsky 2004:5).[52] UG specifies the properties of sound, meaning, and structural organisation of language. It imposes conditions that narrowly restrict the variety of languages. UG can be assumed to be all those properties of language that can reasonably be assumed not to have been learned and, in principle, it should be accounted for in terms of human biology (30). The language faculty is envisioned as constructing an abstract formal skeleton, which is invested with meaning by interpretive rules that specify how sound and meaning relate for specific grammars (Chomsky 1975:55). Each specific grammar provides

---

[50] Mental faculties should be distinguished from *capacities*; capacities are human abilities such as common-sense understanding or the science-forming capacity that allow them to gain understanding and deal with the world (McGilvray 1999:49). Both faculties and capacities are thought to be universal.

[51] It is this concern with general properties of human intelligence that leads Chomsky to assert on numerous occasions that linguistics is a subfield of psychology in that it deals with aspects of mind (e.g. 1972:28, 103; 1975:160): The theory of language is simply that part of human psychology that is concerned with one particular 'mental organ,' human language" (Chomsky 1975:36; 1980:39).

[52] The existence of some kind of language faculty in the human brain seems to be widely accepted within the scientific community, although its autonomy is still controversial (Fukui and Zushi 2004:10).

the formal and semantic properties of an infinite array of sentences for a particular language. These sentences and their structural organisation constitute the specific language that the grammar generates. Different languages can be learned when the language faculty is stimulated in an appropriate way. When the grammar is constructed, the person "knows" the language generated by the constructed grammar and can apply this knowledge to express thoughts and understand what is heard.[53] Language understanding and production takes place within the constraints of the internalised principles of grammar.

All children are born with a system of universal grammar in place in the language faculty – universal grammar is the initial state of the language faculty. This initial state changes in early childhood, as a result of both maturation and input from external stimuli. Eventually the language faculty arrives at a relatively stable state, which is a particular language (L). Chomsky thinks of the initial state as a procedure or an algorithm that has an array of data as its input and yields the language (L) as its output. "Output" here is understood as the internal state of the mature language faculty (Chomsky 1994:159).[54] It should be kept in mind that Chomsky constructs a computational (formal) theory of language – he is only concerned with the mental processes (computations) that link linguistic mental events, and not with what happens "outside" of the head (McGilvray 1999:4). Furthermore, he has a modular theory of such processes in that they occur in relative isolation from other mental events.[55]

According to Chomsky (1994:158), the state attained in the language faculty and its course of development are “internally directed” in that external conditions are too impoverished to account for the development of the “steady state” of the language faculty, which is characterized by highly articulated and intricate structures.[56] The steady state of the language faculty consists of a “cognitive system” and “performance systems”. The cognitive system stores information, which the performance systems can access and retrieve. The performance systems use this information for articulation, interpretation, expression of thought, referring, etc. Chomsky argues that the cognitive system accounts for our “infinite knowledge”, in other words, for our seemingly infinite capacity for relating sound and meaning and employing these to express a seemingly endless range of thoughts.

From analysing the structure of the language faculty, Chomsky proceeds to explain many of the properties of sound and meaning and the relations between expressions in a variety of languages. He argues that many of these properties derive from our "inner nature"

---

[53] See Chomsky (1975:78-136) for a comprehensive account of how the language faculty mediates between UG and the grammar of a particular language, resulting in a person knowing a language.

[54] Chomsky repeatedly cautions that an empirical enquiry would then be needed to explain how such a process occurs, but argues that his assumptions should be accepted in terms of a "best-theory" scenario, in other words, because of the explanatory possibilities that his assumptions yield.

[55] This internalist approach extends to Chomsky’s approach to meaning, where meaning is not a function of relation or reference to the external world, but a function of the intrinsic meanings of words and expressions (McGilvray 1999:4).

[56] Chomsky prefers the term “language growth” to “language learning”, believing it to be less misleading (1994:158; 2004:6).

as determined by the initial state of our language faculty, which accounts for the universal aspects of the different human languages (Chomsky 1994:159). He is sceptical of the commonly held belief that language learning takes place through applying general learning capacities.[57] He insists that there is no reason to suppose that learning theories exists (Chomsky 1975:20; 119 ff), and argues that "no one would take seriously a proposal that the human organism learns through experience to have arms rather than wings, or that the basic structure of particular organs results from accidental experience" (Chomsky 1975:9). But when it comes to the development of personality, behaviour patterns and cognitive structures in higher organisms have generally been approached (within the empiricist tradition, at least) with the assumption that the social environment is the dominant factor in these domains. Hence, "human nature" is an accidental, specific, historic product. Chomsky suggests that a more fruitful approach would be to study the acquisition of cognitive structures in the same way that we would study a complex bodily organ (although he does not endorse a evolutionary approach) (10).

Although he can only characterise the properties of language and of grammar in abstract terms, Chomsky believes that, with the progress of science, it might become possible to know something about the physical representation of grammar and of the language faculty.[58] He argues that his theories are not without empirical content, however, because it is foreseeable that discoveries in neurophysiology or in the study of behaviour and learning might cause us to revise or abandon a given theory of language or particular grammar (Chomsky 1975:36-37, 223). He emphasises that the place of the language faculty within human cognitive capacities is a matter for "discovery, not stipulation" (43). So too is the place of grammar within the system of acquired cognitive structures.

The properties of the forms of grammar that are discovered theoretically are not necessary properties of a system in order for it to be a human language, yet Chomsky observes that many properties seem to hold true for languages in general and contribute to the linguistic competence of the speaker-hearer. He is led to conclude that there exists a restrictive initial schema of some sort within the language user that directs language acquisition, and which contributes a universal structure to language. Hence, the principles of

---

[57] He characterises the empiricist conception of language learning as follows (Chomsky 1975:178):

> Thus, on these assumptions, “cognitive life in general “ is a system developed incrementally by means of association, conditioning, habit formation, generalisation, induction, abstraction of certain specific kinds that have been proposed… Any highly specific system developed on the basis of other principles will therefore be “separated from cognitive life in general”.

[58] Despite this hope, Chomsky is cautious about the capacity of science to answer complex questions. He states for example (2000:2):

> Science is a very strange activity. It only works for simple problems. Even in the hard sciences, when you move beyond the simplest structures, it becomes very descriptive… The idea that deep scientific analysis tells you something about problems of human beings and our lives and our interrelations with one-another and so on is mostly pretence in my opinion…

He goes on to argue that the one study of a complex organism where science does seem to be making headway is in the study of human language.

universal grammar provide the structure to which human language must conform, and the conditions that determine how the grammar of a particular language can be used (Chomsky 1972:61-63).[59] Universal grammar translates into a complex system of rules that involve highly abstract mental operations (Chomsky 1972:64):

> If a scientist were faced with the problem of determining the nature of a device of unknown properties that operates on data of the sort available to a child and gives as "output" (that is, as a "final state of the device," in this case) a particular grammar of the sort that it seems necessary to attribute to the person who knows the language, he would naturally search for inherent principles of organisation that determine the form of the output on the basis of the limited data available. There is no reason to adopt a more prejudiced or dogmatic view when the device of unknown properties is the human mind; specifically, there is no reason to suppose that, in advance of any argument, that the general empiricist assumptions that have dominated speculation about these matters have any particular privileged claim.

The conclusion is then that a generative grammar as envisioned by the Cartesian Linguists does exist, and that it is a system of different types of rules that are organised according to fixed principles and which contains a fixed substructure with general principles of organisation. These general rules or principles make up a "universal grammar", common to all languages. The theory of universal grammar specifies the minimal substructure that a language must posses as well as a variety of formal and substantive conditions that a further elaboration of the grammar must meet. Chomsky's argument is not that the organisation of the system is "natural" any more than any other system of the body is organised according to "natural principles". He suggests that it may be possible to restrict the schema of universal grammar to such an extent that it would be consistent with the "meagre" data available to the language learner. The learner can then be construed as learning a language by searching among the possible grammars available to him and "selecting one that is not definitely rejected by the data available to him" (Chomsky 1972:88). If this were the case, a language learner would not construct a grammar from scratch, based on data available to him, but would determine to which language the data he is exposed to belongs. The argument is not that a child is genetically predisposed to learning one particular language rather than another. Any conclusions that the theorist draws regarding the language-acquisition device of an

[59] Hilary Putnam (1980:107-116) attributes possible significant uniformities between languages to their "common origin", arguing that such an explanation is simpler than postulating the existence of an innate grammar. Chomsky's answer is that the common origin of language would still not account for the ability of a child to acquire language on the basis of insufficient input. It would seem that a theory that can account for the acquisition of language by children, based on the input an average learner would be exposed to, would provide the evidence that Chomsky (1972:87, 182 ff) insists is not available to us and without which we cannot but conclude that innate grammar exists.

individual are conclusions about universal grammar (114). Such conclusions can be falsified by being shown to be contradicted by the construction of grammar in other languages, or verified if they support such facts about other languages. This can lead to strong empirical hypotheses regarding universal grammar.[60]

Chomsky (1975:35) also argues that the faculty of language contributes to the construction of a "common-sense understanding", which yields the common-sense world that all people have in common.[61] The parallels that this assumption has with Peirce's argument that our minds evolved in a way that makes the abduction of natural laws possible should be clear. Chomsky argues that common-sense understanding, like language, is ubiquitous within the human species and seems to be acquired with as little effort (as opposed to scientific understanding, for example). A common-sense understanding situated in the language faculty would explain how our ideas seem to "anticipate" experience, in that it is possible that the internal concepts provided by the language faculty primarily have to do with those human interests that relate to our biological constitution and requirements. Our common-sense understanding is geared to navigate the world in terms of our biological constitution and requirements (Chomsky 1995a:20). Hence, Chomsky speculates that our common-sense knowledge is innately specified, with concepts such as *give* and *to*, while scientific knowledge is invented, with great effort and difficulty. The two capacities may also come into conflict, partly because, Chomsky holds, the way in which we see the world as partly a function of the concepts that are used to describe it (McGilvray 1999:91). McGilvray argues that Chomsky's conception of a common-sense understanding, rooted in the evolution of the language faculty, fulfils the function that Descartes attributed to God, "clarity", and "distinctness", namely to ensure that there is a "complying world" outside our concepts. Latent, inbuilt principles are brought to the "interpretation of data of sense", which determines experience (Chomsky 1966:65).[62] The sets of characteristics would then be shared across the human population, regardless of language. It is important to emphasise that Chomsky does not argue that the structure of language or meaning mirrors a deeper "language of thought" – on more than one occasion Chomsky argues that he knows that language doesn't mirror thought, because he can think without language.

---

[60] It should be noted, however, that Chomsky does not allow for falsification outside of the structure of language. He seems to discount the possibility that his conjecture that a language faculty with universal grammar exist could be refuted by means other than studying language structures.

[61] Chomsky (1995a:28) characterises common-sense understanding as a universal human property; it appeals to principles that are implicitly held in common by all human beings: "…how people interpret object constancy, the nature and cause of motion, thought and action and so on ('folk science' in one of the senses of the term)". As such, common-sense understanding is innate and is contributed to by natural language. Pinker (e.g. 1997:352-362) argues that such common-sense understandings are the result of evolutionary biases that have been inherited from our ancestors because they are beneficial for our survival. This possibility will be discussed in greater detail in Chapter 5.

[62] McGilvray (1999:174) compares our shared innate concepts to our shared perception of colour and sound, which we then project onto the world. As we shall see in Chapter 5, Deacon accounts for colour universals without recourse to any innately determined colour concepts.

The lack of necessary knowledge in this early period leads Chomsky to restrict himself to those aspects of language that "can be sensibly pursued", namely the study of the nature, acquisition, and use of linguistic competence. He believes that the state of knowledge available to him will allow for progress to be made in the way that the classic problems of language can be formulated and hence make novel approaches possible. He believes that his theory of UG has gone a long way towards accomplishing an explanatorily adequate theory of language. However, his conception of what UG looks like and how particular grammars come to be specified, undergo various permutations throughout his career. As we shall see, his early approach does not address the problem of the creative use of language, nor does it contribute to the study of universal semantics (Chomsky 1972:98-99).[63] But in his attempts at correcting these shortcomings in subsequent theories the probability arises that the shortcomings in his earlier theories might not only be due to a lack of empirical evidence to substantiate his theories. It is possible that those shortcomings are partly due to some of the fundamental assumptions that Chomsky makes with regard to UG and his internalist approach to language. It might be that some of his cherished assumptions about language fundamentally inhibit his aim of providing both a descriptively and explanatorily accurate account of language, and explanation for the creative aspect of language use. Perhaps they need to be either modified, or radically overhauled in order to meet Chomsky's requirement of a naturalistic theory of language (and hence of mind).

The following three sections will give a brief overview of some of the modifications that Chomsky subjected his theories to in order to render a theory of language that is both explanatorily and descriptively adequate. The subject matter is extensive and intricate and what follows in no way does justice to the intricacy of Chomsky's thought. It will merely be an attempt to orient the reader with regard to fundamental changes to Chomsky's assumptions and approach to his subject matter and the reasons for these changes. The final section will be an attempt to evaluate his progress and assess the importance of his findings with regard to the purposes of this work.

6. An Early Attempt at a Rationalist Theory of Language: A System of Rules

As we have seen, the study of Cartesian linguistics leads Chomsky to propose that the most appropriate general framework for the study of problems of language and mind is

[63] In its Chomskyan formulation the "creative use of language" principle encompasses the following facts: i) the sentences that people use are stimulus-free (it is impossible to identify or even envision any stimulus that would determine the formation of a particular sentence); ii) the sentences that it is possible to produce are unbounded; and iii) the sentences that are produced are usually coherent enough to be recognised and are appropriate to the situation in which they are produced (McGilvray 1999:6). Given that it is possible to produce an infinity of sentences in an endless variety of situations, and the absence of any determinable causes of given sentences, it seems a miracle that people generally and effortlessly produce appropriate and understandable sentences in every situation that they find themselves in (barring mental illness or other incapacities). The reader will recall that it is this observation, along with that of the "poverty of stimulus", that leads Chomsky (in line with the Cartesian linguists) to conclude that language must rely on an innate or native human capacity. Chomsky believes that a successful linguistic theory must in some way allow for these observational facts.

the one developed within the rational psychology of the seventeenth century. To the extent that Chomsky believes there have been advances in understanding linguistic competence, the assumptions and approaches that underlie them are traditional in character, so much so, that he believes the Cartesian approach to matters of language and mind has been significantly revived (Chomsky 1972:5). However, he does not believe that contemporary theory has advanced very much beyond that of the seventeenth century in determining the characteristics of intelligent behaviour, how it is acquired, the principles that govern it, or the nature of the structures that underlie it (Chomsky 1966:12).

In the 1950's Chomsky proposed a new approach to the study of language, proposing that language is acquired on the basis of innate capacities, and an innate mental structure. Structuralist grammar, the dominant theoretical approach at that stage, was limited to studying phonology and morphology. This proposed approach indicated a shift of focus in the study of language from "behaviour or products of behaviour to states of the mind/brain that enter into behaviour" (Chomsky 1986:3). The central concern then becomes our knowledge of language, where it originates, what it consists of, and how it is used. Chomsky's approach to linguistics was characterised by his substitution of traditional grammars with *generative grammar* as the primary object of linguistic study. "Generative grammar" is the system of rules that specifies sound-meaning relations for a given language (Chomsky 1972:104). Hence, the object of study became the "rules" that underlie language and that a child intuits and internalises in order to acquire language.

Once we have a tentative understanding of what a generative grammar would consist of, we can attempt to formulate the problem of the origin of language, which Chomsky formulates as: "What initial structure must be attributed to the mind that enables it to construct such a grammar from the data of sense?" (1972:79). Such a system cannot be formulated in an a priori manner; it is an empirical question.[64]

In the earlier versions of his linguistic theories, Chomsky proposed that a system of propositions, which express the meaning of a sentence, is produced in the mind as the sentence appears as a physical signal. The physical signal and the system in the mind are related by means of certain formal operations, called *grammatical transformations*. Furthermore, he distinguishes between the *surface structure* of a sentence (organised into categories and phrases that are associated with the physical signal) and the underlying *deep structure* (the abstract system of categories and phrases that are represented in the mind and which determine semantic interpretation).[65] In practise, the deep structure of a given language will generate a deep structure for each sentence, and will contain rules on how the deep structure is to be related to surface structure. The rules that express the relation of deep structure to surface structure are called *transformational rules*. In addition to these features,

---

[64] Chomsky (1972:80) argues that studies of problems of mind have been hampered by an "*a priorism*" in their approach.

[65] Searle (1994:71) sees the introduction of the deep structure of sentences, which can differ from surface features as a "crucial element" in the Chomskyan revolution.

the grammar of a given language will contain further rules that relate paired deep and surface structures ("syntactic objects") to phonetic representations, and to representations of meaning. Such *phrase structure rules* indicate how to generate deep structure representations for sentences, which yields an overall grammatical structure for a particular language, which then transforms meanings into sounds and particular phonetic and syntactic shapes. A person understands a language when he has internalised these rules and can apply them to an infinite range of sentences (Chomsky 1972:106, 125, 162 ff; 1986:56; McGilvray 1999:71).

To know a language means having the ability to assign deep and surface structures to sentences, to relate these structures, and to assign a semantic and a phonetic interpretation to the paired deep and surface structure. Both deep structure and surface structures contribute to meaning. Deep structure provides the grammatical relations of predication and modification, which contribute to meaning, while surface structure determines, in part, matters of focus and presupposition, topic and comment, the scope of logical elements, and pronominal reference (Chomsky 1972:111). The surface structure might differ completely from the deep structure and there need be no point-by-point correlation between the deep structure and its physical signal (Chomsky 1966:30-51; 1972:28-30). The rules by which deep structure is transformed into surface structure are not well understood. Furthermore, it is difficult to distinguish between the contribution of grammar to meaning, and the contribution made by contextual phenomena.

The broad outlines of Chomsky's early attempt at constructing a framework for linguistic study run as follows: The grammar of a language is a system of rules that determines the pairing of sound and meaning. Grammar consists of a syntactic component, a semantic component, and a phonological component. The syntactic component defines a certain class of abstract objects, namely deep structure and surface structure. The deep structure contains information relevant to semantic interpretation, while the surface structure contains information relevant to phonetic interpretation. The grammar that tries to explicitly indicate how phonetic form and semantic content are intrinsically associated in a particular language is called a generative grammar. Furthermore, the semantic and phonological components are interpretive. Grammar as a whole relates semantic and phonetic interpretations, mediated by the rules of the syntactic component that define paired deep and surface structures (Chomsky 1972:125-126).[66] Universal semantics and phonetics also make up a part of universal grammar and should be included in the study of language structure.

There seems to be no *a priori* reason that language should be designed in this way and not any other, and Chomsky suggests that one reason among many possible others for its structure is the existence of empirical constraints that linguistic communication has to meet. Constraints would include the fact that language is also spoken and as such has to

[66] Chomsky (1972:126) qualifies this framework by saying that it is an informal first approximation of what such a framework should look like. A precise theory of grammatical structure will provide a "technical" meaning for the terms "deep structure" and "surface structure", and would allow the empirical question to be raised of how these structures contribute to semantic and phonetic interpretations.

allow for short-term memory constraints (156). This initial view of language as a rule-based system would undergo some modifications in subsequent years: rules of semantic interpretation were introduced that operated on deep structure (cf. Chomsky 1966), and later linguistic expressions were taken to be represented at four levels (D-structure, S-structure, Logical Form (LF) and Phonetic Form (PF) with features of semantic interpretation represented at levels closer to the surface level [cf. Chomsky 1986:66-68]). Chomsky (1972:112) emphasises that his approach to studying language takes place by means of abstraction, which is not illegitimate in itself, but it must be kept in mind that it rests on certain a priori assumptions about the mind:

> We are reduced to platitudes, or to observations, which, though perhaps quite interesting, do not lend themselves to systematic study by means of the intellectual tools presently available to us. On the other hand, we can bring to the study of formal structures and their relations a wealth of experience and understanding…I feel fairly confident that the abstraction to the study of formal mechanisms of language is appropriate; my confidence arises from the fact that many quite elegant results have been reached on the basis of this abstraction. Still, caution is in order. It may be that the next great advance in the study of language will require the forging of new intellectual tools that permit us to bring into consideration a variety of questions that have been cast into the waste-bin of "pragmatics", so that we could proceed to study questions that we know how to formulate in an intelligible fashion.

With regard to possible theories of grammar, Chomsky ran into various difficulties. For example, because a natural language has a finite number of lexical items that must be able to generate an infinite number of possible sentences, grammar must work according to recursive procedures. Hence, a grammar that is descriptively adequate for a natural language must allow for recursion. According to Chomsky, finite state grammar – a very simple recursive grammar – is likely to appeal to those that have a naïve view of what a language is, such as the 1950's engineers and mathematicians of information and computational systems, who believed that sentences of language could be treated as sequences of words generated by a simple Markovian computational system (McGilvray 1999:128 ff).[67] However, finite state grammars are inadequate for generating natural languages, or even formal symbol systems. They are not descriptively adequate, nor do they provide the possibility of explanatory adequacy.

Pure phrase structure grammars, on the other hand, recognise grammatical categories of constituents and generate strings of elements with grammatical structure (McGilvray 1999:129 ff). Phrase structure grammars consist of a set of rewrite rules, which

[67] This approach to grammar will be discussed in Chapter 4.

takes formal category symbols as input and produces other category rules as output. They allow for a large number of constructions, but cannot deal with complex grammar constructions without a large body of additional rules and exceptions. It cannot be explanatorily adequate for all natural languages, and it is difficult to explain how children could come to learn all of these, possibly language specific, rules. Hence Chomsky rejects his early finite state grammars, because they are not descriptively adequate, and the idea of pure phrase-structure grammars, which are not explanatorily adequate (127).

The question arises as to how much of Chomsky's proposed theory is empirically viable. We saw that Chomsky proposed to alleviate the tension between descriptive and explanatory adequacy in his linguistic theory. In these terms, his rule-based linguistic approach had to be shown to be descriptively adequate (Chomsky 1986: 69ff; 1995b:3ff; 2004:x). But many problems arose in this regard in various attempts to construct explicit grammars. There were far too many possible systems available within this format, which meant that it could not answer Plato's problem. Descriptive adequacy within this model would have to sacrifice explanatory adequacy. In Chomsky's (1995b:5) words:

> Recognition of the unsuspected richness of complexity of the phenomena of language created a tension between the goals of descriptive and explanatory adequacy. It was clear that to achieve explanatory adequacy, a theory of the initial state must allow only limited variation: particular languages must be largely known in advance of experience. The options permitted in Universal Grammar (UG) must be highly restricted. Experience must suffice to fix them one way or another, yielding a state of the language faculty that determines the varied and complex array of expressions, their sound and meaning; and even the most superficial look reveals the chasm that separates the knowledge of the language user from the data of experience. But the goal of explanatory adequacy receded still further into the distance as generative systems were enriched in pursuit of descriptive adequacy, in radically different ways for different languages. The problem was exacerbated by the huge range of phenomena discovered when attempts were made to formulate actual rule systems for various languages.

As we shall see, his proposed solution to this dilemma is to show that the apparent complexity of grammar is epiphenomenal – resulting from the interaction between fixed and probably abstract principles, which can vary in application. Hence, the complexity of language is only apparent, and the phenomenal differences between different languages stem from the same basic principles of UG. The implications of such an understanding led (in the 1980's) to a conception of language that, in Chomsky's (1994:161) formulation, "departed radically from the 2500-year tradition of study of language".

7. A Principles and Parameters Model of Language

The period from the 1960's to the 1980's was characterised by an attempt in linguistics to abstract relatively simple general principles and properties of rule systems from natural languages and to attribute these generalities to the language faculty. In the 1980's the approach converged in the principles and parameters (P&P) approach, which Chomsky (2000:14) describes as a "radical departure from the tradition of several thousand years of linguistic research". From the 1950's to the 1980's generative grammarians were mainly concerned with constructing a descriptively adequate grammar of language (Fukui and Zushi 2004:8). With the advent of the principles and parameters approach, focus shifted to explanatory adequacy (16).[68] The complexity that resulted from the rule-based theories of grammar made language appear to be too complex to learn, especially by a child within the time constraint that we know accompanies language acquisition. Chomsky's response was to propose that many more aspects of language are innate than were originally thought (Smith 2000:x). Empirical evidence and novel technical tools led to what Chomsky (2000:ix) terms "a breakthrough", in linguistics in the 1980's. A new language model was formulated, which brought the principles of UG and the language faculty "to the centre stage of language research" (Mukherji 2000: x).The biology of the language faculty became a topic of enquiry, as well as the question of how much of UG theory was empirically viable, and how much of it was based on "theoretical convenience". This new language model was termed, a "principles-and-parameters" (P&P) model (Chomsky 1994:161) and Chomsky hoped it would be capable of addressing the question of explanatory adequacy.[69]

The P&P approach rejects the concept of rule and grammatical construction entirely (Chomsky 2000b:8). Essentially, such a model suggests that actual and possible states of natural languages are states of UG or general principles, with certain options of variation (or parameters). The general principles are independent of language and its construction. Traditional grammatical constructions (interrogative, passive, etc.) are now thought to be "taxonomic artefacts," and grammatical categories are thought to be the result of the interaction between the fixed general principles of UG and the particular settings of the parameters in natural languages.[70] Hence, there are no rules of grammar in the traditional sense, only principles and parameters that Chomsky describes as "indifferent" to traditional grammatical constructions. The new approach holds the possibility that the variety of

---

[68] It remains an open question, however, whether Generative Grammar is any closer to an adequate explanatory theory of language (cf. Fukui and Zushi 2000:17 ff).

[69] The P&P model underwent substantially different versions and we shall attempt to give a general description of what this model entails, while concentrating on its more recent versions.

[70] According to Chomsky (1995b:25 ff) descriptive statements are reduced to two categories in the P&P approach, namely language-invariant and language-particular categories. Language-invariant statements are principles, and language-particular statements are specifications of particular values of parameters. The notion of construction in terms of grammar disappears, so that there are no such constructions such as Verb Phrase, or interrogative and relative clause, or passive and raising constructions. "Rather, there are just general principles that interact to form these descriptive artefacts" (26).

languages can be infinite, in that the interaction between fixed principles and limited parameters could, in principle, lead to an infinite variety of languages. However, Chomsky (2000a:14) argues that the parametric variations seem to be a finite space, which means that there can only be a finite number of languages that satisfy them. Languages have superficial differences but are at base identical, hence this approach offers an explanation of the traditional insight that language involves "infinite use of finite means" (Chomsky 2004:149, ix). There are fixed universal principles common to all languages, with the variation between languages largely restricted to parts of the lexicon, i.e. phonology and morphology.

Metaphorically Chomsky (1994:161) describes P&P as holding that language is a "network" that is "not completely wired-up at birth":

> It is associated with a switch-box, with a finite number of switches, which can be set on or off. When they are set, the network functions, different settings may yield quite different phenomenal outputs. To the extent that the picture is spelled out, we can "deduce" Hungarian or Yoruba by setting the switches in one or another way… Though much less is understood, something similar is must also be true of lexicon, with the links it provides to the space of humanly accessible concepts and signals.[71]

With the P&P model, Chomsky envisages the infant as possessing an innate and universal grammar, with the structural options of all possible natural languages represented. Very early in the child's life, certain parameters are set in the language faculty (the mechanism by which this takes place is not entirely clear) and a specific natural language is acquired. In this manner the parameters for a specific language is set within the options provided by UG. It is envisioned that UG under one choice of "settings" could render Japanese, while under another choice of settings it will render Swahili. Each possible human language is identified with a specific setting of the switches or parameters.

The settings need to be acquired on the basis of the very limited information available to the child learning a language. However, the speculation is that very few parameters are necessary to cover all broad structural differences between natural languages (McGilvray 1999:114). So the idea is that a single innate system (UG) can contain all of the possible structural variations necessary to generate any of the possible human languages.[72] The parameter switches need, somehow, to be set virtually automatically at an early age, based on a child's early experience (115). The child also needs to be able to "set" the switches in his neural architecture through experience (this is not thought to be a conscious activity). Chomsky asserts that small changes in the settings will lead to large variety in output as the

[71] Readers familiar with connectionism will be examined in Chapter 4. Chomsky's model is theoretical and he does not speculate on actual manifestations of his model in the neuronal structures of children.

[72] McGilvray (1999:114) calculates that if we were to assume 14 switches in UG, with a maximum of two possible, independent positions each, it would be possible to generate 16,384 languages.

effect reverberates through the system[73] (Chomsky 2000b:8). The parameter switches are not restricted to structural aspects of UG, but also apply to the phonological component of UG. In this language model meaning is innate, and common to all languages.

The P&P approach assumes that for each particular language, the cognitive system involved consists of a computational system (CS) and a lexicon (Chomsky 1995b:6 ff). The lexicon specifies the entities that CS selects and integrates in order for it to form linguistic expressions. P&P stipulates that the lexicon should not posses any redundancy – it should provide the information that CS requires in an optimal form, "excluding whatever is predictable by principles of UG or properties of the language in question" (6). Items of the lexicon belong to *substantive categories*, i.e. nouns, verbs, adjectives, and particles. Categories such as tenses and complementisers are assimilated under *functional categories*. Now language variations and typologies can be attributed to the "choice" of values of parameters on the part of the language learner. Chomsky is vague about which components of language these parameters can be found in; he argues that too little is understood about them to venture a strong hypothesis.

Hence, the language (L) consists of a lexicon and a computational procedure that uses lexical materials to construct linguistic expressions with their sound and meaning. It may be that the computational procedure is fixed and identical for all languages, while variation is restricted to the lexicon (Chomsky 1994:161). Language acquisition is defined as the process of fixing the parameters of the initial state in a specific way, chosen from a selection of possible settings (Chomsky 1995b:6). Each choice determines a language, identified as an "I-language" (I stands for "internal", "individual", and "intensional"). I-language is the natural language as it is manifested in a particular language user's language faculty. It is defined in terms of UG and contains a list of the specific lexical items in the user's language faculty.[74] I-language is distinguished from E-language, where language is approached as an external phenomenon, and hence studied in terms of linguistic behaviour quite apart from the internal workings of the mind that may accompany language. As such, an E-language approach works in terms of common-sense understandings of language and is usually accompanied by the conception of the mind/brain as a blank slate. E-language approaches include theories of language games[75], or speech acts[76]. Chomsky' questions whether a coherent theoretical conception of E-language is possible (Chomsky 1995b:16; McGilvray 1999:107).

---

[73] Readers familiar with theories of complex systems will recognise the overtones of such theories in this approach. It seems that Chomsky does not explicitly make a link with complexity theory; nor does he give an explanation of why these changes in parameters will have such great effects on the output of the language system.

[74] It is important to note that I-language cannot be acquired without the necessary stimuli from the environment (Fukui and Zushi 2004:3). However, the richness of I-language suggests that much of the information in I-language is not necessarily included in environmental stimuli, and could not have been learned by general learning strategies, such as induction, generalisation, and analogy.

[75] See Wittgenstein (1958).

[76] See Austin (1980).

The computational system (CS) generates the forms of an infinite range of symbolic objects, called structural descriptions (SD's) (Chomsky 1995b:14). Each linguistic expression (SD) is generated by the I-language and includes instructions for the performance systems in which I-language is embedded.[77] By virtue of integration into performance systems a brain-state, I-language, qualifies as a language (Chomsky 2000b:27). Such systems fall into two general types: articulatory-perceptual systems, and conceptual-intentional systems (28).[78] This leads Chomsky to conclude that an expression includes two interface levels, one which provides information and instructions for the articulatory-perceptual systems, and one which does the same for the conceptual-intentional systems. The first interface is then assumed to be phonetic representation (PF), and the other logical form (LF). Sounds and meanings occur at two separate interfaces, where the language faculty interacts with other systems or faculties in the brain. Each SD specifies the full array of phonetic, semantic, and syntactic properties of a linguistic expression. The lexicon specifies the lexical items that appear in the SD's (Chomsky 1995b:20). Each lexical item pairs specific meanings to specific sounds.

Hence, one aspect of an SD is a system of representation, or *D-structure*, and lexical items are inserted into the *D-structure*.[79] The SD provides information about properties of linguistic expressions, including their sound and meaning (21). These features provide instructions to other faculties in the brain involved in language production (the "phonetic representation interface interacts with sound production and reception faculties, while the "semantic representation interface" (SEM) interacts with the conceptual and intentional systems).[80] PF and LF must be universal so that any expression of any actual or potential language can be represented within them; they must be interfaces, in that their elements can interact with the sensorimotor systems and other systems involved in thought within the brain, and they must be uniform, in that all languages must interact with the other language systems within the brain in the same manner, so that they capture all and only the properties of the system of language as a whole (Chomsky 1995b:21). In attempting to envision what the interactions among the various levels might look like, Chomsky adds the caveat that this is an "empirical issue", which will have to be tested as such, even though the relevant empirical evidence is hard to come by (22 ff). However, he assumes that empirical research will confirm these postulations.

---

[77] Performance systems play a role in articulation, interpretation, expression of beliefs and desire, referring, etc

[78] "Intentional" here is understood in the traditional philosophical way of "aboutness", or being about something (Chomsky 2000:8).

[79] It is expected that languages will be very similar at D-structure and LF levels.

[80] It is not clear what the "conceptual" and "intentional" systems are, but as McGilvray (1999:97) points out, Chomsky believes that recent work has been shedding light on this question, although Chomsky does not believe that it is the task of linguistics to concern itself too much with questions about systems and faculties beyond the language faculty. The central task of the linguist is explaining how expressions are produced. Chomskyan theorists, Fukui and Zushi (2000:9), are very aware of the idealised nature of his model of mental faculties, and argue that so little is known about "the other cognitive systems that interact with the language faculty" that it is not at all certain that a scientific theory of language use will ever be possible.

With the semantic representation interfaces (SEM's), Chomsky attempts to make theory of meaning part of "serious science" (McGilvray 1999:100). SEM's are something akin to Frege's senses[81] and thus makes sense part of syntax (words). Reference and truth ("the rest of semantics") become part of pragmatics (100). Pragmatics concerns itself with context, i.e. the identity of the speaker, the intentions and desires of the speaker, the conditions of use of a word, etc. Theorists who are concerned only with the "semantics" of the word on the other hand, restrict themselves to "syntax plus machinery", as McGilvray puts it. In this manner it is possible to generalise beyond single cases and avoid the difficulties posed by introducing the variable psychological states of individual speakers into the mix. Chomsky (1995b:30-51) doubts whether the pragmatic aspects of language use can be incorporated into a coherent theory of language, and restricts himself to the syntactical aspects of language use, the intrinsic and invariable features of words, and the biological computational systems that produce language. Free and creative language use occurs in the domain of pragmatics, and as such might be beyond the reach of science. With the "sense" of a word being part of syntax, the syntactic description of a word includes formal semantic features, through "feature-specifying" terms, which help determine the "meaning" interface part of an expression (cf. McGilvray 1999:103).These meanings do not determine referents, but guide their own use and interpretation in that they affect linguistic "computations" (105). In this manner, meaning becomes biological, universal, and amenable to science.

We have seen that Chomsky aims for an explanatorily adequate theory of language with the P&P approach. The P&P model seems equal to this requirement in that it seems able to account for UG and is a mechanism for selecting between different possible natural languages, despite the poverty of available stimulus to any one child. It also allows for a distinction between a common natural language between a community of speakers of a particular language and the individualistic "internal" language of a particular speaker-hearer, or I-language. Furthermore, the approach fulfils Chomsky's requirement that linguistics restricts itself to the "scientific" aspects of language studies in that it seems to have an analogue in research done on other aspects of brain development.[82] Even meaning seems to be accounted for, in that it can be approached in terms of syntax, rather than the contingencies of context specific states of the speaker-hearer. In order to fulfil its destiny as an explanatorily adequate linguistic theory, theorists will have to determine how parameter

[81] In Frege's theory, sense is the idea in the mind, while reference is the thing in the world which corresponds with the term and its sense. (Morning star, evening star – has the same reference (the planet Venus), but different senses (which are the result of an historical context in which they were thought to be two different bodies). The common-senseidea here (as in much subsequent language philosophy) is that what the word refers to results from the idea that it invokes. The question that inevitably arises, of course, is what are "ideas"? And how do ideas and their "objects" correspond, if at all.

[82] McGilvray (1999:121) points out the similarities between the parameter switches available to a child and the postulated process by which a kitten's visual system develops the capacity to see horizontal lines for example, by "setting the switches" of its inherited visual system. In the absence of the relevant input, the capacity does not develop.

values are set by a finite amount of universal parameters, through experience. Chomsky (1995b:7) describes this as "a problem" that can be "constructively pursued".

An overview of the literature would suggest that there are certain difficulties that Chomsky cannot seem to overcome within this framework. Chomsky is candid in his assessment of the success of the P&P approach (Chomsky 2000b:9):

> This is, of course, a program, and it is far from a finished product. The conclusions tentatively reached are unlikely to stand in their present form; and, needless to say, one can have no certainty that the whole approach is on the right track As a research program, however, it has been highly successful, leading to a real explosion of empirical enquiry into languages of a broad typological range, to new questions that could never have been formulated before, and to many intriguing answers…whatever its fate, the program suggests how the theory of language may satisfy the conflicting conditions of descriptive and explanatory adequacy. It gives at least an outline of a genuine theory of language, really for the first time.

Before we evaluate the success or failure of the P&P approach, it will be useful to explore the changes that Chomsky subjected it to, in terms of his theoretical development.

8. THE MINIMALIST PROGRAMME

Since the P&P model's development in the 1980's, it has undergone several modifications. In the 1990's emphasis shifted to what Chomsky (1995b) has termed the Minimalist Programme. In terms of this approach Chomsky adopts a "principled framework" and holds that a linguistic theory must "meet general considerations of conceptual naturalness that have some independent plausibility, namely, simplicity, economy, symmetry, non-redundancy, and the like" (52). Chomsky (2004:151) suggests that the principles, on which language is based, might be "general principles of an organic nature", which might indicate that there "is a principled reason why things are the way they are". We have seen that Chomsky thinks of the language faculty as a "computational system" and as such he proposes to approach it in terms of general computational principles. The main question that the minimalist programme posts in this regard is: "How good a solution is language to certain boundary conditions that are imposed on the mind?" (Chomsky 2002:17). This approach leads him to do away with many of the structures postulated in earlier generative grammar, claiming that "Nature can't be that ugly" (Chomsky 2004:151). The main thesis of the minimalist programme is that "human language is a computational system that connects the sensory-motor system and the conceptual system in an optimal way" (Fukui and Zushi

2004:16).[83] In other words, emphasis has shifted to the question of how close language comes to "optimal design".[84] This approach is aimed at further reducing the tension between descriptive and explanatory adequacy. The minimalist programme is an attempt to radically rethink the foundations of linguistics and hence giving up all linguistic constructs that are not conceptually or empirically necessary (Smith 2000:xi).[85] The result has been that much of the theoretical constructs of earlier generative grammar approaches have been abandoned, such as deep and surface structure, D-structure, S-structure, and LF, for example (cf. Chomsky 2004:152). The only structures in the language faculty that are retained are phonetic representation (PF) and the semantic interface. New explanations, more in line with current empirical findings (especially in terms of the computational approach to mind), have been

---

[83] This requirement makes more sense when one takes into account Chomsky's view on what determines growth in any organ. Chomsky (2004:xii) believes there are three such factors, namely genetic endowment, environment, and "general principles of growth and development that are not specific to the organism, perhaps not even to the organic world". Thus, in terms of the P&P approach genetic endowment fixes principles and parameters and the environment sets the parameter values. The third factor, "general principles of growth and development" should also play a fundamental role in the development of the language faculty as an organ. Chomsky argues that the most far-reaching general principles of growth and development that we have are computational. Hence, he expects that principles of computational efficiency would have a crucial effect on the growth and development of organs; particularly the acquisition of I-language in the case of language.

The approach is "principled" in as far as the properties of I-language rely on the "general principles of growth and development" and on the conditions that the language faculty must meet in terms of interface legibility. Also see Fukui and Zushi (2004:12ff).

[84] "Optimal design" refers to the optimal solution that language can achieve in terms of principles of computational efficiency. As Fukui and Zushi (2004:17) caution, the origin of the optimality principles is not clear. Chomsky concedes that the conditions "external" to the language faculty that determine the optimal solution are only partially understood. Hence the research task of the minimalist programme becomes to clarify the nature of the interfaces and optimal computational principles "through investigation of how language partially satisfies the conditions they impose (Chomsky 2004:xii). He often refers to the work done by Turing (1952) and Thompson (1942) as a possible approach to the optimality of language.

Chomsky (2004:170ff) briefly explains his view by juxtaposing their position with a position that is prevalent in biology, namely, in François Jacob's famous phrase: "Nature is a tinkerer". In this view, evolution works in a somewhat haphazard fashion, creating organisms that are due to evolutionary opportunism and accident, rather than conforming to any sort of "plan", optimality, or perfection. Chomsky describes Turing and Thompson as believing in a "Galilean style". Galileo believes that nature was perfect and that scientists can discover and describe this perfection. In this approach biology might ultimately be reduced to physics and chemistry. Chomsky (1980:8 ff, 2004:180) seems to agree with this view. Chomsky characterises Galileo's understanding of "perfect" as holding that it would be possible to build a perfect machine that can duplicate nature. According to Chomsky, Newton shared this conception of what "perfection" in nature would entail, i.e. a mechanistic system that follows simple laws, although he ended up proving that such a conception was an "absurdity" (much to his own consternation).

Chomsky argues that two modern manifestation of the Galilean style can be found in 1) methodological minimalism (which is taken to yield "the best theory"); and 2) substantive minimalism (which seeks perfection in objects by reducing them to principles, such as chemical or biological principles, or computational complexity, etc.). He holds that we do not have a clear conception of what methodological minimalism would entail. We do not really know what a good theory would entail. He argues that many questions that should be asked are not, because they are not confined to the mechanisms or mathematical models of mechanism. Chomsky argues that these approaches work for physics, "because physics deals with very simple things", but when phenomena become too complicated, "physics hands it over to somebody else" (Chomsky 2000:2; 2004:174).

[85] Minimalist theorists are careful to stress that the Minimalist Programme is just that, a programme, and not a theory. It defines possible directions that research might take, rather than proposing a definite hypothesis (e.g. Smith 2000:xi).

sought. Much of the assumptions and the methodology of the Minimalist Programme are still attributable to the attempt to present the available data in an understandable form, rather than to empirical motivations. Chomsky (2000a:16) argues that our understanding is too limited to eliminate certain assumptions about language and mind, even if some of these assumptions contradict one another.

The assumption that the language faculty exists as some part of the mind-brain that is dedicated to the knowledge and use of language is retained (Chomsky 2000a:17; 2000b:3 ff). So too is the assumption that the language faculty is a species-property with little variation across the species, and unique to the species. The language faculty is regarded as including a cognitive system, which is defined as "a system that stores information" (Chomsky 2000b:4, 117). Chomsky still postulates the existence of further systems that access that information, what he calls "performative systems", which would include the sensorimotor systems, and articulatory-perceptual systems in the brain. The cognitive system of the language faculty changes state in terms of "internal" maturation processes and external experience, i.e. language acquisition. It is also assumed that the states attained by a specific language faculty due to language acquisition are superficial, in that they are still largely determined by the common language faculty.[86] Language provides instructions to the performative systems in the form of a potentially infinite set of linguistic expressions (or generative grammar) (8). Chomsky still distinguishes between two categories of performance systems[87]: those that represent sound (accessed by the sensorimotor-systems) and those that represent meaning, or conceptual-intentional systems (which Chomsky describes as "mostly mysterious"). Conceptual-intentional systems access aspects of expressions which enable one to talk about the world and express thoughts and feelings, and so forth. The performance systems access different parts of expressions, which means that expressions have two kinds of symbolic objects as parts (9).[88]

It terms of the minimalist programme Chomsky (2000b:10) questions whether there are additional “interface levels” necessary to language processing active in the mind/brain. He also questions whether there are other levels internal to language, such as deep and surface structure, for example. The minimalist programme holds that everything that has been postulated in modern linguistic theory with regard to internal structure and rules in language can be better accounted for in terms of “legibility conditions” at the interface between the language faculty and faculties in the brain that have to do with sound and meaning production

[86] Chomsky (2000:7) likens the "growth" of the language faculty to the growth of wings by an embryo, or the onset of puberty at a certain stage. Hence, the changes are largely genetically determined and uniform, with some influence from environmental factors. Like Pinker (1995; 2002) he argues that those that treat the language faculty as different from other biological "organs" are enamoured with an irrational form of dualism.

[87] Chomsky (1995b:3) cautions that this assumption is likely to be false in that it is unlikely that articulation and perception involve the same interface, but deeply rooted in the western intellectual tradition .

[88] Chomsky (2000:9) is again, careful to caution that he is "into pretty far-reaching empirical assumptions about the architecture of the mind", but insist that they are reasonably plausible and a good basis on which to build.

and interpretation. It also holds that the only computational processes are those that are required by “the weakest assumptions about interface properties” (10). One such assumption is that language consists of “word-like units” that have to be interpreted by the relevant systems in the brain. Secondly, it is also assumed that these units are organised into expressions, and thirdly, that these items have the properties of sound and meaning. The minimal assumptions about language under this approach are: i) language has properties of sound and meaning (“features”), ii) items are assembled from these features (“lexical items”), iii) complex expressions are constructed from these items (10).

From this basic framework, it is postulated that the computational system in the brain that generates expressions has two basic operations: i) assembling features into lexical items, and ii) forming larger syntactic objects out of these items. Chomsky thinks of the first operation as a lexicon that indicates inflections, which plays an important role in computation. He argues that optimal design would require that no additional features be added in the course of computation, which contradicts much of his previous theories, where grammar rules are invoked. Furthermore, he aims to show that no structural relations are brought into play during computation, other than those necessary for legibility or “induced in some natural way by the computation itself” (such as local relations between features, and elementary relations between syntactic objects) (11).

Chomsky notes that there exists much empirical evidence that proves the proposals of the minimalist programme wrong, and that these proposals contradict a core assumption of the principles-and-parameters framework, which proposes that language is highly “imperfect” (11). However, he believes that his new approach sheds the “excess baggage” of earlier theories and improves on explanatory and descriptive adequacy (11). He argues that his recent work indicates that variation among languages might be of a limited nature, and can be reduced to properties of inflectional systems. He proposes that language variation is limited to a narrow part of the lexicon. Legibility requires that the features that are assembled into lexical items are divided in three ways: i) semantic features (interpreted at the semantic interface; ii) phonetic features (interpreted at the phonetic interface); iii) features not interpreted at either interface. Hence, each feature is a device to facilitate computation.

In a perfectly designed language, there would be no uninterpretable formal features, which Chomsky (2000b12; 2000a:21) argues is not the case with natural language. Furthermore, at the level of syntactic computation language is characterised by the displacement property, which is also an imperfection in language design (Chomsky 2000b:12; 2000a:23).[89] If we were to assume optimal design, we would expect these two apparent imperfections in natural language to be related. Chomsky argues that they are related in terms of the legibility conditions imposed on the structure of language by the structure of the

[89] Chomsky describes the “displacement property” as the phenomenon in language where phrases are interpreted as if they were in a different position in the expression where similar items sometimes appear. Phrases are interpreted in terms of natural local relations (Chomsky 2000b:12). He argues that this property is universal in language and has "enormous" consequences for interpreting sound and meaning (Chomsky 2000a:23ff).

mind/brain. In his words: "…the displacement property is, indeed, forced by legibility conditions: it is motivated by interpretive requirements that are externally imposed by our systems of thought, which have these special properties (as the study of language indicates) (Chomsky 2000b:13).

We have seen that the computational operations themselves had originally been assumed to be of two kinds: i) phase-structure rules that form larger syntactic objects from lexical items, and ii) transformational rules that express the displacement property (Chomsky 2000b:13). According to Chomsky, research has shown that the computational operations to do with language are much more complex and various than had originally been supposed. The minimalist programme postulates that the variety and complexity in phrase-structure rules are only apparent and that a perfect solution to this problem would be to eliminate them in favour of an irreducible operation that attaches two objects (that are already formed) to one another, thereby forming a larger object with the properties of the "target of attachment" (Chomsky 2000b:13; 2000a:21ff). Chomsky names such an operation "Merge" and argues that it is the only computational operation active in the language faculty. He characterises it as a free operation along the lines of: "Build bigger things out of smaller things" (Chomsky 2004:152). Its application to a specific case is determined by general principles that interact with the specific parameter choices of a particular language. Furthermore, he argues that this theory is the simplest assumption that accounts for the mechanism of displacement in that the only operation is the attachment operation, which means that "language is perfect enough so that the simplest assumption is correct, which has lots of consequences" (Chomsky 2000a:25).

At some point during the "Merge" operation we will have a syntactic object (phase) which is passed on to the interpretive systems (phonology and semantics). Chomsky (2004:152) rather obtusely explains that "they do whatever they do to it, and then they're finished with it, and they forget about it". This process can go on indefinitely and can produce sentences of any complexity, until "at some point maybe you stop and apply the final Spell-Out, interpret the maximal syntactic object so far constructed, and you're done." The idea is that an optimal computational procedure would consist of such a "Merge" operation and operations to construct the displacement property – some kind of transformational operation (a copying operation). The transformational operation should be reduced to its simplest form. Chomsky suggests that there are overt and covert transformations that apply at the phrase level and differ from one another in terms of ordering (2004:152). Transfer operations "hand things over to phonology and semantics".

Fukui and Zushi (2000:13) suggest that the minimalist programme raises the possibility that the language faculty is, in fact, something analogous to the postulation of a "speech organ" within the brain. A speech organ consists of various other organs and structure within the human body, which interact to produce complex linguistic sounds. These organs exist for purposes other than language production, but are secondarily used for speech production. They suggest that such a scenario is feasible with regard to the "language

organ" within the brain. This approach could have far-reaching implications for an evolutionary approach to language.[90] They argue, however, that it is not possible to even speculate about this question at this point, mainly because the "essential properties" of the language faculty are not known in detail (15). They suggest that generative grammar has been approaching "a good approximation" of the nature language faculty, which needs to be taken into account by any theory about the origin and evolution of the language faculty.[91]

In this regard, Chomsky (2004:157) observes that language has to interact with at least two other "systems of the body" (in accordance with the traditional assumption that has its origins with Aristotle), namely the sensory-motor and computational-intentional systems. These systems have their own properties, independent of language, which means that they will impose their own requirements on the language faculty where it interacts with them (the interface). Hence, if some of the properties of the internal state of the language faculty satisfy conditions imposed at the interface, those aspects of language can be explained in terms of the external requirements placed upon it; something which Chomsky calls a "principles explanation" (158). In the MP there are two sources of principled explanation, namely interface conditions, and "language external" conditions (which are the general principles of organisms that have already been discussed). All other influences are "unprincipled", by which he means that "there is no current explanation" and there "may never be an interesting explanation", in that it may be due to an evolutionary accident, or some unfathomable principle according to which the brain works. In these terms, Chomsky describes the minimalist programme as an effort to "make as many things principles as possible and see how far you can go" (158).

---

[90] In spite of his apparent aversion to evolutionary theory, Chomsky considers the origin and evolution of language to be of fundamental importance to the study of language (Chomsky 1975:252; 2004:x, 49). He is sceptical, however, about our current knowledge of evolution in general, and the evolution of the human mind, believing the theory to be limited and insufficient for addressing the important questions with regard to our language capacity, namely Plato's problem and the creative aspect of language use. As he has stated more than once, "You study what you can study, not what you would like to study" (Chomsky 2004:150). At the moment, it seems, theorists steer clear of many of the issues that can be raised in terms of evolutionary theory, because "it's too hard" (150). Furthermore, he emphasises (as both Peirce and Darwin had and Gould still does) that there is no reason to assume that natural selection is the only mechanism through which modifications in species takes place (Chomsky 2000b:163; see also Otero 1994b:707 ff)). He argues that the more we learn about complex systems, the operation of physical law, and factors involved in spontaneous self-organisation in systems, the easier it would be to accord natural selection its proper place in evolutionary theory. He also argues that the status of internalist approaches (such as his) remains unaffected by such considerations. In Chapters 4 and 5 some of these assumptions will be challenged, and it will become clear that an evolutionary approach does challenge the status of internalist approaches to language and mind.

In contrast to Searle, for example, Chomsky does not believe that functional explanations of language are more accurate than the formal explanations by which his work is characterised (cf. Chomsky 2004:55).

[91] Although they do not reject the possibility that the language faculty evolved "purely on the basis of natural selection" they remain sceptical and declare:

> On the other hand, one of the tasks of approaches like minimalism is to give enough substance to the claim, which is still quite vague at this point, that certain physical and mathematical laws of nature, which define the "channels" through which natural selection works, are reflected in the core properties of the language faculty (Fukui and Zushi 2004:18).

Semantically, Chomsky now argues that some features of a lexical item like "book" or "cat" that are internal to it will determine the modes of interpretation for it. He holds that words are interpreted in terms of factors such as material constitution, design, intended and characteristic use, institutional role, etc. (15). Such properties are used to identify objects and determine the categories to which words will be assigned. Chomsky holds that these categories are semantic features that parallel the phonetic features that determine sound (16). He leaves open the question as to the extent to which the specific properties of language are the consequence of "biochemical laws applying to objects with general features of the brain" (16).

Chomsky holds that a possible approach to semantic interpretation is one which adopts Hume's proposal that the identity that we ascribe to things is established by virtue of the way that people think, rather than through any properties inherent to things.[92] "The semantic properties of words are used to think and talk about the world of perspectives in terms of the perspectives made available by the resources of the mind, rather in the way phonetic interpretation seems to proceed" (16). Accordingly, Chomsky argues that whereas contemporary (analytic?) philosophy asks what a word refers to and gives different answers, the question itself has no clear meaning. The answer to such a question would vary, depending on how "the semantic features" are used when we think and talk (17). Words do not pick entities out of the world. Hence, Chomsky deems "conventional approaches" to semantic matters to be "very dubious"[93] (17).

Within the minimalist framework the operations of the language faculty are minimal and conceptually simple. Expressions are sets of features present in lexical items and operations are local and limited to the present (McGilvray 1999:149). Chomsky's suggestion is that languages are "surprisingly 'perfect'", in that they satisfy general conditions imposed on them in a "near-optimal" way, in contrast to other biological systems, which seem to be "messy" solutions to evolutionary problems, given particular sets of constraints, the materials available, and accidents (Chomsky 2000a:18 ff).[94] This new framework approaches language as a design problem that the language faculty solves in terms of the biological constraints set by the other cognitive systems involved in the production of language. The argument is that the solution that has come about is surprisingly elegant, especially in terms of making sounds and meanings legible at the interfaces (McGilvray 1999: 150; Chomsky 2000a:19). For example, an expression would have to have a certain temporal order, phonetic properties,

---

[92] With regard to Hume's philosophy, Chomsky (1980:31) argues that he was wrong in his empirical assumptions about how the mind comes about, but right in as far as he formulated his questions in a manner analogous with physics. Chomsky argues, however, that Hume's empiricism was far less extreme than has been presented in contemporary philosophy (Otero 1994a:4).

[93] In the following Chapter a similar criticism made by Jacques Derrida will be examined.

[94] As will become clear in the discussion of evolutionary theory in Chapter 4, viewing language as a "perfect" or "near-optimal" system would lead evolutionary theorists to question its suitability as a candidate for evolution by means of natural selection. Evolution works in terms of various environmental and physiological constraints, as well as by appropriating accidental mutations. Such a process rarely, if ever, renders elegant and logically pleasing results. Evolutionary products can only be considered "perfect" in terms of a very specific and contingent context.

and rhythmic properties, etc., in order to be legible by the sensorimotor system (Chomsky 2000a:19). These features are not logical necessities for language. Chomsky speculates that the conceptual-intentional systems require certain kinds of information about words and phrases, with certain kinds of relations among them (19). If language is perfect in a strong sense, then sound-meaning relations and a range of empirical properties of expressions will follow from an optimal solution to the legibility conditions. In this case, the "best-theory" for studying language would be one that considers just satisfaction of the legibility conditions. Hence, "if we knew enough about the legibility conditions, the sound-meaning relations would follow (20).[95] Chomsky (1995b:317) hopes that the minimalist approach preserves descriptive adequacy, and deepens explanatory adequacy.

Chomsky (2004:163) argues that if the strongest version of minimalism were to be true, language would be the optimal solution to the interface conditions that are part of primate biology. The "sole innovation" of evolution in this instance would be that which is domain-specific for language, namely the recursive linkage of the interface systems. He is careful to reiterate that this scenario would only be true if the strongest version of minimalism were found to be correct, which he strongly doubts. Part of an evolutionary approach to the language faculty would be to identify which structures and properties are "domain specific" (in a weak sense) to the language faculty over and above such a basic linkage of the interface systems. The argument is not that there are aspects of the human brain that are exclusively dedicated to language; components of the language faculty could be directed toward other functions as well (164). But there would be a unique interaction dedicated to language, which should follow principles of optimal design to some extent.

It is important to keep in mind the MP is a programme and, as Chomsky (2004:154) states, "a good deal remains under various modifications (Chomsky 2004:157):

> But once you begin to look at it carefully, insisting on fixed generative rule systems, you find that the apparent variety and complexity is extraordinary. And now we're back to the paradox again. And the question is whether you can show that you can not only resolve the paradox and the "what" and the "how" questions in some serious way, but maybe you can even go beyond and show that there's something about that that approaches perfection, that really arises from the way things like principles of efficient computation operate on a system like this. If we look a step further, if you take the final state, the I-language – notice there's really no difference between looking at the initial and the final state, the difference between explanatory and descriptive adequacy is just terminology. You're just looking at states of the language faculty, and you look at all of them the same way…

[95] Although Chomsky concedes that there does not seem to be anything in biology that indicates that perfect design in Chomsky's (2000a:20) sense is a possibility, he insists that there is reason to suppose that language is "surprisingly close to perfect in that very curious sense; that is, it is a near-optimal solution to the legibility conditions, or what are sometimes called the 'bare output conditions'".

If conclusions about the "perfection" of language are correct, Chomsky (2000a:29) holds that it will be possible to pursue the "basic quest of the Minimalist Programme", which is to try and show that the universal properties of language are explicable in terms of the principles of optimal design, given the requirements of interface legibility. He argues that the intuition that things are perfectly designed has often been productive in the hard sciences (for simple systems), because they have been discovered by "far-reaching abstraction from the phenomenon of ordinary life" (Chomsky 2000a:30). Chomsky's own feeling is that language has surprising and interesting properties for a biological system[96] and that those are where the interesting questions about language lie (30).

As far as the creative aspect of language use goes, Chomsky argues that it "remains as mysterious" to us today as it seemed to the Cartesians (17). He believes that there remains much work to be done in linguistics and that a lot of what will be done in the field will be descriptive (Chomsky 2004:186-187). With regard to the study of language he argues:

> Language can and should be studied from many different points of view, hence in many different disciplinary contexts. The disciplinary lines are fairly artificial at best, mostly for convenience, and over time they should dissolve and surely not be regarded as straight jackets for enquiry and educational programmes. The parts of the study of language that we have been talking about seem to me to fall, in principle, into human biology. If there is sufficient progress, I think there is a fair chance that institutional arrangements will evolve in that direction. But to predict the course of thought, ideas, discoveries, that's even more vain than predicting human affairs generally.

Fukui and Zushi (2004:24) note that the minimalist programme (though more an extension of the P&P framework than an alternative framework) did not have the same kind of impact on linguistics that the P&P approach had. They argue that this might be due to the fact that minimalism deprives theorists of the "tools for linguistic description". They argue that the MP nevertheless has great significance for linguistics, in that it has the effect of eliminating "pseudo-explanations" that have long been accepted within linguistics. Importantly, minimalism does not accept UG as a touchstone, but views it as an "explicandum whose properties call for further explanations" (24). In this understanding, minimalism has the potential to transform linguistics and creates the possibility of linking linguistics to other sciences of the mind/brain.

[96] Chomsky (2000a:42) argues that there is nothing of significance known about community and culture that relates to questions about the nature of language as a biological system, which is why he doesn't pay any attention to those factors in his theories. His point is not that culture and community are unimportant, but that there is little scientific understanding of them.

9. WHAT CAN WE LEARN FROM CHOMSKY?

Pylyshyn (1994:597) argues that Chomsky's greatest contribution to cognitive science was to bring back scientific realism. We can extend this observation to language philosophy as well. He argues that the Chomskyan turn made it possible to treat some aspects of the human capacity for language as compatible with mechanism.[97] Chomsky moved scholarly research away from its empiricist and positivist bias which had dominated research on language since the nineteenth century (Kampf 1994:842) and delivered decisive criticism on language learning theories that used results from research into stimulus-response actions in animals in highly restricted environments and extrapolated them to human verbal behaviour (cf. Bracken 1994:878). Furthermore, Newmeyer (1994:926) argues that there has been a Chomskyan revolution in linguistics because "anyone who hopes to win general acceptance for a new theory of language is obliged to show how the theory is better than Chomsky's". He quotes C.F. Hockett as viewing the publication of *Syntactic Structures* as 'one of only four major breakthroughs' in the history of modern linguistics, along with the publication of Sausurre's 1916 *Cours de linguistique général* (928).

In this sense it seems unproblematic to argue that Chomsky has made an invaluable contribution to the study of language and mind. He has managed to point the way out of some of the conceptual traps that have been the legacy of the assumptions and preoccupations of generations of scholars that have focussed on these phenomena in the past, especially as they had manifested in behaviourism and structuralism. He does not hold that all inherited insights be summarily abandoned, but that they be re-evaluated, both in terms of scientific method and new empirical findings. Furthermore, he seems to have set the agenda for research into the relation between language and mind for some time to come. At present it does not seem to be possible to approach the question without taking the UG hypothesis into account. Plato's problem is a very real problem when it comes to accounting for language acquisition and it is clear that there are few options open to us at present in terms of answering the problem through unmodified language-learning theories.

One criticism that is levelled at UG is that of neurologist A. R. Luria (1994:321) who "highly disapproves" of Chomsky's hypothesis on, as he puts it, the "inherited nature" of linguistic structures. He argues that UG leads to a philosophical dead end, and that the further development of linguistics will depend on setting the theory aside and studying the epistemological issues and the psychological roots of speech and language. He argues that Chomsky's assumption of the innate nature of deep linguistic structure is based on philosophical defects and leads to scientific shortcomings. Thus, his main criticism against Chomsky is his innate postulations and suggests that we look to evolution and we look for the roots of basic linguistic structures in the interaction between active subject and reality, and not

[97] Pylyshyn's (1994:598) position is that the Chomskyan turn was one of two developments that brought scientific realism back into the realm of the mental. The other was the work begun by Hilbert and developed by Turing, Church and Markov which formulated abstract notions of mechanism and of "information processing" (cf. Chapter 4).

to the mind itself (327). While Chomsky has defended his reasons for not focussing on the evolution of language, an evolutionary perspective seems to be unavoidable in the study of language and mind. In a following chapter we will explore an attempt at an evolutionary explanation for the structures of both language and mind.

Another criticism that is frequently levelled at the UG hypothesis is exemplified by Tomasello (2004:642-645) who argues that there seems to be no consensus on what would constitute UG, nor on what would be sufficient counter-evidence to falsify it. Furthermore, he argues that all of the empirical phenomena that are usually cited as proof for innate UG are consistent with "the existence of biological adaptations for more general skills of human cognition and communication."[98] In the spirit of Chomsky, Wunderlich (2004:646) replies to this objection with the observation that the UG hypothesis should be seen as "a general idea" directing scientific investigation, rather than a specific hypothesis that can be refuted by means of particular evidence. Like Chomsky, he holds that the UG hypothesis is defendable in terms of its "fruitfulness" in that it yielded valuable insights and led to extensive and interesting research into the phenomenon of human language. At the very least, he argues, external evidence points to the conclusion that "There must be *some* UG" (647). It must be emphasised that Chomsky always argues that the UG theory is subject to falsification through empirical evidence. And as we shall see in Chapter 5, in the light of some empirical evidence, it seems that Wunderlich is right when he states that there must be *some* UG.

Searle (1994:80) emphasises how "peculiar and eccentric" Chomsky's approach to language is. Much of Chomsky's theory seems to run counter to ordinary, plausible and common-sense assumptions about language, such as the assumption that the purpose of language is communication.[99] Searle has reservations about Chomsky's "counterintuitive" approach to language, arguing that it is pointless "and perverse" to study structure independently of function, because both of these aspects of language interact. Furthermore, he argues that Chomsky's system cannot accommodate semantics, which hampers the development of an accurate and useful theory of language. Compare this criticism with Moravcsik's (1994:189) position that that Chomsky's focus on syntax is an important corrective in language studies, since too much of recent philosophical work on natural languages, notably in positivism and structuralism, focused on semantics, and on meaning and reference in particular.[100] Searle's suggested remedy for this shortcoming is to graft Chomsky's study of syntax onto the study of speech acts (91).

Contrary to Searle's position, it can be argued that Chomsky's "counterintuitive" approach to language is one of the strong points of his theory. One of Chomsky's major contributions to language and mind is his insistence that neither phenomenon needs to

---

[98] See Newmeyer (2004) examples of such alternative explanations.

[99] Searle (1994) understands communication in much the same way that we saw Peirce did, i.e. as encompassing communication with others and with oneself through "internal dialogue".

[100] See Fodor and Katz (1994:31-67) for a comprehensive discussion of the limitations of such approached to language.

adhere to our common-sense understanding of them. Nor do inherited philosophical assumptions, like the value of introspection in studying the mind, need to be uncritically adhered to. Chomsky is correct in arguing that much of contemporary physics is "counterintuitive" from the point of view of our common-sense understanding of the world, but is accepted in terms of its ability to account for the behaviour of phenomena.[101] There is no guarantee that our common-sense understanding renders a more accurate conception of phenomena or a more useful one for that matter. There is no good reason to assume that language and mind phenomena are more amenable to common-sense or intuitive scrutiny than to scientific methodology. Chomsky's approach explicitly breaks away from the remnants of assumptions that mental phenomena are open to introspection and that cognitive structures somehow mirror language structures. This is a necessary break, seeing that approaches that work in terms of these assumptions do not seem to be yielding much by way of useful hypotheses.

With regard to Searle's position on semantics, we have seen Chomsky is highly sceptical of the study of semantics, given his criticism of behaviourist and structuralist approaches to language. It seems that Chomsky cannot envision how semantics can be studied in a fruitful manner, which has led him to subsume the subject with syntax and linguistic structure. One major objection that Chomsky has to a theory of semantics is that semantics would probably require studying E-language, a problem that he believes to be too difficult to be profitable in terms of current knowledge. This seems to be a major shortcoming in Chomsky's work, one which he might yet address. It is difficult to guess how he might do that in terms of the P&P approach and the Minimalist Programme. The lack of a theory of semantics and the exclusion of community and culture from Chomsky's theories, while substantiated, are failings in his work. While he does not hold his work up as a comprehensive language theory and, especially in terms of his latest output, cautions that he is merely proposing a "research programme", these aspects seem to be deficiencies that run through the entire body of his work. In the following chapters we will explore possible approaches to the problem of semantics and the influence of the world external to the language user on language.

It is striking that Chomsky substantially reviews his own theories on language to the extent that he makes some significant about-turns in the course of his career. There is a case to be made for the argument that he pares his conception of the structure of grammar and the

---

[101] The approach of hypothetically postulating mechanisms in theory construction in a manner similar to that employed in the other sciences is summed up by Katz (1994:114) as follows:

> If the logical consequences of the model match the observable behaviour of the system and would not do so without the hypothesis, the scientist may say that this hypothesis accounts for the behaviour of the system in terms of the behaviour of the unobservable but causally efficient component. If the model is the simplest one which enables the scientist to derive all the known facts and predict previously unknown ones as the effects of the hypothesized component, he can assert that his model correctly pictures the structure of the system and its unobservable components. In this way, a linguist can assert that his theory correctly represents the structure of the mechanism underlying the speaker's ability to communicate with other speakers.

corresponding language faculty in the mind down to such an extent that by the time we get to the Minimalist Programme Chomsky has effectively erased any substantive conceptions of these structures that he might have had in his previous theories. What we are left with is a "research programme" which needs to be fleshed out to a significant extent to have any real value. Furthermore, the necessity of a modular language faculty in the brain becomes more and more questionable, especially when considered in terms of developments within neurobiology. As will become apparent in Chapter 5, evidence suggests that language is highly distributed throughout the brain and there seem to be no brain structures that are exclusively dedicated to processing language. Chomsky's insistence on the existence of a language faculty in the brain seems to be fundamentally influenced by the computational approach to mind. This approach, and its shortcomings, will be discussed in more detail in the following chapter.

It is interesting to note that Chomsky's changing view of both language and his postulated language faculty has kept in step with major changes in approaches to mind throughout the latter half of the century. The Minimalist Programme has much in common with the connectionist approach to mind, for example. Another way in which Chomsky's work has reflected attitudes in contemporary sciences of mind is in his stance on evolution, and the role that it plays in his conception of both language and the language faculty. Chomsky (e.g. 2000b:163; 2004:150) is sceptical about the possible explanatory applications of natural selection to language and the language faculty, believing that other principles may be discovered that are just as relevant to the evolution of complex structures as the principle of natural selection.[102] This sceptical position seems to be prevalent among linguists (Newmeyer 2003)[103] and is preserved in Chomsky's contemporary work. He tempers his scepticism in later works, but still seems to shore up his theories against evolutionary interpretations, thus strengthening Dennett's (1995) claim that Chomsky is antagonistic towards evolutionary theory.

In "The faculty of language: what is it, who has it, and how did it evolve?" (Hauser, Chomsky and Fitch 2002) Chomsky is again adamant that evolution may not be able to account for the structure of the language faculty. Hauser, Chomsky, and Fitch particularly question the applicability of "the argument from design" which is used in neo-Darwinism as an indication that an entity is a plausible product of natural selection (see Chapter 5). However, they revise their conception of the language faculty to such an extent that it becomes almost

---

[102] See also Dennett (1995:384-400) where he discusses what he perceives as Chomsky's hostility towards evolutionary approaches to the mind and language (389):

> …although he [Chomsky] insisted that the "language organ" was innate, this did *not* mean to him that it was a product of natural selection! Or at least not in such a way as to permit biologists to pick up the buck and analyse the way in which the environment of our ancestors had shaped the design of the language organ over the eons. The language organ, Chomsky thought, was *not* an adaptation, but…a mystery, or a hopeful monster. It was something that *perhaps* would be illuminated some day by physics, but not biology.

[103] Other prominent sceptics include Gould and Lewontin (1979) and Kauffman (1993).

unrecognisable in comparison to Chomsky's earlier theories (2002:1573). Chomsky, Hauser and Fitch argue that the only component of the faculty of language (in a narrow sense) that is "uniquely human" is its capacity for recursion. They then speculate that recursion might not have evolved "for" language but "for" other cognitive abilities. Pinker and Jackendoff (2005:204) note that this position seems to be quite a radical departure from Chomsky's earlier positions that i) only the human brain is specialised for the complex ability of language, and ii) language is distinct from other cognitive systems. Thus, the Hauser, Chomsky, and Fitch paper can be construed as a significant recantation on the part of Chomsky. The position is, however, consistent with regard to Chomsky's sceptical attitude towards natural selection. Whereas Chomsky, Hauser and Fitch accept that various aspects of language can be explained in evolutionary terms, their recursion theory raises the possibility that "language proper" is not an adaptation. As we shall see in Chapter 4 a major shift takes place in theories on mind and language in the last decade of the twentieth century and the Darwinian theory of natural selection is reintroduced as a scientifically legitimate (and prolific) approach to language and mind. The question arises as to why Chomsky is willing to change his mind on so many aspects of his theories (perhaps in line with new developments within various fields pertaining to language and mind), but remains resistant to natural selection as applied to language as adaptation, which has been met with enthusiasm by various theorists. What is to be gained from viewing language as a by-product of other evolutionary adaptations, rather than an adaptation in its own right?

In this regard, Pinker and Jackendoff (2005:201-217) argue that the recursion-only hypotheses of Hauser, Chomsky and Fitch is extremely weak and they question the motivation behind the hypothesis, given its disparity with both the facts of language and with Chomsky's earlier work (Pinker and Jackendoff 2005:218-219). They conclude that the rationale for the recursion-only hypothesis is Chomsky's Minimalist Programme. Given that this programme is based on the belief that only representations of sounds and representations of meaning are truly indispensable to language, all the other linguistic structures and principles should be eliminated from the theory (219). *Merge* and its principles of economy are the only elements necessary for creating a system of language and the core of language can be characterised as a "perfect system". The complexity of linguistic phenomena can be attributed to the need for language to interface with the systems for thoughts and speech organs. Pinker and Jackendoff argue that this seemingly elegant, simple, and optimal system of language does not support evolutionary theory and particularly the argument from design, because "if language per se does not consist of very much, then not much had to evolve for us to get at it" (219). The auditory, vocal, and conceptual systems may have evolved, but they are not part of the language structure that is unique to humans, and could have evolved for various reasons other than language "proper". "Merge" would be the only modification that needs to have been added to enable these already existing structures to support human language, and it may "even have been affected by a single genetic change that became fixed in the population through drift or other random processes"

(219). It thus becomes unnecessary to appeal to adaptation to explain the apparent adaptive complexity of language. Thus, it seems that Chomsky is loathe to give up some of the assumptions underlying his minimalist programme, and interprets evolutionary evidence accordingly. Pinker and Jackendoff point out that Chomsky himself admits that the major problem of the Minimalist Programme is that "All the phenomena of language appear to refute it" (220). In terms of scepticism toward the ability of science to explain complex phenomena, Chomsky chooses to adhere to the programme which appeals to his own intuitive sense of what an elegant solution to the problem would look like. It is not at all clear that the Minimalist Programme can be empirically vindicated, which Pinker and Jackendoff argue is the presumption that lies behind the recursion-only hypothesis (2005:220-222). This places the proposal that the Minimalist Programme refutes the evolution of language hypothesis on very shaky ground.

In addition to their recursion-only hypothesis, Hauser, Chomsky, and Fitch (2002) give three more reasons for why the language faculty could not have evolved by means of natural selection: i) the purpose of language is not communication and is in all probability badly designed for communication, ii) language is the "optimal" mapping between sound and meaning, unlike other biological systems, which are not perfect, and iii) the language faculty (in the narrow sense) originated in other cognitive abilities, and hence did not evolve "for" language. These arguments are aimed at nullifying the contemporary move towards explaining the existence and structure of language in terms of its gradual development because of the adaptive benefits that it conferred its possessors. As will become apparent in Chapters 4 and 5, these arguments by Hauser, Chomsky, and Fitch exhibits fallacies that are common in objections to theories of natural selection.

Pinker and Jackendoff (2005:223-231) address each of these objections in detail and show them to be unconvincing. Firstly they argue that the communicative capacities of language are unmistakable, if not always infallible. Chomsky’s position that language is "for" inner speech is implicit in the claim that the function of language is not communication. Pinker and Jackendoff (224-225) argue that language displays many features that are unaccountable for if one assumes that it arose solely for the purpose of inner dialogue, such as phonology, phonetics and linear order, among others. Furthermore, the tenet of the Minimalist Programme that language consists of mapping from meaning to sound would not be a necessary feature of language, as Chomsky claims it is. One of the very few plausible reasons for being able to map meaning to vocal sounds is to be able to communicate meaning "externally", as it were. Furthermore, we do not invent the language that we use in inner speech; we learn it from the community.

Secondly, the assumption that language is an optimal and inevitable mapping between sound and meaning is highly suspect (Pinker and Jackendoff 2005:225-226). It seems that Chomsky considers invented formal symbol systems to be examples of perfect languages. Pinker and Jackendoff argue that the "perfection" or "optimality" of natural languages and symbol systems cannot be compared, because they are not designed to

satisfy the same desiderata. Language needs to be used in real time by real agents with limited computational capacity and knowledge, and are gained from experience in situations that are subject to vagaries. Formal languages are designed and are stipulated by formal arbiters, which are implemented in machines with relatively vast computational capacities and relatively less restrictions on tome. Furthermore, the "optimality" of the Minimalist conception if language is far from apparent, and the criteria in terms of which "optimality" can be measured are unclear (cf. 226). It should be added that Darwin (1985:217-224) has shown that structures need not be "perfect" or "optimal" in order to be of benefit to their possessors. In fact, evolutionary speaking, optimality and perfection are highly contingent notions, varying in accordance with a highly variable environment. As Pinker and Jackendoff (2005:229) declare in this regard:

> There is a simpler resolution of the apparent incompatibility between biology and Minimalism, namely that Chomsky's recent claims about language have it backwards. Rather than being useless but perfect, language is useful but imperfect, just like other biological systems.

Finally, with regard to the contention that the language faculty did not necessarily evolve in order to support language, Hauser, Chomsky, and Fitch make a mistake that is often made by opponents of evolutionary theory. Even if the language faculty was originally selected for functions other than language, there is no reason to preclude the possibility that it was not subsequently selected because of its language function. Hence, even if their contention is accepted, there is no reason to believe that the language faculty was not subsequently shaped by selection for language.

Pinker and Jackendoff (2005) present a much more comprehensive and erudite critique of the recursion-only hypothesis than is suggested by this summary. The concern here is more to do with apparently anti-evolutionary stance taken up by Chomsky and his co-writers and the reasons for it. The Chomsky, Hauser, and Fitch paper seems to be a case in point, seeing that, as Pinker and Jackendoff point out, the thesis expounded in it runs into tremendous difficulties without seeming to gain any explanatory power over previous hypotheses. In contrast (Pinker and Jackendoff 2005:231):

> The alternative in which language is an adaptation for the communication of language and its intentions faces none of these problems. It is consistent with behavioural and genetic evidence that language shows multiple signs of partial specialisation this task rather than grafting one component (recursion) onto a completely unchanged primate base. It is based on defensible conclusions about the nature of language established by existing linguistic research rather than a promissory program that is admittedly incompatible with the facts. It does not require tendentious claims such as that language is non-redundant,

perfect, unsuited for communication, or designed for beauty rather than use. It meshes with other features of human psychology that make our species unusual in the animal kingdom, namely a reliance of acquired technical know-how and extensive cooperation among non-kin. And it does not imply that linguistics poses a crisis for biology but rather helps to bring them into consilience.

10. CONCLUSION

It can thus be argued that Chomsky runs up against his own common-sense conceptions of what an acceptable model of language and its manifestation in the mind would look like. Apart from his dedication to abstract theorising, which disregards much of the empirical evidence produced in neurology, psychology, and so forth, the greatest shortcoming in Chomsky's work can be argued to be his dedication to a computational approach to mind. In such an approach the mind is regarded as something akin to symbolic processors, i.e. computers. As will become apparent in Chapter 4, the computational approach to mind disregards major dissimilarities between minds and computers. It is only when these dissimilarities are acknowledged and incorporated into such theories that models can be constructed that resemble actual natural languages and the actual workings related to mental processes.

First, however, we will examine another attempt at overcoming some of the shortcomings of behaviourist and structural approaches to languages and the mind. We have seen that Chomsky failed to account for meaning (semantics) in his theories, which he attributes to the fact that the problem is "too difficult". He also disregarded contextual aspects of language for similar reasons. The following chapter will be dedicated to a post-structural theorist's attempt to overcome these difficulties in terms of a novel theoretical approach to language.

CHAPTER 3

DERRIDA ON MEANING IN LANGUAGE

"[S]tructuralists and semioticians optimistically elaborate theoretical metalanguages to account for textual phenomena; poststructuralists skeptically explore the paradoxes that arise in the pursuit of such projects and stress that their own work is not science but more text (Culler 1994:25).

First, I would like to say, even if it shocks certain amongst you and even if I myself took my head in my hands when Richard Rorty said that I was sentimental and that I believed in happiness, I think that he's right (Derrida 1996:77).

1. INTRODUCTION

In the previous chapter the argument was made that Chomsky's analysis of language focuses primarily on the structure of language, and is partly a reaction to the shortcomings that he attributes to structuralist and analytical approaches (positivism and ordinary language philosophy) to language, all of which focus on meaning. He accuses these disciplines of imposing arbitrary methodological restrictions and of employing many of the unjustified, non-naturalist assumptions that are part of our theoretical heritage, probably due to our (often erroneous) common-sense understanding of language. However, we have also seen that Chomsky despairs of the ability of science (in its current from, at least) to account for the everyday use of language, both because of the difficulty of the problem and the limitations of scientific methodology and human cognitive capabilities when it comes to explaining complex problems. One of Chomsky's major criticisms of structuralism was that it failed to account for the creative aspect of language use, and for its semantic content. However, despite various attempts throughout his body of work, Chomsky cannot account for either the creative aspect of language use or for meaning. And as we have seen, he remains sceptical of the possibility to do so. Furthermore, Chomsky objects to having to account for meaning through studying E-language (aspects of language external to the language faculty of the individual language user) and attempts to subsume semantics under syntax and grammar. He believes that

studying E-language would be too difficult in terms of current knowledge. For similar reasons he excludes community and culture from his language theories. In the previous chapter the possibility was addressed that Chomsky's own common-sensepresuppositions lead him to conceptual barriers that cannot be surmounted unless he revises some of those assumptions. In this regard it might be useful to examine another attempt at identifying and overcoming entrenched notions about language and cognition.

Jacques Derrida is both heir to, and very critical of, the structuralist tradition and of behaviourism. As with Chomsky, Derrida is greatly concerned with the structure of language, especially as it pertains to meaning. In many respects he touches upon the same questions that Chomsky does, but with somewhat different results. As will become apparent, the very possibility of meaning and the manner in which it arises in language are central concerns for Derrida. In fact, he seems to have found a way in which meaning in language can be explained, thus disproving Chomsky's scepticism. Another feature that these two theorists have in common is the appropriation of the insights of Peirce, although Derrida concentrates on slightly different aspects of Peirce's work.

Derrida, as a continental philosopher, worked within a setting very different from that of Chomsky. Where Anglo-American philosophy made a "linguistic turn" early in the twentieth century and philosophical problems came to be regarded primarily as problems of language, phenomenology continued to dominate the philosophical scene in continental Europe. However, French philosophical thought shifted to the analysis of structure and language during the 1950's and 1960's under the profound influence of structuralism (Johnson 1993:1). Much of this shift was characterised by a shift from the hegemony of Hegel to a re-appropriation of the work of Heidegger, as structuralism came to exert a powerful influence over the social sciences. Johnson (2) puts the "linguistic turn" in French thought down to the combined influence of these two theoretical orientations – Heideggerianism and structuralism. He highlights the renewed dialogue with science in the social sciences, in contrast to the preceding existentialist-humanist paradigm, based on the rejection of idealism in favour of philosophy of the concrete. The general shift in the paradigm to structuralism, which had originated in the late nineteenth century, came to manifest itself in contemporary discourse as a whole, driven in part by the interdisciplinary reach of the structuralist programme (primarily in the form of the appropriation and application of Saussurean linguistics).

Johnson (1993:4) asserts that a minor epistemic shift took place within structuralism itself during the 1960's, where emphasis in the linguistic analogy shifted from a focus on the structure of "language" as such to focus on the logic of "writing", as reflected in the work of Derrida. He is quick to point out, however, that, despite its major influence, it would be a mistake to consider the work of Derrida to be the origin of this conceptual shift. He argues that Derrida himself recognised that the shift in emphasis was symptomatic of developments within a wider interdisciplinary context (cf. Derrida 1976:6-10). The shift in emphasis is often

characterised as indicative of a split between "structuralism" and "post-structuralism"[1], although Culler (1994:25) argues that it is not always useful to cast structuralism and post-structuralism as opposite and distinctive programmes. He argues that the acute critical self-awareness that characterises post-structural theory was evident from the beginning of the structuralist enterprise: "Enterprises now deemed post-structuralist, such as critiques of the sign, of representation, and of the subject, were manifestly already under way in the structuralist writings of the 1960's." And, for all his criticism of Saussure, Derrida "is pursuing with the greatest possible rigour the structuralist principle that in the linguistic system there are only differences without positive terms" (28) which is an insight that Derrida appropriates from Saussure's work on language (Derrida 1986:5). Hence, Culler argues, the distinction between structuralism and post-structuralism is highly arbitrary (30).

Derrida's work should not only be read against the background of structuralism[2], but also against the phenomenological tradition to which he was an heir. Some of his most enduring themes and extensive writings are engagements with Husserl and Heidegger. Furthermore, Derrida's work seems to be significantly influenced by (or at least in step with) developments in the information sciences, cybernetics, biology, and (indirectly, at least) systems theory.[3] Although he is accorded only brief mentions in Derrida's work, the influence of Peirce is evident in many aspects of Derridean theory and we will explore some of the similarities between Derrida and Peirce's treatment of language and signs.

Derrida is of the opinion that the emphasis on language and the subsequent shift of emphasis from language to writing are inevitable within the context of the history of Western

---

[1] Culler (1994:22) gives the following useful, if narrowly literary, description of the respective positions held by structuralism and post-structuralism:

> In simplest terms, structuralists take linguistics as a model and attempt to develop "grammars" – systematic inventories of elements and their possibilities of combination – that would account for the form and meaning of literary works; post-structuralists investigate the way in which this project is subverted by the workings of the texts themselves. Structuralists are convinced that systematic knowledge is possible; post-structuralists claim to know only the impossibility of this knowledge.

Culler stresses the pervasive links between the two positions and cautions that it is all but impossible to unambiguously divide theorists between the two camps (24-30).

[2] It is clear from his early work that Derrida is critical of structuralism in that he believes that it had introduced a damaging, and restricting methodology into the human sciences (1978: 6, 19-20; 26-28). As Johnson (1993:16) puts it,

> Derrida therefore criticises a certain kind of structuralism which adopts, consciously or unconsciously, a certain kind of geometry and a certain conception of time as the determining framework of its analysis. Though he links these two models (geometrical and temporal) via the example of Leibniz, his correctives of them are made separately, and are respectively 'force' and 'durée'. In fact, these correctives have the effect of promoting the models, as Derrida presents them, by one dimension.

[3] According to Johnson (1993:6), Derrida only alludes to the emergence of the information sciences that are contemporaneous to his "philosophy of writing" in several of his earlier works. Nevertheless, Derrida was aware of this "epistemic shift" in various disciplines, and considered it to be an inevitable result of what he diagnoses as the "end" of a certain metaphysics. It is for this reason that Johnson (1993:10) describes Derrida's philosophy as interacting with a wider contemporary episteme, and especially with contemporary life sciences.

philosophy.[4] He argues that the problem of language has never been "simply one problem among others" in that a "historico-metaphysical epoch *must* finally determine as language the totality of its problematic horizon" (1976:6).[5] Thus, Derrida views the semiological bent in twentieth century theory as an inevitable development, given the metaphysics that had traditionally dominated western thought. In recognising language as "the totality of its problematic horizon" the tradition comes to realise that our knowledge of the world cannot escape the play of language/writing. As will become clear, "writing" takes on a distinctive meaning in Derrida's work and it is important to understand what he means with this term, in that virtually his entire life's work can be related back to this concept. The subsequent discussion will focus extensively on Derrida's derivation of his conception of "writing".

It is imperative to emphasise that Derrida makes a distinction between language and "text". Derrida's argument does not as much relate to language as a system, as it does to the structure of writing (which he also refers to as text), as he conceives of it. Thus, despite his association with the linguistic turn in continental thought, Derrida's argument is *not* that "everything is language" or that "all philosophical questions boil down to problems of language" (as the analytic approach may be described). As Glendinning (2004:6) puts it: "…Derrida's thought, far from being part of a distinctively linguistic turn in philosophy, is actually working already beyond it." Derrida's concern is with the structure of "writing" that makes language possible in the first place (Derrida 1976:7). As such, his philosophy can be described as structuralist. His position can also be seen as a phenomenological one, and even an epistemological one, where language enables our access to the world and the structure of language can also be considered to be the structure of cognition (9). Thus his work is in step with the movement toward a semiological understanding of the world, where the world is approached in terms of information and codes (Johnson 1993:2-3). As Glendinning (2004:6) points out, Derrida considers the linguistic turn in philosophy to be symptomatic of a general *ailment* within Western metaphysics, rather than a philosophical

---

[4] Derrida gives a broad definition of "writing" and his conception of writing exceeds language (1976:8-9):

> And thus we say "writing" for all that gives rise to inscription in general, whether it is literal or not and even if what it distributes in space is alien to the order of the voice; cinematography, choreography, of course, but also pictorial, musical, sculptural "writing". One might also speak of athletic writing, and even with greater certainty of military or political writing in view of the techniques that govern those domains today. All this to describe not only the system of notation secondarily connected with these activities but the essence and the content of these activities themselves. It is also in this sense that the contemporary biologist speaks of writing and *pro-gram* in relation to the most elementary processes of information within the living cell. And, finally, whether it has essential limits or not, the entire field covered by the cybernetic program will be the field of writing.

[5] This "hystorico-metaphysical" epoch started with Plato and the shift to a transcendental realm of ideas, and was perpetuated by Aristotle and his definition of truth as correspondence (Norris 2004:17). The transcendental remained central to the entire subsequent Western philosophical tradition. In various works Derrida specifically refers to Descartes, Kant, Hegel, Nietzsche, Freud, Husserl, and Heidegger as heirs to this philosophical epoch.

advance.[6] Hence, the linguistic turn is inevitable because "everything that seemed solidly to render its status as essentially *un*problematic, everything that had assured us that it *is* what we thought it *should* be, namely the system of signification of an order of pure intelligibility (classical "meaning"), an order traditionally grasped in terms of the divine word or *logos*, has begun to melt into air (Glendinning 2004:6). The reasons for this "melting" can, among other things, be traced to the burgeoning of the sciences of mind, which were starting to undermine the traditional philosophical certainties, such as the primacy of mind, the accessibility of the mental through introspection, and the reliability of the cogito.

2. Deconstructing the Metaphysics of Presence

In his engagement with the western philosophical tradition, Derrida highlights this tradition's conception of, and preoccupation with, a certain 'metaphysics'. Derrida characterises metaphysics in this tradition as always assuming some form of *presence*, which leads him to speak of a *metaphysics of presence* (1978:279-280):

> The history of metaphysics, like the history of the West, is the history of these metaphors and metonymies. Its matrix – if you pardon me for demonstrating so little and for being so elliptical in order to bring me more quickly to my principle theme – is the determination of being as presence in all senses of this word. It would be possible to show that all the names related to the fundamentals, to principles, or to the centre have always designated the constant of presence – *eidos*, *archè*, *telos*, *energeia*, *ousia* (essence, existence, substance, subject), *aletheia*, transcendentality, consciousness, or conscience, God, man, and so forth.

The metaphysics of presence assumes that a primary or essential phenomenon such as being or mind or meaning is something that is present to the subject/agent/observer and, as such, can be contrasted to its opposite, which is characterised by the absence of presence and the possibility of accessibility. Such an assumption leads to a world that is viewed in terms of dichotomies – presence/absence, in/out, signifier/signified, male/female, transparent/ opaque, etc. In this understanding, disclosure of the truth is the presentation of the entity itself (e.g. Derrida 1986:6, 10). Often, these dichotomies are understood as the result of the

[6] Derrida sees systematic links between the traditional conception of the sign (and the resultant understanding of writing), and the metaphysics of Christian creationism "and infinitism", with its roots in classical Greek thought (Derrida 1976:13). Here the intelligible signified can be linked to the divine *logos* of a "creator God" (e.g. Peirce), or to the idea of the purely intelligible *logos* in general (Glendinning 2004:7). Hence, Derrida's assertion that "the sign and divinity have the same place and time of birth. The age of the sign is essentially theological (1976:14).

"natural" order of priority among phenomena and leads to normative judgements based on their perceived inevitability.[7] Usually, the second term in the dichotomy is considered to be a negative, corrupt or undesirable derivative of the first term (Johnson 1981:viii). As Critchley (1996:19) describes it, the Derridean practise of detecting the "metaphysics of presence" within the philosophical tradition can be assimilated to an anti-foundationalist critique of philosophy. Derrida (in the spirit of Nietzsche) cultivates a systematic mistrust of metaphysics although, as we shall see, his position is not that such metaphysical dichotomies can ultimately be overcome (cf. 1978:197).[8] In Critchley's formulation, there is a "foundationalist claim in deconstruction[9] which cannot be pragmatised". In terms of the objectives of this work, Derrida's position with regard to the metaphysics of presence is best exemplified by his analysis of language as a whole, and writing in particular, within this tradition.

Derrida is thus careful to emphasise that the "historico-metaphysical" epoch in which we find ourselves cannot but operate in terms of a dichotomous structure. Metaphysical dichotomies are not historical contingencies – he argues that their existence "was absolutely necessary" (Derrida 1976:7-8). Privileging terms such as the *phonė* (the spoken) over the written is inevitable because of its perceived presence. One can hear oneself speak, which inevitably leads to the conclusion that meaning is unproblematically conveyed through the spoken word (7-8):

> The privilege of the *phonė* does not depend upon a choice that could have been avoided. It responds to a moment of economy (let us say of the "life" of "history" or of "being as self-relationship"). The system of "hearing (understanding)-oneself-speak" through the phonic substance – which presents itself as the nonexterior, no mundane, therefore no empirical or no contingent signifier – has necessarily dominated the history of the world during an entire world-epoch, and has even produced the idea of the world, the idea of world-origin, that arises from the difference between the worldly and the non-worldly, the outside and the inside, ideality and nonideality, universal and nonuniversal, transcendental and empirical, etc.

This hierarchical way of thought is prevalent in a philosophical tradition which Derrida terms *logocentric* (1976:4, 12). Logocentric metaphysics assigns the origin of truth in general

---

[7] Derrida questions the assumption that there are "inherent" or "essential" as apposed to "accidental" or "improper" properties of phenomena (cf. Derrida 1997b:141).

[8] Rorty (1996:13) points out that Nietzsche and Derrida's suspicion of the western philosophical tradition is shared by the American pragmatists. As discussed in Chapter 1, Peirce was an important influence on the pragmatic tradition. This suspicion of received wisdom is shared by Chomsky in is scepticism towards common-senseexplanations of the world.

[9] As we shall see, Derrida approaches metaphysical oppositions by means of "deconstruction" in an attempt not to be completely caught up within the metaphysical structure.

to the *logos* (reason or thought) (Derrida 1976:3).[10] Thus, within a metaphysics of oppositions that which originates in the mind, the truth, takes precedence over and serves to suppress its opposite, falsehood (untruth/sham/pretence/perception, etc.), which originates in all that is not logos/mind.[11] Derrida describes this opposition of terms as a "violent hierarchy", where one term dominates the other (Derrida 1981b:41). Within the "epoch of the logos", "speaking" and "writing" are placed in such a hierarchical opposition, as well as the "signifier" and the "signified" (Derrida 1976:13). Derrida argues that within this epoch, these inherited oppositions are accepted as self-evident, even by those "who believe that the scientificity of their work begins where metaphysics ends" (13).[12] What Derrida aims to show is that what is historically seemingly self-evident within a given metaphysical tradition, need not be self-evident at all.[13] Situated within the logocentric tradition, Derrida does not propose to overcome or surpass this metaphysical stance, but proposes to disrupt it "from within". His position is that it is not, in fact, possible to "step out of" a given metaphysical position. He argues that "nothing is conceivable without a certain metaphysical stance". The most that one can hope to do is to identify the historical and systematic origins of certain metaphysical assumptions and to realise that these assumptions are not inevitable or necessary (13-14).

Thus, Derrida proposes to challenge the simple alternatives posited by logocentrism and seek an *economy* that escapes (at least some of) the impositions of the metaphysical system of thought (cf. Derrida 1978:19-20).[14] Derrida envisages the concept of economy as a

---

[10] Derrida (1976:11) describes the relationship between truth and logos in the metaphysical tradition as follows:

> All the metaphysical determinations of truth, and even the one beyond metaphysical onto-theology that Heidegger reminds us of, are more or less immediately inseparable from the instance of the logos, or of a reason thought within the lineage of the logos, in whatever sense it is understood: in the pre-Socratic or philosophical sense, in the sense of God's infinite understanding or in the anthropological sense, in the pre-Hegelian or the post-Hegelian sense

[11] In Derrida's words (1997b:93):

> …the enterprise of returning "strategically", in idealisation, to an origin or to a "priority" seen as simple, intact, normal, pure, standard, self-identical, in order then to conceive of derivation, complication, deterioration, accident, etc. All metaphysicians have proceeded thus, from Plato to Rousseau, from Descartes to Husserl: good before evil, the positive before the negative, the pure before the impure, the simple before the complex, the essential before the accidental, the imitated before the imitation, etc.

[12] In this regard we have seen that both Chomsky and Peirce seem unable to overcome the requirement of the primacy of mind in their attempts to formulate scientific accounts of mind and language.

[13] In his own words (Derrida 1976:13) he states that "to these metaphysical-theological roots many other hidden sediments cling." Hence (1976:13),

> The semiological, or more specifically, linguistic "science" cannot therefore hold on to the difference between signifier and signified – the very idea of the sign – without the difference between sensible and intelligible, certainly, but also without retaining, more profoundly and implicitly, and by the same token the reference to a signified able to "take place" in its intelligibility, before its "fall," before any expulsion into the exteriority of the sensible here below. As the face of pure intelligibility, it refers to an absolute logos to which it is immediately united. This absolute logos was an infinite creative subjectivity in medieval theology: the intelligible face of the sign remains turned toward the face and the word of God.

[14] A good definition of what Derrida means with the concept of 'economy' is given by Johnson (1993:20):

corrective for the deficiencies of structuralist assumptions (Johnson 1993:32-33) in that it would lead such theories to include the forces that constitute structures into their analyses of such structures. Derrida calls the strategy through which he aims to accomplish this metaphysical disruption *deconstruction*. Deconstruction works through a double gesture or "double writing" where it firstly reverses the classical opposition of given terms (speaking and writing, for example), and secondly proceeds to "displace" or disrupt such oppositions in order to gain a new perspective on them (Derrida 1976:14; Derrida 1997a:21). Deconstruction never overcomes the metaphysical hierarchy; it always takes place and remains in a certain metaphysics (Derrida 1976:19). However, it manages to disrupt the economy or the logic of that metaphysics, and thus to reveal what the metaphysics necessarily represses in its dichotomous structuring.[15] In this manner, Derrida envisages the strategy of deconstruction as providing the means for intervening in the field of oppositions which it criticises.[16] Derrida aims to show that logocentrism (unintentionally, but inevitably) simultaneously asserts and undermines the oppositions in terms of which it operates (7).

It is important to reiterate that deconstruction is not, as it has often been portrayed, a mere negation of philosophical ideas, nor is it a simple rejection of ideas that philosophy holds dear. Deconstruction is firmly situated within the western philosophical tradition and is an intimate engagement with the ideas and writings within this tradition; it is a tool by which philosophical text can be closely analysed and critiqued, with the aim of rendering a

---

> Derrida's 'economy' is a way of conceiving the more general, abstract notion of the dynamic interplay, or exchanges, between the elements of any system or complex. The economy harnesses a force without neutralising it, since force and location in a sense intersect in the economy (differences are both ('*à la fois*') spatial and dynamic), they cohabit in a necessary double bind.

Derrida envisages the concept of economy as a corrective for the deficiencies of structuralist assumptions (Johnson 1993:32-33) in that it would lead such theories to include the forces that constitute structures into their analyses of such structures.

[15] In his (Derrida 1997a:21)

> Very schematically: an opposition of metaphysical concepts (e.g. speech/writing, presence, absence, etc.) is never the confrontation of two terms, but hierarchy and the order of a subordination. Deconstruction cannot be restricted or immediately pass to a neutralisation: it must, through a double gesture, a double science, a double writing – put into practice a reversal of the classical opposition and a general displacement of the system. It is on that condition alone that deconstruction will provide the means of intervening in the field of oppositions it criticises…

[16] A useful and simple example of how deconstruction could work is given by Culler (1994:86-88) in Nietzsche's deconstruction of causality: In *The Will to Power* Nietzsche argues that the traditional conception of causality – that certain causes lead to certain effects – is, in fact, an inversion of the structure of causality, as it appears *to us*, at least. Hence, the principle of causality gives logical and temporal priority to a cause over its effect. Nietzsche argues that the principle of causality is not a result of experience, but rather of a chronological reversal of experience. If we feel pain, for example, we might look around to a discernable cause. If we happen to see a pin, we invert the *phenomenal* structure of events, and infer that the pin caused the effect of us feeling a pain. Phenomenologically speaking, the pain caused us to look for the pin, and we have to infer the causal sequence *pin … pain*. Nietzsche is not arguing that the principle of causality is illegitimate. His aim is rather to draw attention to the fact that the cause-effect hierarchy is the result of a phenomenological reversal that we are not necessarily aware of. In a manner of speaking, the cause is secondary to, or reliant on, the effect, *for us*. We do not directly perceive the principle of causality in the world, but logically construct it from experience. The aim is to gain a new perspective on a given principle, through using the logic of that same principle, and not by appealing to superior reason or a higher logical principle.

"positive", though provisional, philosophical position and a new perspective at the end of the process. The process of deconstruction is never complete, however, since otherwise one would run the risk of falling back into the metaphysics one has sought to disrupt.[17] The new perspective that has been provided by the deconstructive analysis of a certain opposition should in turn be deconstructed, and so on *ad infinitum* (Derrida 1981b:40-47).

Derrida formulates the deconstruction as a philosophical strategy in terms of a particular view of the understanding of knowledge (*episteme*) within the western philosophical tradition. It is not that the tradition itself is inevitable, but the implications of its structure and its assumptions (its metaphysics) are. Metaphysics, here, is understood to mean the conditions of possibility of a particular epistemology. As such, Derrida's position has much in common with Peirce's theory of abduction and Chomsky's appropriation of it and his subsequent critique of common-sense understanding. However, Derrida keeps within a more traditional understanding of metaphysics and he should not be understood as analysing the biological or physiological conditions that structure knowledge, but rather the traditional assumptions within a particular metaphysics that structure knowledge within that tradition. Hence, his theory has more in common with Chomsky's argument that science should be employed as a corrective for our common-sense understanding of the world. As we shall see, in terms of Derrida's view Peirce's ultimate recourse to theology in order to ensure the possibility of definite knowledge and universal truths is not an anomaly but an inherent part of his, and Derrida's, metaphysical tradition. Derrida does not take the same recourse, however. Instead of trying to ground ultimate, objective and certain truth in a phenomenon external to human cognition, Derrida settles for the "provisional" truth of meaning constituted within a system of differences However, ultimate truth, independent of human situatedness in the world, will always remain a metaphysical pipedream.

3. Writing – a Case Study and More

As we have already stated, language takes centre-stage in twentieth century philosophy and Derrida's philosophy is no different. His deconstruction of the logocentric spoken-language vs. written-language dichotomy is not only one of his most famous deconstructions, nor simply an interesting case study of deconstruction in action; it is the foundation of his distinction between *language* and *text* – a distinction that is fundamental to understanding his philosophy, and not merely relegating it to the realm of "literary criticism".

As we have already seen, Derrida sees the linguistic turn in philosophy as a historical inevitability within Western metaphysics. He considers this turn to be symptomatic of the realisation that the certainties provided by *logos* in the philosophical tradition have begun to

---

[17] As Spivak (1976: xiii) puts it:

> The most that can be said, and Derrida has reminded us to say it anew, is that a certain view of the world, of consciousness, and of language has been accepted as the correct one, and, if the minute particulars of that view are examined, a rather different picture … emerges.

waver. His position is that the logic of essence and identity, which had characterised traditional approaches to language and meaning, failed to account for the conditions of possibility of language and meaning (Patton 2004:28). In *Of Grammatology* (written in 1967) Derrida states (Derrida 1976:6-7):

> However the topic is considered, the problem of language has never been simply one problem among others. But never as much as at present has it invaded, *as such*, the global horizon of the most diverse researches and the most heterogeneous discourses, diverse and heterogeneous in their intention, method, and ideology…This inflation of the sign "language" is the inflation of the sign itself…Yet, by one of its aspects or shadows, it is itself still a sign: this crisis is also a symptom. It indicates, as if in spite of itself, that a historico-metaphysical epoch *must* finally determine as language the totality of its problematic horizon. It must do so not only because all that desire had wished to wrest from the play of language finds itself recaptured within that play, but because, for the same reason, language itself is menaced in its very life, helpless, adrift in the threat of limitlessness, brought back to its own finitude at the very moment when its limits seem to disappear, when it ceases to be self-assured, contained, and guaranteed by the infinite signified which seemed to exceed it.

It seems that Derrida makes the same connection between epistemology and the logic of the sign that Peirce makes. Derrida sees a connection between the philosophical conception of the sign and the metaphysics of Christianity and its appropriation of classical Greek thought (Glendinning 2004:7; Derrida 1976:13-14).[18] The connection lies in the conception of *logos*, in one sense or another, as a realm of pure perception or intelligibility, uncontaminated by worldly contingencies and the fallibility of the senses. The sign is conceived as consisting of a worldly and possibly fallible signifier, while the signified is the true and unambiguous meaning to which the signifier refers. The (intelligible) truth underlying (sensible) appearances or the meaning underlying the signifier is supposedly accessible through rationality, that ideal state of the mind which is not fooled by fallible worldly appearances. Logos is a necessary condition for true knowledge. Derrida links the signifier/signified distinction to the conception of logos and to the parallel idea of a creator God (Derrida 1976:13):

---

[18] "The difference between signified and signifier belongs in a profound and implicit way to the totality of the great epoch covered by the history of metaphysics, and in a more explicit and more systematically articulated way to the narrower epoch of Christian creationism and infinitism when these appropriate the resources of Greek conceptuality…And this distinction is generally accepted as self-evident by the most careful linguists and semiologists, even by those who believe that the scientificity of their work begins where metaphysics ends" (Derrida 1976:13).

> The semiological or, more specifically, linguistic "science" cannot therefore hold on to the difference between the signifier and the signified – the very idea of the sign – without the difference between sensible and the intelligible, certainly, but also not without retaining, more profoundly and more implicitly, and by the same token the reference to a signified able to "take place" in its intelligibility, before any expulsion into the exteriority of the sensible here below. As the face of pure intelligibility, it refers to an absolute logos to which it is immediately united. This absolute logos was an infinite creative subjectivity in medieval theology: the intelligible face of the sign remains turned toward the word and the face of God.

Just as God is present to the ultimate reality and *logos* is present to the realm of ideas, the signified should be "present" to ensure unambiguous meaning.[19] Derrida argues that logocentrism would conceive of the determination of the being of an entity as *presence* (Derrida 1976:12). Applied to language, such logocentrism is also a *phonocentrism* (11). In phonocentrism, spoken language (the *phonè*) is considered to be present to the *logos*. In this understanding, as is seen with Aristotle, spoken words are symbols of mental experience, with an immediate and essential proximity to the mind. Such proximity ensures an accurate portrayal of "thoughts" through *phonè* – spoken words are symbols of mental thoughts, which mirror things in the world by some sort of natural resemblance. And mental experiences are universal to all humans: "The feelings of the mind, expressing things naturally, constitute a sort of universal language… it is the stage of transparence" (11). Hence, Derrida illustrates that the distinction between written and spoken language is made in terms of the pervasive metaphysics of presence: "The formal essence of the signified is presence, and the privilege of its proximity to the logos as *phonè* is the privilege as presence. This is the inevitable response as soon as one asks: 'what is the sign?', that is to say, when one submits the sign to the question of essence" (Derrida 1976:18). Within the logocentric tradition, the signifier and the signified are differentiated on the grounds of their proximity to presence/logos/mind (13).

As opposed to spoken symbols, written symbols are a further derivation from pure presence, being the symbols of spoken words. Written symbols, therefore, are even further removed from logos than spoken symbols are, and are thus regarded with distrust. Derrida holds that this suspicion of writing spans the history of western philosophy, from Plato onwards.[20] This is the reason that Plato (1981) bans the poet from his ideal state; the poet

---

[19] Hence, Derrida (1976:14) states: "The sign and divinity have the same place and time of birth. The age of the sign is essentially theological. Perhaps it will never *end*. Its historical *closure* is, however, outlined." And (1976:98): "God is the name and the element of that which makes possible an absolutely pure and absolutely self-present self-knowledge."

[20] In his words(Derrida 1976:11):

deals in signifieds that are twice removed from logos, and hence are too far removed from the truth to be trusted. Derrida defines the written signifier as "always technical and representative" (Derrida 1976:11). The written signifier has no constitutive meaning, and is the origin of the general notion of the "signifier" (11).

Derrida proceeds to apply his analytical strategy of deconstruction to the metaphysical tradition's speaking-writing dichotomy. Not, as we have already seen, to negate these concepts or to reject the metaphysical tradition, but in order to "mark the conditions, the medium, and the limits of their effectiveness and to designate rigorously their intimate relationship to the machine whose deconstruction they permit"; and, in the same process, designate the crevice through which the as yet "unnameable glimmer beyond the closure can be glimpsed" (Derrida 1976:14). He suspects that the conception of the sign within the metaphysics of presence is systematically and genealogically determined by that history and that the sign can be understood differently within that same history (14). Such a different understanding may yield new insight into the nature of the sign, and hence into the nature of our knowledge, which is always mediated by signs. What Derrida will show with his deconstruction is that there cannot be a logos or meaning, an order of pure intelligibility, which is an ideal presence, existing prior to being signified (Glendinning 2004:7).

So, within the traditional western philosophical hierarchy, the written is secondary to the spoken, because it does not carry the guarantee of presence to the mind and its thoughts that the spoken does. Within this philosophical tradition, speech represents the internal, logic, reason, the transcendental, and truth, while writing represents the external, sophistry, the physical, the non-transcendental, and the possible misrepresentation of truth (Culler 1994:91). Furthermore, with speech the utterer is present to clarify any misunderstandings, while with writing the writer/utterer is absent and thus unable to rectify any misinterpretations

---

> All signifiers, and first and foremost the written signifier, are derivative with regard to what would wed the voice indissolubly to the mind or to the thought of the signified sense, indeed to the things itself (whether it is done in the Aristotelian manner…or in the manner of medieval theology, determining the *res* as a thing created from its *eidos*, from its sense thought in the logos or the infinite understanding of God.

And (Derrida 1981:109-110):

> And writing appears to Plato (and after him to all of philosophy, which is as such constituted in this gesture) as that process of redoubling in which we are fatally (en)trained: the supplement of a supplement, the signifier, the representative of a representative… The structure and the history of phonic writing have of course played a decisive role in the determination of writing as the doubling of a sign, the sign of a sign. The signifier of the phonic signifier. While the phonic signifier would remain in animate proximity, in the living presence of *mnēmē* or *psuchē*, the graphic signifier, which reproduces or imitate it, goes one degree further away, fall outside of life, entrains life out of itself and puts it to sleep in the type of its double…If it were purely external, writing would leave the intimacy or integrity of psychic memory untouched. And yet, just as Rousseau and Saussure will do in response to the same necessity, yet without discovering the *other* relations between the intimate and the alien, Plato maintains *both* the exteriority of writing *and* its power of maleficent penetration, its ability to affect or infect what lies deepest inside it. The *pharmakon* is that dangerous supplement that breaks into the very thing that would have liked to do without it yet lets itself at once be breached, roughed up, fulfilled and replaced, completed by the very trace through which the present increases itself in the act of disappearing.

of his meaning. The absence of the utterer, which writing implies, leads to the possibility that written signifiers might distort meaning; hence writing is considered secondary and undesirable, and even rejected. Within a metaphysics of presence, absence is undesirable, perhaps even dangerous (cf. Derrida 1997a:1-10).

Within Derridean terminology writing can be characterised as traditionally being conceived of as a *supplement* (Derrida 1976:144). His argument is that the logocentric philosophical tradition views writing and the production of texts in general as essentially being "secondary" activities (14). The argument is that writing is usually conceived of as having a structure similar to that of a supplement, subordinate to the truth and meaning that precedes it in the guise of the spoken.[21] Hence, writing is artifice, a human creation, and thus fallible (15-18). Within the logic of the supplement, writing becomes dangerous, because it indicates a possible deficiency in the phonological sign which it supplements. On the one hand it only enhances what it supplements: "The supplement adds itself, it is a surplus, a plenitude enriching another plenitude, the *fullest measure* of presence" (144-145). On the other hand, the supplement also supplements in the sense that it replaces, or fills a void: "Somewhere, something can be filled up of itself, can accomplish itself, only by allowing itself to be filled through sign and proxy. The sign is always the supplement of the thing itself" (145; 166).

In terms of the metaphysics of presence, Derrida argues that the process of supplementarity has always already infiltrated presence. In other words, representation arises from the desire to hide the inevitable absence of presence, or what Derrida calls *the abyss* (Derrida 1976:163). If the supplement adds nothing to that which it supplements, it should be useless and superfluous. If it does add something, it threatens the integrity or "purity" of that which it supplements, possibly contaminating it. Hence, designating something as a supplement and underplaying the importance of that supplement is an attempt at suppressing a lack in the supplemented. In the case of spoken language, it is the lack of presence that is hidden in viewing spoken language as a primary and writing as a secondary phenomenon. If one were to attempt to go back from the supplement to the source, one would find that "there is a supplement at the source" (304).[22] Derrida concludes that writing, which has traditionally been seen as a supplement, standing in for the originary presence of speaking (which has its source in the logos), turns out to be no more supplementary than spoken language itself. All of language exhibits the logic of the supplement. Writing takes place *before* and *within* speech (315), which means that all language (spoken and written) is structured according to the logic of writing.[23]

---

[21] Keep in mind that in French the word *supplément* has two meanings: "an addition" and "a substitute" (Johnson 1981:xiii).

[22] Derrida (1976:303-304) argues:

> The supplement comes in the place of a lapse, a nonsignified or a nonrepresented, a nonpresence. There is no presence before it, it is not preceded by anything but itself, that is to say by another supplement. The supplement is always the supplement of a supplement

[23] Again, Derrida demonstrates his semiological bent, where the world is understood in terms of signs (Derrida 1976:6-7):

Derrida proceeds to analyse the concept of "absence" (anathema to the metaphysics of presence) as it is generally applied to writing. According to the traditional understanding of writing, writing is an instrument that makes it possible to communicate with "persons who are absent" (Derrida 1997a:5-6). In the case of writing, the concept of absence is that of absence as "the modification of presence" (16). Hence, writing is traditionally thought of establishing a "presence" between the writer and the reader, eliminating the physical distance between the two parties. This conception seems straightforward enough. However, Derrida argues that the absence in question cannot merely be viewed as the literal absence of the receiver or reader of a piece of writing from its author (7-9, 17). If this were the case, such absence would in actual fact only be a case of deferred presence, and writing would not be different from spoken communication, except for a time-gap between the production and the perception of meaning. He points out that writing by its very structure is not limited to an exchange between specific parties, as spoken communication is. Writing can function in the event of "absolute absence"; a piece of writing can still have meaning even if the author and the intended reader were radically absent. Furthermore, a written communication must in principle be able to function in the absence of any specific readers (Derrida 1997a:7):

> In order for my "written communication" to retain its function as writing, i.e., its readability it must remain readable despite the absolute disappearance of any receiver, determined in general. My communication must be repeatable – iterable – in the absolute absence of the receiver or of any empirically determined collective of receivers. Such iterability…structures the mark of writing itself, no matter what particular type of writing is involved (whether pictographical, hieroglyphic, ideographic, phonetic, alphabetic, to cite the old categories). A writing that is not structurally readable – iterable – beyond the death of the addressee would not be writing.

Hence, in his deconstruction of the classical conception of writing Derrida identifies the following essential predicates in a minimal determination of writing (Derrida 1997a: 9-10):

Firstly, a written sign cannot "exhaust itself", or cease to function in the moment in which it is inscribed (this characteristic is in implicit opposition to the essential predicate of a spoken sign, where it is "lost" at the moment of being uttered; no physical reminder of it exists external to the minds of the utterer and hearer of the sign). Because the written mark is not lost, it can be iterated (repeated). Secondly, a written mark can break with its context.[24]

> By a slow movement whose necessity is hardly perceptible, everything that for at least twenty centuries tended toward and finally succeeded in being gathered under the name of language is beginning to let itself be transferred to, or at least summarised under, the name of writing. By a hardly perceptible necessity, it seems as though the concept of writing…is beginning to go beyond the extension of language…This, strictly speaking, amounts to destroying the concept of "sign" and its entire logic.

[24] Derrida defines "context" in this regard as: "the collectivity of presences organising the moment of its inscription (1997a:9).

Derrida is emphatic that the ability to break away from the context in which it is constituted is not an accidental property, but "the very structure of the written text". Context would include the presence of the writer, the writer's experience, as well as the intention the writers has when writing. Written signs can function in the absence of all of these factors. Thirdly, written signs operate in terms of "spacing" – "spacing…constitutes the written sign" (Derrida 1997a:9; Derrida 1976:203). Spacing separates the sign from other signs, from which it can be disengaged, and it separates the sign from its present reference, which makes the written sign possible in the first place. Derrida identifies a further level of rupture which applies to the written sign, in that any given sign can break away from the collection of signs in which it functions, without losing its possibility of signification. In other words, "no context can entirely enclose it" (Derrida 1997a:9). This observation will be very important in our discussion of Derrida's conception of the sign and signification in general.

Derrida's deconstruction leads him to the following two conclusions (1997a:7):

i. Since every sign, in whatever form (thus spoken of written) presupposes a certain absence, the absence within a particular field of writing will have to be "of an original type", and
ii. if it proves to be true that this "absence" is universal to all instances of signs and communication the result will be a general shift in our understanding of writing. Perchance the predicate thus introduced to characterise the absence peculiar and proper to writing were to find itself no less appropriate to every species of sign and of communication, the consequence will be a general shift; writing would no longer be one species of communication, and all the concepts to whose generality writing had been subordinated (including the concept itself qua meaning, idea or grasp of meaning and of idea, the concept of communication, of the sign, etc.) would appear to be non-critical, ill-formed, or destined rather, to insure the authority and the force of a certain historical discourse.

Derrida concludes that the structure of the spoken sign matches that of the written sign: "…the other of the signified is never contemporary, is at best the subtly discrepant inverse or parallel – discrepant by the time of a breath – of the order of the signifier" (Derrida 1976:18) Presence in all its metaphysical forms, presence to the logos or presence to meaning, is illusory. All signification takes place in terms of absence – the "absence" of both the utterer/writer and the receiver/reader. As Spivak (1976:xvi) points out: "Derrida seems to show no nostalgia for a lost presence. He sees in the traditional concept of the sign a heterogeneity…" It is the nostalgia for presence that causes the Western metaphysical paradigm to identify the spoken sign with the presence of the signified (in the logos). Derrida does not reject the notions of the signifier and the signified, nor those of presence and

absence. He does believe that it is possible to "overcome" the logocentric metaphysical tradition in which he finds himself. In fact, he goes as far as to hold that, at present at least, nothing is conceivable outside of the notions of signifier and signified (Derrida 1976:13). As such, his deconstruction is an attempt to gain new insight into the structure of signifier and signified and to re-evaluate their interaction and interdependence. As we shall see, he comes to the conclusion that there can be no linguistic sign prior to writing (14). And the structure of writing, far from being secondary and supplementary to spoken language, turns out to exhibit the very conditions of the possibility of signification in general.

4. *DIFFÉRANCE*, TRACE, ARCHI-WRITING

Derrida's deconstruction of the spoken-written dichotomy leads him to conclude that all forms of signification are structured according to the logic of writing, and hence he extends the concept of writing to refer to all instances of signification. The fact that these two structures are identical has traditionally been successfully masked by the illusion of presence that the context in which speech takes place allows. However, for Derrida, presence is always an illusion. Derrida argues that for presence to be possible at all, it needs to function according to the structure usually attributed to absence. This insight pertains to his analysis of the concept of the sign in that he concludes that thought, object, and sign are never identical. Derrida thus concludes (as does Saussure) that it is not identity that allows for the possibility of reference in the sign, but *difference*. "The sign marks a place of difference" (Spivak 1976:xvi).

Derrida's conception of the role that difference plays in establishing *meaning* can be explained in terms of his relationship with structuralism. Derrida is the heir of the Saussurean insight that the relationship between the signifier and the signified is arbitrary (Saussure 1983:116-118).[25] Saussure's (1983:118 ff) contention is not that the relationship between the two is entirely random, but that the relationship is not based on any "natural" correlation between the two. The signifier-signified relationship is entirely conventional, and hence dependent on the system of signification within which they find themselves.[26] The "arbitrariness" or "unmotivatedness" (Derrida 1976: 47) of the sign implies that the meaning of a sign does not originate in its correlation to the signified, but in its difference from other signs

---

[25] Derrida notes, however, that this insight can already be seen in Plato's *Sophist* (Derrida 1976:53).

[26] To quote Saussure (1983:118) on this point:

> Everything we have said so far comes down to this. In the language itself, there are only differences. Even more important than that is the fact that, although in general a difference presupposes positive terms between which the difference holds, in a language there are only differences, and no positive terms. Whether we take the signification or the signal, the language includes neither ideas nor sounds existing prior to the linguistic system, but only conceptual and phonetic differences arising out of the system. In a sign, what matters more than any idea or sound associated with it is what other sounds surround it. The proof of this lies in the fact that the value of a sign may change without affecting either meaning or sound, simply because some neighbouring sign has undergone change.

within the linguistic system.[27] Saussure's insight serves to undermine the traditional "phonologism" and the privilege it accords to speech over writing (53)[28]:

> By definition, difference is in itself never a sensible plenitude. Therefore, its necessity contradicts the allegation of a naturally phonic essence of language. It contests by the same token the professed natural dependence of the graphic signifier. That is a consequence Saussure himself draws against the premises defining the internal system of language. He must now exclude the very thing which permitted him to exclude writing: sound and its "natural bond" with meaning.

Derrida maintains the structuralist focus on the structure of phenomena, but radicalises the concept of structure and its implications, as well as the concept of difference. He argues that the source of linguistic value (meaning) lies in difference (Derrida 1976:53), and coins the term *différance* to illustrate his conception of how meaning is structured though difference (Derrida 1976:23; Derrida 1986:3-27). Phonologically, the term *différance* is indistinguishable from its French root-concept, *différence*, hence illustrating Derrida's intention with the concept while emphasising the difference between the meanings of the two concepts.[29] *Différance* encompasses three possible meanings; "difference-differing-deferring" (Derrida 1976:66-67; Derrida 1986:3-27). Hence it can refer to i) the passive differences within a system of signification, as characterised by Saussure; ii) the act of differing, or the systematic play of differences within a system of signification, and iii) deferral, where meaning can never be present and is thus always deferred, always to come. With this new concept, Derrida emphasises that the language system is not a passive structure, but one of dynamic movement. Thus, meaning is continuously being established and re-established and never fixed. Meaning is transitional and transient. This leads him to assert (Derrida 1981b:27):

---

[27] Derrida (1981b:28) explains his position as follows:

> Since there is no presence before and outside semiological *différance*, one can extend to the system of signs in general what Saussure says about language: "Language is necessary for speech to be intelligible and produce all its effects; but speech is necessary for language to be established; historically, the fact of speech always comes first." There is a circle here, for if one rigorously distinguishes language and speech, code and message, schema and usage, etc., and if one wishes to do justice to the two positions thus enunciated, one does not know where to begin, nor how something can begin in general, be it language or speech. Therefore, one has to admit, before any dissociation of language and speech, code and message (and everything that goes along with such a disassociation), a systematic production of differences, the *production* of a system of differences – a différance – within whose effects one eventually by abstraction and according to determined motivations, will be able to demarcate a linguistics of language and a linguistics of speech, etc.

[28] As we shall see, Derrida makes use of this Saussurean insight and revises it with Peirce's semiotics.

[29] Derrida also defines *différance* as follows: "an economic concept designating the production of differing/deferring" (Derrida 1976:23).

> …différance, then, is a structure and a movement no longer conceivable on the basis of the opposition presence/absence. Différance is the systematic play of differences, of the traces of differences, of the *spacing* by means of which elements relate to one another. This spacing is simultaneously active and passive (the *a* of différance indicates this indecision as concerns activity and passivity, that which cannot be governed by or distributed between the terms of this opposition) production of the intervals without which the "full" terms would not signify, would not function.

Hence, with his amendment of difference to différance Derrida emphasises that différance is neither active, nor passive (Derrida 1986:9). It incorporates temporisation and spacing into the referential process (Derrida 1981b:27). Derrida argues that différance is what makes the "presentation of the being-present possible" (Derrida 1986:6)."Presence" is an effect of the system.[30] The sign "stands for" meaning or the referent; hence it "represents the presence in its absence. It takes the place of the present" (9). The sign (verbal, written, monetary, political presentation, etc.) is deferred presence in that it defers the moment in which we can "encounter the thing itself". This is what Derrida means with signification as the différance of temporisation. Furthermore, Saussure's principle of difference as the condition for signification affects the sign both as signifier and as signified (10). The signified concept is never present to the logos; every concept is constituted within a system in which it refers to other concepts by means of a systematic play of differences (10). Différance is the structural possibility in language that makes the "presentation" of both the signifier and the signified possible (6).[31]

Derrida (1986:6) cautions against conceiving of différance as an existence or an essence.[32] Différance is "the possibility of conceptuality, of a conceptual process and system in general" (11). In a system of language, there are only differences and the play of differences, and these differences are in themselves effects of the system. Derrida is adamant that they "have not fallen fully formed from the sky", nor are they "prescribed in the grey matter of the brain" (11). In this understanding, it becomes impossible to identify the

---

[30] Derrida states in this regard (Derrida 1986:16-17):

> Thus one comes to posit presence – and specifically consciousness, the being beside itself of consciousness – no longer as the absolutely central form of Being but as a "determination" and as an "effect." A determination or an effect within a system which is no longer that of presence, but of différance, a system that no longer tolerates the opposition of activity and passivity, nor that of cause and effect, or of indetermination and determination, etc., such that in designating consciousness as an effect or a determination, one continues – for strategic reasons that can be more or less lucidly deliberated and systematically calculated – to operate according to the lexicon of that which one is delimiting.

[31] "Retaining at least the framework, if not the content, of this requirement formulated by Saussure, we will designate as *différance* the movement according to which language, or any code, any system of referral in general, is constituted "historically" as a weave of differences" (Derrida 1986:12).

[32] "*Différance* is not only irreducible to any ontological or theological – ontotheological – reappropriation, but as the very opening of the space in which ontotheology – philosophy – produces its system and its history, it includes ontotheology, inscribing it and exceeding it without return" (Derrida 1986:6).

origin of meaning; the argument is that the "playing movement" that allows for the possibility of the differences that make up the language structure is one of différance.[33] Hence, Derrida rejects the idea that the differences in language can be attributed to a subject, or a substance, "a being somewhere present" that can be pinpointed as a cause or an origin of language. And if it is impossible to point to origins in language in meaning, it becomes difficult to account for causes and effects within the system.

Derrida argues that a way to overcome the closure of a traditional cause-and-effect framework with regard to language, and to formulate a concept of cause and effect appropriate to a system of différance, is to adopt the idea of the *trace*. A trace is neither a cause nor an effect, but the condition of possibility for signification (Derrida 1986:12).[34] Hence, the movement of signification is possible because of différance, in that every element (*trace*) present to a signification is related to elements other than itself. Elements, or traces, always incorporate elements or traces of those signs to which they were related in the past, and traces to which they can be related in the future. The "present" is structured in terms of (or constituted from) a synthesis of marks or traces, both retentive and "protentive".[35] Any sign is understood in terms of signs to which it was related in the past, and in terms of other signs it may conceivably be related to in the future. Derrida envisages that signs can thus be seen in terms of a relational system of traces – elements primary to signs which relate signs not only to other concurrent signs in the system of language, but also to past and future instances of those signs. In a way, it can be argued that Derrida conceives of signs as being

---

[33] *Différance* is the non-full, non-simple, structured and differentiating origin of differences. Thus the name "origin" no longer suits it (Derrida 1986:11).

[34] In *Of Grammatology* Derrida describes the *trace* as follows (1976:63):

> Without a retention in the minimal unit of temporal experience, without a trace retaining the other as other in the same, no difference would do its work and no meaning would appear. It is not the question of a constituted difference here, but rather, before all determination of the content, of the pure movement which produces difference. *The (pure) trace is différance*. It does not depend on any sensible plenitude, audible or visible, phonic or graphic. It is, on the contrary, the condition of such a plenitude. Although it *does not exist*, although it is never a *being-present* outside of all plenitude, its possibility is by rights anterior to all that one calls sign (signified/signifier, content, expression, etc.), concept or operation, motor or sensory. This différance is therefore not more sensible than intelligible and it permits the articulation of signs among themselves within the same abstract order – a phonic or graphic text for example – or between two orders of expression.

[35] To quote Derrida (1986:13):

> An interval must separate the present from what it is not in order for the present to be itself, but this interval that constitutes it as present must, by the same token, divide the present in and of itself, thereby also dividing, along with the present, everything that is thought on the basis of the present, that is, in our metaphysical language, every being, and singularly substance or the subject. In constituting itself, in dividing itself dynamically, this interval is what might be called spacing, the becoming-space of time or the becoming-time of space (*temporization*). And it is this constitution of the present, as an "originary" and irreducibly nonsimple (and therefore, *stricto sensu* nonoriginary) synthesis of marks, or traces of retentions and protentions (to reproduce analogically and provisionally a phenomenological and transcendental language that will soon reveal itself to be inadequate), that I propose to call archi-writing, archi-trace, or différance. Which (is) (simultaneously) spacing (and) temporisation.

constituted within a system of traces, which incorporates a temporal dimension into the process of signifying.[36] Derrida calls his new concept of writing "archewriting" (Derrida 1976:56) in an attempt to emphasise that language "prior" to writing (in his new conception of writing) is an impossibility.[37] Meaning prior to traces is impossible. Hence he concludes that language is a *form* and not a substance (57).[38]

5. Meaning and Context

We have seen that Derrida problematises the idea of an originary, phonic language and deconstructs this conception of language in order to show that the essential principles of written language are in fact the essential (enabling) principles of all language. Language is constituted through a certain structure, which operates in terms of différance. Inextricably linked to the function of language is meaning[39] and we have seen that Derrida argues that meaning is constituted through différance and traces, which "permits the articulation of signs", hence "permits the articulation of speech and writing" (Derrida 1976:62-63). Derrida argues that the products of language are *texts*, which are systems and chains of traces (65). In terms of his deconstruction, "texts" should not be taken to mean "written texts" in the traditional sense, but any entity formed within (or, perhaps more accurately from) the structure of language (Derrida 1981a:5).[40] It should be kept in mind that Derrida has a very broad (semiological) conception of language, designating any process of signification and interpretation as an instance of "language". Within the subjective, lived experience of human

---

[36] In Chapter 5 we will see how Deacon conceives of the constitution of meaning in similar terms, although it can be argued that Deacon formulates his theory in a much more straightforward manner than Derrida does here.

[37] "An archewriting whose necessity and new concept I wish to indicate and outline here; and which I continue to call writing only because it essentially communicates with the vulgar concept of writing. The latter could not have imposed itself historically except by the dissimulation of the arche-writing, by the desire for a speech displacing its other and its double and working to reduce its difference. If I persist in calling that difference writing, it is because, within the work of historical repression, writing was, by its situation, destined to signify the most formidable difference. It threatened the desire for the living speech from the closest proximity, it *breached* living speech from within and from the very beginning" (Derrida 1976:57).

[38] Derrida's argument can be put in terms of systems theory, as is done by Johnson (1993:49-50):

> The classical conception sees the structure as possessing a definite, fixed centre which, paradoxically, is both inside and outside of that structure. The paradox arises from the fact that, while the centre permits a combinatorial play of elements within the structure, it commands and limits that play and does not as such enter into the process of combination and substitution. This means that the system, or structure, is 'closed', as Derrida puts it. The traditional (and reassuring) conception of the closed structure is, however, disturbingly questioned when it is recognised that there is not necessarily any one privileged centre for a given structure, and that the play of substitution is therefore an infinite process. This would mean a certain 'opening' or 'openness' of the system, totality or structure … an openness whose condition of possibility is, however, a certain closure.

[39] As we have seen, Chomsky disregarded meaning as a topic that was "too difficult" for our current understanding. Derrida's position is that meaning and language structure cannot be separated.

[40] Hence, Derrida's infamous assertion that "there is nothing outside the text" (Derrida 1976:158) should be understood as saying that we understand the world through language; a rather less foreign philosophical position than "everything is writing" or some such interpretation.

beings (which is neither "internal" nor "external" to the mind) differences between elements in the world make them apparent to us and "constitute the *texts,* the chains, and the systems of traces" (65).[41] Derrida argues that it is the difference between "world" and "experience" that is the condition of all differences in general; of all other traces (Derrida 1976:65)[42]:

> *The trace is in fact the absolute origin of sense in general. Which amounts to saying once again that there is no absolute origin sense in general. The trace is the difference* which opens appearance and signification. Articulating the living upon the nonliving in general, origin of all repetition, origin of ideality, the trace is not more ideal than real, not more intelligible than sensible, not more a transparent signification than opaque energy and no concept of metaphysics can describe it.

Derrida is often understood as making an ontological claim here, conflating linguistics and ontology (a common semiological move as was demonstrated by Peirce in Chapter 1) to claim that "the world is language" or "everything is writing". But such interpretations overlook the fact that Derrida is in fact making a *phenomenological* claim: *our* understanding/experience of the world within space and time is structured in terms of language and understood as written language. Human beings (and all sentient organisms) understand the world through interpretation (Derrida 1976:70-71):

> If the trace, arche-phenomenon of "memory," which must be thought before the opposition of nature and culture, animality and humanity, etc., belongs to the very movement of signification, then signification is a priori written, whether inscribed or not, in one form or another, in a "sensible" and "spatial" element that is called "exterior." Arche-writing, at first the possibility of the spoken word, then of the "*graphie*" in the narrow sense, the "usurpation," denounced from Plato to Saussure, this trace is the opening of the first exteriority in general, the enigmatic relationship of the living to its other and of an inside to an outside: spacing. The outside, "spatial" and "objective" exteriority which we believe we know as the most familiar thing in the world, as familiarity itself, would not appear without the grammè, without difference as temporalisation, without the

---

[41] In terms of the logic of the supplement Derrida formulates his position as follows (Derrida 1976:159):

> …there has never been anything but writing; there have never been anything but supplement, substitutive significations which could only come forth in a chain of differential references, the 'real" supervening, and being added only while taking on meaning from a trace and from an invocation of the supplement, etc.

Again, Deacon makes a similar argument (Chapter 5) but in decidedly more prosaic terms.

[42] Compare this to Peirce's conception of "Secondness" as manifesting itself when we "come up against" a reality of the world that confounds our expectation in terms of Firstness (cf. Chapter 1).

nonpresence of the other inscribed within the sense of the present, without the relationship with death as the as the concrete structure of the living present.

Derrida is not making an ontological claim beyond the subjective experience of living in the world as a human being. In this sense it can be argued that he conflates the structuralist and phenomenological projects. His argument is thus that: language is constituted by traces, and traces are the absolute origin of sense *in general*. The meaning of the concept of language is extended to include all activities of imposing meaning and structure upon our experience of the world; it imposes an interpretable order on the stimuli that may otherwise be a chaotic mass of meaningless "traces". This meaningful structure can be "articulated" and communicated to other human beings, who presumably share a similar phenomenological make-up and language.

Derrida uses this conclusion to criticise modern linguistics and structuralism. Linguistics after Saussure views the signifier as a trace, but persists in thinking of the signified as having a meaning "thinkable in principle within the full presence of an intuitive consciousness" (Derrida 1976:73).[43] In this understanding, the signified can be separated from the signifier and does not need the signifier "to be what it is". In other words, the signified remains, in principle, objectively accessible in the "external world" independently from human cognition. Derrida argues that this insistence on the sanctity of the signified, which is "thinkable" outside of all signifiers, remains dependent on the *onto-theology* of our metaphysical tradition.[44] He believes his deconstruction of language and the sign undoes this onto-theology and exposes its insecure foundations.[45] Signification (meaning) in general is revealed to be dependent upon human situatedness in the world and the contingencies of our phenomenological (cognitive) structure. There can be no external, transcendental assurances in terms of which we can substantiate knowledge and meaning. It is perhaps discomfort with this uncertainty that led to Chomsky's difficulty in accounting for meaning and Peirce's seemingly incongruous grounding of his semiotic in the existence of God. Derrida evokes a universe where absolute objectivity and absolutely certain knowledge are impossible. And this world view, however well-substantiated, is an abomination in our metaphysical tradition.

---

[43] As with Chomsky, Derrida compares the "mechanical" treatment of text by the structuralists to the science and metaphysics of the seventeenth century (Derrida 1976:16; Johnson 1993:14). He relates his criticism of structuralism to Leibniz's criticism that Descartes ignored force, when he attempted to explain natural phenomena, confusing it with "quantity of movement" (Derrida 1986: 16). Structuralism, as part of a particular theoretical lineage, assumes a *metaphysics of presence*, as does its predecessor, phenomenology (26-27). Derrida accuses structuralism of favouring the simultaneous and suggests the concepts of "force" and "durée" as correctives for to the structuralist approach (4, 27).

[44] Derrida borrows the term "onto-theological" from Heidegger, who borrows the term from Kant, and uses it to characterise the highly determined way in which Western metaphysics is structured (Hart 2004:54). Kant conceives of onto-theology as a way of theoretically conceiving of God, i.e. in terms of being. In this conception God is the *highest* and the *original* being; he is the "*being* of all being" (54).

[45] "One is necessarily led to this from the moment that the trace affects the totality of the sign in both its faces. That the signified is originarily and essentially (and not only for a finite and created spirit) trace, that it is always already in the position of the signifier, is the apparently innocent proposition within which the metaphysics of the logos, of presence and consciousness, must reflect upon writing as its death and its resource" (Derrida 1976:73).

Derrida's argument is not one of relativism, however. He does not argue that signification and meaning are completely subjective and irreducibly arbitrary. In fact, within his theory, despite the absence of an onto-theological grounding principle, meaning can be more or less determined. This is because of the structural character of the signifier, which is both embedded in, and constituted by, a *system* of signifiers.[46] A signifier outside of such a system is not a signifier at all. The iterability (repeatability) of the signifier and the fact that it is constituted within a system of differences (traces) highlights the importance of *context* in establishing meaning. Derrida extensively engages with the importance of context in terms of Austin's *speech act* theory, and again in reply to Searle's criticism of his reading of and commentary on Austin (cf. *Signature, Event, Context* (1997a) and *Limited Inc.* (1997b)).

In *How to do Things with Words*, Austin (1980:8, 11) argues that context determines the meaning of utterances, in that the intention of the speaker can provide a clue as to the meaning of an utterance. With regard to how meaning relates to context in general, Derrida (1997a:2), however, argues that while a word may seem to be understood in terms of a context of convention, the context is always only "implicit and structurally vague." This raises the question of whether the context is ever determinable. As might be expected, Derrida argues that a context can never be completely determinable, or more specifically that the determination of a context can never be saturated (3):

> Is there a rigorous and scientific concept of context? Or does the notion of context not conceal, behind a certain confusion, philosophical presuppositions of a very determinable nature? Stating it in the most summary manner possible, I shall try to demonstrate why a context is never determinable, or rather, why its determination can never be entirely certain or saturated.

If Derrida's conception of context is correct, it would mean (in terms of his analysis of writing in the general sense) that meaning can never be completely determined. We have seen that Derrida argues that the "absence" (of the addressee and the writer) that is traditionally associated with the logic of writing, is in fact implicit in the structure of language as a whole. We are no more "present" to the meaning of spoken language than we are to that of written language – interpretation of meaning takes place in both instances. Moreover, the process of interpretation is the same process in both instances. Derrida argues that the possible absence of the writer (utterer) implies that his communication must be "repeatable" (*iterable*) in a situation completely different from the one in which it is uttered. Hence writing (language) is organised by a code; it is constituted in its identity as a mark by its iterability, in the absence of "every empirically determined 'subject'" (Derrida 1997a:8-9).[47]

---

[46] "…the person writing is inscribed in a determined textual system" (Derrida 1976:160). Both "writing" and "textual system" should be understood in the general sense as formulated by Derrida.

[47] "The possibility of repeating and thus of identifying the marks is implicit in every code, making it into a network that is communicable, transmittable, decipherable, iterable for a third, and hence for every possible user in general. To be what it is, all writing must, therefore, be capable of functioning in the

Derrida likens writing to a machine, which can, in principle, continue to function in the absence of its creator. With his emphasis on non-presence, Derrida emphasises the absence of the *intention* with which the speaker might have spoken. Writing must continue to function even if the "intention" of its speaker cannot be verified. In keeping with his critique of metaphysics of presence, Derrida argues that the speaker/writer is never absolutely and unambiguously "present" to his meaning in the way that it is envisioned in the metaphysics.[48] The situation of the "writer" and the "addressee" are the same with regard to the written text – they can both interpret and reinterpret the text and arrive at a slightly different understanding with every understanding. Hence, Derrida (1997a:8) argues that writing, as an iterative structure, exhibits an *essential drift* "cut off from all absolute responsibility, from consciousness as the ultimate authority, orphaned and separated at birth from the assistance of its father." A given interpretation of a text can never be guaranteed. Derrida holds that it is this essential drift that leads Plato to condemn writing in the *Phaedrus*.

If the meaning of a sign were completely determinable within the context in which it is uttered, the sign cannot be interpretable within another context, because the possibility that exactly the same context will be recreated with each iteration is virtually nil. Derrida has shown that a sign breaks with its context ("with the collectivity of presences organising the moment of its inscription") – this is an essential characteristic which enable written signs to function as "writing", and we have seen that he extends the structure of writing to all language. "Context" in this regard is also extended to include the "internal semiotic context", where any given sign can be removed from the given chain in which it is inserted at a particular time, and it would still be able to function. "One can perhaps come to recognise other possibilities in it by inscribing it or grafting it onto other chains" (Derrida 1997a:9). No sign can be entirely enclosed by a particular context. Furthermore, a sign can function in the absence of a referent, seeing that it does not require a fixed context in order to be interpretable (11-12). Every sign can still function, even if it is cut off from the original "intention" with which it was uttered, especially because every sign can be *cited* (Derrida 1997a:12). A cited sign can break with any given context and cited in a potentially infinite arrays of new contexts.

Hence, Derrida argues that no signifier (written or spoken) has a unique and singular reality (Derrida 1976:91). A signifier can only function as such if it is *repeatable* and in order for writing to function as writing, a written sign needs to be repeatable in a different context. Such iterability is the only way in which we can have more than one user of the structure of

---

radical absence of every empirically determined receiver in general. And this absence is not a continuous modification of presence, it is a rupture in presence, the 'death' or the possibility of the 'death' of the receiver inscribed in the structure of the mark…The perhaps paradoxical consequence of my here having recourse to iteration and to code: the disruption, in the last analysis, of the authority of the code as a finite system of rules; at the same time, the radical destruction of any context as the protocol of code" (Derrida 1997a:8).

[48] "…he has not employed his absolutely actual and present intention or attention, the plenitude of his desire to say what he means, in order to sustain what seems to be written "in his name" (Derrida 1997a:8).

language (Derrida 1997a:12). Derrida's argument is thus that iterability is central to all instances of signification. Iterability is a precondition for signification; a sign must be recognisable as "the same" in a different context. However, he does not draw the (perhaps tempting) conclusion that a sign must have a fixed, pure, and readily intelligible meaning that is somehow linked to it and which can be transferred between contexts. On the contrary, he draws the conclusion that the meaning of a sign is always dependent on *a* context (Derrida 1997a:12):

> This does not imply that the mark is valid outside of a context, but on the contrary that there are only contexts without any centre or absolute anchoring. This citationality, this duplication or duplicity, this iterability of the mark is neither an accident nor an anomaly, it is that (normal/abnormal) without which a mark could not even have a function called "normal." What would a mark be that could not be cited? Or one whose origins would not get lost along the way?

It should be kept in mind that Derrida extends his conclusion with regard to writing to all experience in general (Derrida 1997a:10). He argues that there is no experience that consists of "pure" presence, but only of "chains of differential marks" (10):

> Are these three predicates, together with the entire system they entail, limited, as is often believed, strictly to "written" communication in the narrow sense of this word? Are they not to be found in all language, in spoken language for instance, and ultimately in the totality of "experience" insofar as it is inseparable from this field of the mark, which is to say from the network of effacement and of difference, of units of iterability, which are separable from their internal and external context and also from themselves, inasmuch as the very iterability which constituted their identity does not permit them ever to be a unity that is identical to themselves?

One of the main criticisms leveled at Derrida with regard to his contention that meaning relies on an indeterminable context is that his theory inevitably leads to relativism, where meaning becomes a radically subjective affair and truth becomes an impossibility. One such criticism is that of Eco, who specifically accuses him of misappropriating Peirce in order to substantiate his view of language as an "unstructured drift of indeterminable meanings". We will use Eco's criticism in order to illustrate that the accusation of relativism is a fatuous one, which overlooks some of Derrida's central arguments. It is also useful to compare Derrida's position to that of Peirce, in order to situate his project within a broader perspective.

6. PEIRCE, DRIFT, TRANSCENDENTALISM AND THE POSSIBILITY OF KNOWLEDGE

There are many parallels between Derrida's treatment of the structure of the sign and its relation to thought and the work of Peirce.[49] Derrida explicitly draws on some of Peirce's work, as can be seen in *Of Grammatology*, although there is an argument to be made that Derrida (perhaps unknowingly) owes a greater debt to the work of Peirce than he acknowledges. As it is, Derrida (1976:48) credits Peirce with greater insight than Sausurre into the necessary structure of the sign (the symbol in Peirce's terminology) that is implied by the rejection of a simple theory of correspondence between the signifier and its signified. He quotes Peirce's argument that symbols grow out of other signs (icons and indices), and that we think only in signs. He reads Peirce (correctly in my opinion) as arguing that there is no ground beyond the structure of signs to give foundation to their meaning. Hence, he agrees with Peirce's contention that semiotics does not depend on logic, but that logic is a semiotic (48). He sees Peirce's argument as going "very far in the direction that I have called the de-construction of the transcendental signified", justifying the conclusion that "(f)rom the moment that there is meaning there are nothing but signs. We think only in signs" (49). Derrida agrees with Peirce that in philosophy, the *theory of things* should be reduced to the *theory of signs*. Furthermore, Derrida argues that Peirce's semiotic demonstrates that the logic of the trace can be understood as an "active movement" and not as a "given structure" (51).

Eco (1990:34) criticises Derrida's appropriation of Peirce as an authority "to legitimise his attempt to outline a semiosis of infinite play, of difference, of the infinite whirl of interpretation." He understands Derrida's theory as one of *infinite interpretive drift*, while he is adamant that Peirce's theory of "Unlimited Semiosis" does not support such a conclusion (28):

> Can we really speak of the unlimited semiosis apropos of the Hermetic ability to shift from term to term, or from thing to thing? Can we speak of unlimited semiosis when we recognise the same technique implemented by contemporary readers who wander through texts in order to find in them secret puns, unheard-of etymologies, unconscious links, dances of "Slipping Beauties," ambiguous images that the clever reader can guess through the transparencies of the verbal texture even when no public agreement could support such an adventurous misreading?

This criticism is levelled at theories of "Hermetic drift", which Eco distinguishes from the "other form of drift" – deconstruction (Eco 1990:32). It is worth examining Eco's argument

---

[49] Significantly, Peirce shares Derrida's suspicion of a "dualistic" world-view (Peirce 1965 Vol. I:26 [1.61]):

> One of the worst effects of the influence of moral and religious reasonings upon science lies in this, that the distinctions upon which both insist as fundamental are dualistic distinctions, and that their tendency is toward ignoring all distinctions that are not dual and especially of the concept of continuity.

with Hermetic drift, because many of Derrida's critics imagine deconstruction to be something that allows for "random" and "drifting" interpretations of texts. This view points to the prevalent misconception that Derrida advocates a relativist position where all meaning is relative and no interpretation carries more weight than any other. Eco carries this misconception over into his reading of Derrida's appropriation of Peirce, and consequently rejects Derrida's interpretation of Peirce's theory. He implies that Derrida has no intention of interpreting Peirce "correctly" given that his (Derrida's) theory does not allow for the possibility of "correct" interpretations (35).[50] Eco's misunderstanding of Derrida leads him to criticise the use that Derrida makes of Peirce's semiosis.

Eco understands Peirce's "Unlimited Semiosis" (his theory that an interpretant is always another representation, which has its interpretant and so forth) as implying that "a sign is always something by knowing which we know something *more*" (Eco 1990:28). In other words, given infinite or unlimited semiosis within a community of interpreters, the interpretation will eventually come to approximate the final logical interpretant. And, Eco adds, at a certain stage of the interpretation "we know more about the content of the representamen which started the interpretive chain" (28). He continues (quoting Peirce) (28):

> To know more does not mean to know everything, but it means that a sign entails all its remote illative consequences and the meaning of a proposition embraces every obvious necessary deduction (1935:5.165).

As opposed to knowing more, Eco argues that *Hermetic* semiosis entails that: "A sign is something by knowing which we know something *else*" (28). This understanding implies a random drift in interpretation, where a "freeflow" (sometimes associated with Derrida's concept of "freeplay" of traces) of ideas lead haphazardly from one to the next, without a discernable structure and hence resulting in random interpretations of equal validity. In elaborating on Peirce's conception of unlimited semiosis, Eco argues that Peirce's approach allows for "knowing more" about a sign, because we are given grounds for knowing its object (in the Peircean sense), namely the "point of view of a given context" (28). As we saw in the chapter on Peirce's semiotic, he explains representation by insisting that the sign always means something *to somebody* and that the context in which an interpretation takes place forms the crucial third branch of his *object-sign-interpretant* triad of reference,[51] where the

---

[50] Eco (1990:35) argues:

> Derrida would be the first to say that his reading makes Peirce's text move forward, beyond the alleged intentions of its author. But if we are not entitled, from the Derridean point of view, to ask if Derrida read Peirce well, we are fully entitled to ask , from the point of view of Peirce, if he would have been satisfied with Derrida's interpretation.

[51] Peirce (1965 Vol. II:135 [2.228]) defines the sign as:

> …something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object*. Its stands for that object, not in all

interpretant is the process that enables one to infer the reference from a sign and its context. An essential feature of the interpretant is that it places the sign in the context *of other signs*. In his view a sign can only be a sign when it is interpreted, hence context forms an intrinsic part of the structure of signs.[52] The implication is that Derrida overlooks this aspect of "knowing more" in his appropriation of Peirce.

Eco is right in insisting that Peirce does not allow for every interpretation to be as good as any other. We have seen that Peirce holds that an erroneous interpretation will be revealed as such by the reality of the context imposing on it. If an interpretation is wrong, the particular context in which it occurs will not support that interpretation; it will lead to discrepancies that do not match the interpreter's knowledge of the world (habits, in Peircean terminology). Eco (1990:28-29) describes this point in "structuralistic terms" as holding that:

> …semiosis is potentially unlimited from the point of view of the system but is not unlimited from the point of view of the process. In the course of a semiosic process we want to know only what is relevant according to a given *universe of discourse*…one is entitled to try the most daring abductions, but if an abduction is not legitimated by further practical tests, the hypothesis cannot be entertained any longer.

In discussing "another form of drift" – deconstruction – Eco describes Derrida as proposing that a written text is a "machine that produces an indefinite deferral" (Eco 1990:32). In his subsequent summary of Derrida's argument, Eco highlights the absence of the intentional subject or author of a text, and the transcendental signified. Eco (33) takes Derrida to be implying that the reader no longer has the duty or the possibility of remaining faithful to the absent intention behind a text, which leads to the conclusion that "language is caught in a play of multiple signifying games" and that a text cannot have an unambiguous and generally agreed upon meaning.

Given this interpretation, Eco vehemently denies that Derrida's theory of language (which he calls "the infinite drift of deconstruction") can be likened to the unlimited semiosis of Peirce. He also denies that Derrida is justified in appropriating Peirce to support his theory that "there is nothing outside the text" (Eco 1990:35).[53] Eco argues that despite Peirce's

---

> respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen [sign].

[52] In Peirce's formulation (Peirce 1965 Vol. V:379 [5.569]):

> A sign is only a sign *in actu* by virtue of its receiving an interpretation, that is, by virtue of its determining another sign of the same object. This is as true of mental objects as it is of external signs. To say that a proposition is true is to say that every interpretation of it is true. Two propositions are equivalent when either might have been an interpretant of the other. This equivalence, like others, is by an act of abstraction (in the sense in which forming an abstract noun is abstraction) conceived as identity.

[53] Eco (1990:33) also criticises Rorty for encouraging such a view by labelling deconstruction and other forms of "textualism" pragmatism.

assumptions that led him to postulate unlimited semiosis, namely that i) we have no power of introspection and all knowledge about our internal state is gained through hypothetical reasoning; ii) we have no power of intuition, and every cognition is determined by previous cognitions; iii) we cannot think without signs; and iv) we cannot have any conception of the incognisable, his theory cannot be equated with "deconstruction" (36). He bases the reason for his contention on what he sees as Derrida's penchant for disregarding "obvious truths", in order to stress "non-obvious truths".[54] Eco proceeds to read Peirce in terms of a "zero-degree" reading.[55]

According to Eco (1990:37), Peirce seems to support the idea of infinite interpretation because of his principle of *synechism*, where "reality" consists of a continuum, rather than readily identifiable individual entities (cf. Chapter 1). In Eco's words "In a continuum where one can isolate infinite undetermined individuals, the possibility of error is always present, and therefore, semiosis is potentially unlimited" (37). Eco argues that Peirce holds that it is contextuality that allows for something to be identified and asserted. An assertion would take place within a "universe of discourse" and "under a given description, but it would not exhaust all possible determinations of that object" (37). Evidently Eco understands Peirce's argument to differ from that of Derrida, however, as we saw in the previous section Derrida proposes a similar argument with regard to context: context is a precondition for signification to take place. So far, Derrida's appropriation of Peirce seems to be justified.

Eco (1990:38) furthermore holds that other ideas in Peirce undermine Derrida's reading of him. He argues that we cannot ignore the "realistic overtones of Peirce's idealism".[56] One of those is the idea of "purpose" which Eco argues lies outside of language, where we interpret something in accordance with its purpose "in the external world" (38). Eco does not elaborate on this example, but seems satisfied that it points to an instance where something "beyond language" contributes to the meaning of a sign. Furthermore, Eco refers to two instances in which "semiosis is confronted with something external to it", namely indices and the production of the *Dynamic Object*. Although Eco does not fully agree with Peirce's definition of what constitutes an index, the very fact that the act of indication (when pointing to something and naming it) is possible, indicates that indices are "in some way linked to an item of the extralinguistic or extrasemiotic world" (Eco 1990:38). Dynamic

---

[54] Eco (1990:36) quotes Rorty in this regard who accuses Derrida of having "no interest in bringing 'his philosophy' into accord with common sense." In the light of Chomsky's view of the role of science, as discussed in the previous chapter, we could argue that this "lack of interest" in "common sense" on Derrida's part contributes to, rather than detracts from, the scientific relevance of his work. In many respects, philosophy as a scientific discipline should challenge our common-senseconceptions of the world.

[55] Eco (1990:36) defines a zero-degree reading as the one authorised by the "dullest and the simplest of existing dictionaries", and by the state of a given language at a given historical moment, and by the healthy native speakers of a given language community.

[56] The reader will recall that Peirce referred to himself as an "objective idealist". He allowed for the possibility that our access to reality is mediated by perception, interpretation and our cognitive abilities, he also held that, in principle at least, it would be possible for us to distinguish between human constructs and objective truths, given a communal process of enquiry and enough time.

Objects, on the other hand, are linked to Peirce's argument that we are compelled by Reality, or as Eco puts it "something external to the circle of semiosis" to produce representamens. Again Eco sees this example as indicating that there is something outside of semiosis that contributes to or inspires meaning. However, he clouds the issue by arguing (39):

> The Dynamic Object cannot be a piece of the furniture of the physical world but it can be a thought, an emotion, a feeling, a belief. We can say that a text can be interpreted independently of the intention of its utterer, but we cannot deny that any text is uttered by somebody according to his/her actual intention, and this original intention was motivated by a Dynamic Object (or was itself the Dynamic Object).

It is difficult to reconcile Eco's position that the Dynamic Object cannot be part of the "physical world" but is something like a thought or a belief external to the circle of semiosis, with his acknowledgment that Peirce holds that we think only in signs. Peirce does hold that we become conscious of the world prior to having knowledge of it (Firstness) and furthermore, that the world constitutes a reality that transcends our conceptions (Secondness), but we only gain knowledge of the world, or accord meaning to it in terms of synthesising experiences by means of signs (Thirdness). His point seems to be that the signs in terms of which we think are tempered and shaped by "reality" or a world external to the semiotic process. But in this regard Derrida cannot be faulted for his reading of Peirce. He argues (Derrida 1976:48):

> It must be recognised that the symbolic (in Peirce's sense: of "the arbitrariness of the sign") is rooted in the nonsymbolic, in an anterior and related order of signification: "Symbols grow. They come into being by development out of other signs, particularly from icons, or from mixed signs." But these roots must not compromise the structural originality of the field of symbols, the autonomy of a domain, a production, and a play: "So it is only out of new symbols that symbols can grow… But in both these cases, the genetic root-system refers from sign to sign. No ground of nonsignification – understood as the insignificance or an intuition of a present truth – stretches out to give it foundation under the play and the coming into being of signs.

In this instance, Derrida's interpretation of Peirce's argument is accurate. Firstness and Secondness are precognitive and do not constitute knowledge or meaning. Knowledge, thought and meaning necessarily involve Thirdness, and hence can only take place in terms of signs. In this regard Firstness and Secondness are very much akin to what Derrida envisions with the concept of *trace*, although an argument can be made that Peirce makes a much more comprehensive attempt at trying to account for the characteristics of Firstness and Secondness than Derrida does in terms of traces. As we shall discuss, this can be

attributed to the fact that Derrida very much stays within the realm of the phenomenological and does not make any serious attempt at accounting for the cognitive structure that underlie our ability to use language. In many ways this omission might be seen as a lack in his work.

Derrida does not view Peirce as a "deconstructionist" but declares that he "goes very far in the direction that I have called de-construction of the transcendental signified, which, at one time or another, would place a reassuring end to the reference from sign to sign" (Derrida 1976:49). Eco (1990:39) goes on to argue that Derrida and Peirce diverge when it comes to Peirce's conception of the "absolute object" or logical final interpretant which limits possible interpretations, in principle. Eco argues that such a limit to interpretation cannot take place within the "deconstructive framework" (he seems to use this term to refer to Derrida's work as a whole), and in this regard he is partly right. The final logical interpretant lies outside the "course of semiosis" (39). We have seen that Peirce believes that, given enough time and a large enough community of thinkers, the ideal state of complete information can, in principle, be reached. Peirce is not a nominalist and believes that it is possible to know the world "as it really is." He is also adamant that the process of sign-interpretation is not an arbitrary process. In this manner, Peirce attempts to preserve both the constructive aspect of knowledge (the interpretative element that "translates" the sign into some form of mind), and the possibility of true knowledge (ensured by the constraints placed upon possible interpretations by an external "reality" (cf. Pierce 1965 Vol. V:324-325 [5.473]).

Hence, Peirce conceives of "a community as an intersubjective guarantee of a nonintuitive, nonnaively realistic, but rather conjectural, notion of truth" (Eco 1990:39). Given Peirce's argument that the meaning of a proposition is nothing more than the conceivable practical effects if the proposition were true, Eco concludes that the process of interpretation must stop "outside language", if only for a while. He argues that every practical effect is not a "semiotic" one (40). Thus Eco argues that the "practical effects" of propositions must happen outside of language, even if the community must agree on particular practical effects through the mediation of language. The "transcendental idea of a community" transcends the intention of the "individual interpreter" (40), which Eco sees as not being Kantian, seeing that the transcendental principle at issue comes after and not before the semiotic process.[57] Such intersubjective meaning acquires privilege over individual meaning. Ultimately, there is an "ideal perfection of language" and the Real (40). Eco declares that "the process of language is not an individual affair" (41). In contrasting Peirce's semiosis to Derrida's theory of language, Eco implies that meaning *is* an individual affair for Derrida. Furthermore, he declares that with Peirce "the truth can at least be reached in the long run", implying of course, that it cannot be reached with Derrida.[58] As we hope to show in the next section, Eco

---

[57] In Eco's words (1990:40):

> …[I]t is not the structure of the human mind that produces the interpretation but the reality that the semiosis builds up.

[58] To quote Eco (1990:41):

errs in this interpretation of Derrida. Derrida's is not a position of relativism, nor does he deny that some interpretations of a given text (in the broad, Derridean sense) are better than others. Part of Eco's misreading of Derrida lies in his misconception of what Derrida means with the concepts "writing" and "text". Eco understands these terms as primarily applicable to "linguistic texts" in Derrida, and not as pertaining to our understanding of the world as a whole.

Eco's is a common error in interpretations of Derrida's work. Richard Rorty is another prominent commentator on Derrida's work, and is generally admiring of Derrida's work. However, Rorty (1996a:15) is dismissive of his "method called deconstruction", and declares that he sees "no real connection between what Derrida is up to and the activity which is called 'deconstruction'". Seeing that Rorty dismissively refers to deconstruction as "the sort of thing Derrida does"[59] and as a maxim that entails the instruction to "(f)ind something that can be made to look self-contradictory, claim that that contradiction is the central message of the text, and ring some changes on it (15)" it is not surprising that Rorty does not see the connection between Derrida the "deconstructionist" and Derrida the "critic of metaphysics". One reason for this misunderstanding of deconstruction is the use that Derridean theory has been put to in literary criticism. However, probably the most fundamental reason for this misunderstanding, and many others of Derrida's theory, is the failure to understand what Derrida means with his concept of "text", as well as the "texts" that deconstruction is aimed at.[60]

Rorty (1996a:14) refers to Derrida as "a critic of metaphysics, in the tradition of Nietzsche and Heidegger". There are many ways in which this appraisal is accurate. However, it may be misleading if one is not clear on Derrida's understanding of the concept

---

> There is community because there is no intuition in the Cartesian sense. The transcendental meaning is not there and cannot be grasped by an eidetic intuition: Derrida was correct in saying that the phenomenology of Peirce does not – like Husserl's – reveal a presence. But if the sign does not reveal the thing in itself, the process of semiosis produces in the long run a socially shared notion of the thing that the community is engaged to take as if it were in itself true. The transcendental meaning not at the origins of the process but must be postulated as a possible and transitory end of every process.

[59] Rorty (1996a: 13-17) points out the similarities between "the sort of thing Derrida does" and that of American pragmatism which is heir partly to the theories of Peirce (16):

> Pragmatism starts out from Darwinian naturalism – from a picture of human beings as chance products of evolution. This starting-point leads pragmatists to be suspicious of the great binary oppositions of western metaphysics as are Heidegger and Derrida. Darwinians share Nietzschean suspicions of Platonic other-worldliness, and the Nietzschean conviction that distinctions like mind-vs.-body and objective-vs.-subjective need to be reformulated in order to cleanse them of Platonic presuppositions and give them a firmly naturalistic sense. Naturalists, like Derrideans, have no use for what Derrida calls 'a full presence which is beyond play' and they mistrust, as much as he does, the various God-surrogates which have been proposed for the role of such a full presence. Both kinds of philosophers see everything as constituted by its relations to other things, and as having no intrinsic, ineluctable nature. What it is depends on what it is being related to (or, if you like, what it differs from).

[60] It is this misunderstanding that leads to Rorty's (1996a:16) rather damning conclusion that Derrida's assault on metaphysics produces "private satisfactions" to those who concern themselves with philosophy, but is politically inconsequential. It also leads him to conclude that Derrida's early work, such as *Of Grammatology*, demonstrates certain theses, and thus contributes to philosophy (17).

"metaphysics" and the role that it plays in our everyday understanding. Rorty, for example, argues that "pragmatists" read both Derrida and Wittgenstein not as having discovered the essential nature of language, or of anything else, but simply as having helped get rid of a misleading, and useless, picture – the one which Quine called "the myth of the museum": the image of there being an object, the meaning, and next to it its label, the word" (1996a:16). According to Rorty, "Pragmatists" are baffled by Derrida's suspicion of empiricism and naturalism, in that he views them as "forms of metaphysics, rather than replacements for metaphysics" (16). Rorty's formulation gets it exactly right, Derrida sees empiricism and naturalism, and all "-isms" for that matter, as forms of metaphysics. He is extremely critical of metaphysics, yet at the same time he denies the possibility of ever overcoming metaphysics and resigns himself to forever caught up in a metaphysics. How does one account for this seemingly paradoxical position? As Rorty (16) also puts it, pragmatists do not understand why Derrida wants to "sound transcendental", or "persists in taking the project of finding conditions of possibility seriously".[61] What should we "replace" metaphysics with, if not empiricism and naturalism?

It is important to note that Derrida does not understand "metaphysics" to be a specialised branch of philosophy, caught up in issues such as the fundamental nature of reality or of mind. Nor does he understand it to be a branch of enquiry that transcends the knowledge obtainable through the sciences (Norris 2004:14). Rorty's discomfort with Derrida's attitude towards empiricism and naturalism can be understood in the light of his conception of metaphysics as the result of "the metaphysicians need for getting behind Appearance to Reality" (Rorty 1996b:43). "Metaphysics" in Derrida's usage usually refers to the range of beliefs and presuppositions that underlie our everyday knowledge and language. Whereas Heidegger envisions the possibility of transcending the Western metaphysical tradition with its particular concepts and categories, and gaining access to an "authentic" experience of truth (Norris 2004:14), Derrida does not attempt to transcend this metaphysical tradition. He does not believe such a transcendence to be possible (e.g. Derrida 1976:14, 24; 1978:20; 1981a:5-6; 1997:116-118). We have already seen that Derrida does not see deconstruction as a technique that allows us to escape from our metaphysical heritage, but rather as a tool for making explicit the meaning/phenomena that the structure of signification necessarily suppresses (e.g. Derrida 1981a: 6-7). The difficulties and ambiguities posed by the structure of language are inevitable and, in a manner of speaking, inescapable. Contrary to the assumptions underlying some aspects of the linguistic turn in philosophy, his argument is not that many of the problems of philosophy can be put down to language puzzles that can be sorted out through the analysis of language.

Derrida does not make a methodological point as much as a structural point with his concept of deconstruction. His propensity to "sound transcendental" and to persist in "taking

[61] Critchley (1996:31) argues that setting up such an opposition between "transcendental" and "pragmatic" philosophy is "unhelpful". He argues that deconstruction is "pragmatist", but "not pragmatist all the way down" (37), because deconstruction is not relativist. For Derrida's view on the similarities between deconstruction and pragmatism see Derrida (1996:78-88).

the project of finding conditions of possibility" should be understood in the context of an epistemology, rather than ontology. Hence, Derrida is not concerned with the possibility of transcending empiricism in order to reach an originary, or authentic truth, nor should he be understood as trying to establish the conditions of possibility of universal truths. What Derrida is trying to establish with deconstruction, is a way to make explicit our metaphysical presuppositions, but also the impossibility of thinking outside of some form or system of metaphysical presuppositions (Derrida 1997:138). In this instance "metaphysical" can be understood as those first principles on which we construct our understanding of the world. Given our epistemological heritage and perhaps our cognitive abilities – although Derrida does not really elaborate on this aspect of our ability to access "reality" – we cannot but have certain preconceived notions that tint our understanding of the world. Derrida's argument is that we cannot escape having such "first principle's" – they are the very conditions of possibility for gaining knowledge – but that we need to be aware of them and of the possibility that they may be erroneous or misleading. The way in which he proposes that we take account of metaphysics by means of deconstruction (Derrida 1976:24):

> The movements of deconstruction do not destroy structures from the outside. They are not possible and effective, nor can they take accurate aim, except by inhabiting those structures. Inhabiting them in a certain way, because one always inhabits, and all the more when one does not suspect it. Operating necessarily from the inside, borrowing all the strategic and economic resources of subversion from the old structure, borrowing them structurally, that is to say, without being able to isolate their elements and atoms, the enterprise of deconstruction always in a certain way fall prey to its own work.

Deconstruction "falls prey to its own work" because it can never move "outside" of the structure of metaphysics, or achieve a neutral position and "untainted" access to reality independent of human language and cognition. A deconstruction always remains deconstructable. Derrida's position can be likened to Chomsky's on our "common sense" understanding of the world. However, Derrida is even more sceptical than Chomsky about the possibility of finding a corrective for such (possibly skewed) assumptions. Whereas Derrida is by no means dismissive of science, nor is he a relativist, he is careful to emphasise that science also operates in terms of an inherited "metaphysics".

7. Derrida on Truth and Science

Although extremely critical of logocentric metaphysics, Derrida does not advocate destroying logocentrism. As we have seen, he does not believe it possible to step outside the boundaries of our metaphysical heritage. If it were possible to do so we would risk destroying science, as well (Derrida 1976:75; 102). In Derrida's words: "we must begin where we are…in

a text that we already believe ourselves to be" (162). With his critique of an entire metaphysical tradition and the rejection of the idea of "presence" as a foundation in terms of which to guarantee truth, one would have to ask whether science and knowledge are possible in the Derridean universe. His critics' unequivocal answer would be "No". However, Derrida's own position seems more equivocal. He seems to view knowledge not as the systematic uncovering of hidden truths, but as a field that involves "*freeplay*" that is to say, which involves the possibility of infinite substitutions constrained by the laws or conventions that make the system of science possible in the first place (cf. Johnson 1993:151 ff).

It is important to emphasise that Derrida does not propagate or desire the "destruction" of science. His is not a world-view of relativism and undecidability as some of his critics might have it. Derrida is greatly concerned with the importance of science and truth and has much in common with both Chomsky and Peirce in that regard. Still, Eco is partly right in his assessment of Derrida's position on meaning and truth. Derrida does not attempt to "guarantee" ultimate truth in something external to the process of signification. He would argue that the belief that this is at all possible is an illusion based on the metaphysics of presence. However, that does not mean that he holds that relativism is inevitable or that any interpretation is as good as any other. Nor does he discredit or discount the importance of community in this regard. What Derrida objects to in the scientific enterprise (as with structuralism) is its totalising tendency – explaining an essentially open system with a model which is a closed system. This gesture of closure is the scientific gesture par excellence (Johnson 1993:151). However, Derrida does not oppose the closed-system epistemology of the western philosophical tradition with the idea of a purely open system. He concedes that some structure is needed to make a system possible in the first place (152 ff).

Critics of Derrida often accuse him of setting up an "all or nothing" choice between pure realisation, or self-presence and complete freeplay or undecidability. Derrida is then taken to argue for a complete freeplay or undecidability of meaning, and hence to exclude the possibility meaning of truth that can be objective (see Eco's criticism in the foregoing section). However, such criticism overlooks much of what Derrida's deconstruction of writing implies. It is his critics, rather than Derrida himself, who set up an "all or nothing" alternative between the extremes of complete realism and absolute relativity. Derrida (1997c:116-117) argues that any concept that lays claim to rigour implies an alternative of "all or nothing." One cannot form a philosophical (or scientific) concept outside of such an "all or nothing", or classical and binary logic. He states: "…it is true, when a concept is to be treated as a concept I believe that one has to accept the logic of the all or nothing" (117). His argument, however, is that the concepts that emerge from such a binary logic are always idealisations[62] – necessary and indispensable idealisations, but idealisations nonetheless.[63] Such idealisation forms part of

[62] For Derrida, "all conceptual production appeals to idealisation" (1997c:117).

[63] Derrida elsewhere (1997b:69) argues that a process of abstraction and idealisation is always a process of objectification; and without abstraction and idealisation, there is no systematisation (Derrida 1997c:118).

the rules and the context of the scientific discourse. And the determination and redetermination of a context is never a neutral, innocent, transparent, or disinterested gesture (Derrida 1997c:131). Furthermore, the analyses of the determination of a context is never a purely theoretical gesture, but always involves a *political* evaluation (132). Derrida argues that the "rules" that determine context are not found in nature, but "symbolic inventions, or conventions, institutions that, in their very normality as well as their normativity, entail something fictional" (134). Derrida does not argue that these rules are arbitrary, or unrelated to "reality", but he emphasises that they are not "natural realities" (134).

Derrida (1997b:70) argues, however, that, in terms of his concept of iterability, this idealizing abstraction is always provisional. Iterability (and the various terms that can supplant it, such as *différance*, trace, etc.) both renders idealisation possible and undermines complete idealisation.[64] "But this does not imply that all 'theorisation' is impossible. It merely de-limits the theorisation that would seek to incorporate its object totally but can accomplish this only to a limited degree" (71). The hierarchy of oppositional values on which western logocentric metaphysics is based is already present in a "positive", supposedly neutral concept that is being analysed. Furthermore, scientific discourse itself cannot escape being implicated in such a hierarchy, which is a normative hierarchy. Hence, scientific discourse should embody such "positive" values such as objectivity, neutrality, etc. Derrida argues that the determination of "positive" values (objectivity, normality, seriousness, etc.) is in itself dogmatic. It does not derive from "common sense", as is often argued, but from "a restrictive interpretation of common-sensewhich is implicit and never submitted to discussion" (Derrida 1997b:91):

> But often while analysing a certain ethicity inscribed in language – and this ethicity is a metaphysics (there is nothing pejorative in defining it as such) – they reproduce, under the guise of describing it in its ideal purity, the given ethical conditions of a given ethics. They exclude, ignore, relegate to the margins other conditions no less essential to ethics in general, whether of *this given* ethics or of another, or of a law that would not answer to Western concepts of ethics, right, or politics. Such conditions, which may be anethical with respect to any given ethics, are therefore not anti-ethical in general. They can even open or recall the opening of another ethics, another right, another "declaration of rights," transformation of constituents, etc. it is such conditions that interest me when I write of iterability and of all that is tied to this quasi conception a discourse and in other texts I cannot reproduce here (122).

---

[64] "There is no idealisation without (identificatory) iterability; but for the same reason, for reasons of (altering) iterability, there is no idealisation that keeps itself pure, safe from all contamination. The concept of iterability is this singular concept that renders possible the silhouette of ideality, and hence of the concept, and hence of all distinction, of all conceptual opposition. But it is also the concept that, at the same time, with the same stroke marks the limit of the idealisation and of conceptualisation" (Derrida 1997c:119).

Furthermore, Derrida (1997b:93) questions the simplicity of the oppositions which are made within metaphysics. Hence, he proposes that we take the limits of the scientific stance into consideration and rethink our received concepts of "science" and "objectivity" (Derrida 1997c:118).[65] Derrida proposes to add a "supplementary complication" to the classical, binary logic of scientific discourse, which takes this logic into account, but is not entirely dependent on it (117). He tries to formulate a *supplementary discourse* which tries to account for the fact that the "ideal purity" of concepts postulated in classical scientific theory is inaccessible, and shows that idealisation inevitably excludes essential traits of what it aims to describe. With iterability, Derrida aims to mark the "possibility and the limit of all idealisation and hence of all conceptualisation" (118). Hence, Derrida's argument is that the "all or nothing" choice is not created by him, but is implied in every distinction between, or opposition of concepts (120). It is a structural precondition for idealising, or simulating objectivity, in the first place.[66] The effect of Derridean analysis is to point out this structural necessity at the base of "the logic of opposition", which inevitably implies scientific discourse as a whole, and language in general. "All discourse involves this effect of idealism in a certain manner" (Derrida 1997b:94).

It is important to emphasise that Derrida does not argue that the idealisation in scientific discourse is illegitimate or superfluous. It is absolutely necessary for scientific discourse, and any discourse in general. Taking up an "objective" stance is essential to science. What he suggests is that the provisionality of this stance also be taken into account, in that one scientific discourse tries to take note of that which it inevitably excludes in idealisation. In other words, that the context in which scientific discourse takes place be taken into account (Derrida 1997:137):

> What is called "objectivity," scientific for instance (in which I firmly believe, in a given situation), imposes itself only within a context which is extremely vast, old, powerfully established, stabilised or rooted in a network of conventions (for instance those of language) and yet which still remains a context. And the emergence of the value of objectivity (and hence of so many others) also belongs to a context. We can call "context" the entire "real-history-of-the-world," if you like, in which this value of objectivity, even more broadly, that of truth (etc.) have taken on meaning and imposed themselves. That does not in the slightest discredit them. In the name of what, of which other "truth," moreover, would it? One of the definitions of what is called deconstruction

---

[65] Derrida declares (1997c:117):

> To this oppositional logic, which is necessarily, legitimately, a logic of "all or nothing" and without which the distinction and the limits of a concept would have no chance, I oppose nothing, least of all a logic of approximation, a simple empiricism of difference in degree; rather I add a supplementary complication that calls for other concepts, for other thoughts beyond the concept and another form of "general theory," or rather another discourse, another "logic" that accounts for the impossibility of concluding such a "general theory".

[66] "…the structural idealism of which we have just spoken constitutes the condition of a certain classical value of what is called scientific truth" (Derrida 1997c:120).

> would be the effort to take this limitless context into account, to pay the sharpest and broadest attention to possible context, and thus to an incessant movement of recontextualisation. The phrase which for some has become a sort of slogan, in general so badly understood, of deconstruction ("there is nothing outside the text") means nothings else: there is nothing outside the context. That does not mean that all referents are suspended, denied, or enclosed in a book … But it does mean that every referent, all reality has the structure of a differential trace, and that one cannot refer to this as "real" except in an interpretive experience. The latter neither yields meaning nor assumes it except in a movement of differential referring. That's all.

The stability of both meaning and knowledge derives from the fact that the system as a whole transcends any individual choices of its users. And the users of the system have their understanding constituted by that same system, which they inherited, hence they cannot break out of the system (Cilliers 1998:39).

Instead of ignoring the "marginal" or "parasitic" elements that do not seem to "fit into" a given structure of discourse, Derrida argues that a theory should try to recognise why a structure which is considered to be normal or ideal inevitably renders such "accidents". In order to do this, "the habitual logic space" needs to be transformed, not in order to establish confusion or indeterminability, but in order to highlight simplifying idealisations that presents the object/s being studied in terms of a false simplicity (Derrida 1997c:127).[67]

Like Chomsky, Derrida has reservations about the capability of science to address complex and difficult problems. However, he sees the limitations of science as inherent to the structure of all human knowledge and meaning. In these terms, he is able to point to a way in which science can overcome its structural limitations to a certain extent, through analysing its structure and the complexities that it simplifies in order to operate at all. Like Peirce, Derrida recognises the abduction implicit in all empiricism. Unlike Peirce, however, Derrida does not attempt to guarantee objective truth beyond empirical structures with recourse to evolutionary laws that apply to knowledge or to God; Derrida "guarantees" truths as far as possible in

---

[67] Derrida (1997c:127) sums up his argument as follows:

> As a philosophy, empiricism is still dominated by a logic I deem it necessary to deconstruct. Doubtless the concept of iterability is not a concept like the others (nor is *différance*, nor trace, nor supplement, nor parergon, etc.). that it might belong without belonging to the class of concepts of which it must render an accounting, to the theoretical space that it organises in a (as I often say) "quasi"-transcendental manner, is doubtless a proposition that can appear paradoxical, even contradictory in the eyes of common-senseor of a rigid classical logic. It is perhaps unthinkable in the logic of such good sense. It supposes that something happens *by* or *to* set theory: that a term might belong without belonging to a set. It is of this too that we are speaking when we say "margin" or "parasite." It is of this as well that an accounting, and a reason, as the enlightened modern thinkers we still want to be, right? But this thought of iterability, if it troubles all exclusion or simple opposition, should not capitulate to confusion, to vague approximations, to indistinction: it leads to an extreme complication, multiplication, explication of "precise and rigorous distinctions."

terms of the structure constitutes knowledge, but emphasises that it is not possible to ground truth and knowledge in terms of any entity perceived to be outside of that structure.

8. WRITING AS COGNITIVE APPARATUS?

A criticism that can be levelled against Derrida is his failure to attempt to account for the role that human cognition plays with regard to the structure of language and meaning in general. Thus, although many of his arguments reflect the insights that caused Peirce to propose his theory of abduction, Derrida seems to confine his analysis to the (extended) structure of language and phenomenological speculation regarding our experience of the world. The closest that Derrida comes to accounting for the human cognitive structure is in his discussion in *Freud and the Scene of Writing* (1978:196-231) of the use that Freud[68] makes of metaphors of writing in order to describe various cognitive functions (although he accuses Freud of remaining caught up in the metaphysics of presence). Notable among these works are Freud's early *Project for a Scientific Psychology* (1950 [1895]); *Interpretation of Dreams* (1901); and *A Note upon the "Mystic Writing Pad"* (1925). Derrida's analysis of Freud's work is involved and fascinating, and he manages to highlight many of his own preoccupations and concepts (such as those of trace, and différance) in Freud's work.[69] He believes, however, that Freud's work belongs firmly to the history of the metaphysics of presence and its "system of logocentric" repression, and that Freud remains caught up within traditional and metaphysical concepts, without reflecting upon them (Derrida 1978:197-198).

Although he is critical of much of Freud's presuppositions, Derrida is clearly taken with his conception of the cognitive in terms of writing, traces, différance, etc. Derrida argues that there exists an underlying *necessity* that induced Freud to compare the cognitive apparatus to structure of writing. Apart from his critique of Freud's attempts at accounting for

---

[68] Rorty (1996a:16) refers to Freud and Heidegger as Derrida's two "grandfathers" which is a very apt description of Derrida's preoccupation with, and position in, the philosophical tradition. Derrida sees Freud's postulation of the unconscious as a defining moment in our metaphysical epoch. Whereas the tradition had previously conceived of the individual subject as fully "present" to himself, and as fully conscious of himself, his motivations, and his "internal dialogue" (a form of phonocentrism), the inaccessible, yet vitally influential unconscious, undermines this view of the "present" subject. Freud, like Derrida challenges the traditional practise of equating meaning with consciousness (Sharpe 2004:64).

[69] "A certain alterity, to which Freud gives the metaphysical name of the unconscious – is definitely exempt from every process of presentation by means of which we would call upon it to show itself in person. In this context, and beneath this guise, the unconscious is not, as we know, a hidden, virtual, or potential self-presence. It differs from, and defers, itself; which doubtless means that it is woven of differences, and also that it sends out delegates, representatives, proxies; but without any chance that the giver of proxies might "exist," might be present, be 'itself" somewhere, and with even less chance that it might become conscious. In this sense, contrary to the terms of an old debate full of the metaphysical investments that it has always assumed, the "unconscious" is no more a "thing" than it is any other thing, is no more a thing than it is a virtual or masked consciousness. This radical alterity as concerns every possible mode of presence is marked by the irreducibility of the after-effect, the delay. In order to describe traces, in order to read the traces of "unconscious" traces (there are no "conscious" traces), the language of presence and absence, the metaphysical discourse of phenomenology, is inadequate. (Although the phenomenologist is not the only one to speak this language.)" (Derrida 1986: 20-21).

cognitive structure, however, Derrida himself does not have much to say upon the subject. He does not generally take the role that human cognitive structure and cognitive abilities could play in terms of the structure of language into account, beyond highlighting the fact that "presence" to meaning and intention is an illusion. In this sense his theory is still firmly caught up within the preoccupations and jargon of phenomenology and structuralism. For all his criticism of the metaphysical tradition and structuralism in particular Derrida does not manage to overcome the unjustified, non-naturalist assumptions that Chomsky argued are part of our theoretical heritage.

Although it can be argued that his aim is precisely to address certain presuppositions within these theoretical traditions and that his project does not include biology, neurology or human anatomy, his arguments could benefit from making his epistemological position more explicit. Such clarification might go a long way towards dispelling interpretations of his work that present him as suggesting that, ontologically, "everything is text". Of course, in terms of his own theory, the epistemological/ontological divide is an idealisation, but it remains a useful clarificatory one at that. The question arises as to whether this conception of the cognitive structure is merely a convenient (philosophical?) metaphor, or whether it has some basis in fact.

9. CONCLUSION

It can thus be argued that despite Chomsky's scepticism and his own failure at accounting for meaning in language in terms of the structure of language, Derrida's analysis shows that it can be done. What was required over and above the structuralist assumption that meaning emerges as a result of the differences within the language system, was the temporalisation of the language structure. Furthermore, it was necessary to incorporate an aspect of language that Chomsky explicitly disregards – those aspects of language that do not take place within the cognitive structure, namely "E-language". Hence, Derrida takes the context in which language is produced and decoded, as well as historical and future aspects of the system, into account when explaining meaning. In this manner he is able to avoid the trap of assuming that the structure of language must conform to a preconceived logical structure or that it needs to be "perfect". Language only needs to be optimal in as far as it facilitates communication between particular parties, given their shared conventions and interpretive capabilities. Interpretation/communication is always a bit of a hit-and-miss affair, subject to some guesswork and continual revision. This is not a defect in our language systems, but an integral part of such systems, given the context in which they need to function. In chapter 4 it will be seen that an adherence to the assumption that language must necessarily be optimal in terms of a specific logic would prove to be a tenacious obstacle to creating "thinking machines" with abilities comparable to that of humans.

In terms of his temporalisation of the language system, Derrida recognises that memory plays a crucial role in establishing meaning within the system. Meaning is thus not

only distributed throughout a differential system which spans both conventional language – Chomsky's I-language – and the context where is manifests – E-language. The meaning of a specific sign is also determined by a network (traces) of past encounters with the term, as well as other possible future applications of it. Because every context is unique, every given instance of a sign has a unique and fleeting meaning. Such meaning is taken up within the memory of the system, be it in the minds of language users or in permanent records, such as recordings, writings, and so forth. All of these past meanings create traces that potentially influence present and future meanings. Derrida thus incorporates Peirce's insight that reference requires three phenomena, a sign, its object, and an interpretant, and adds a fourth element to the structure of reference. The following two chapters will highlight the importance of adding memory, time, and context to the process of establishing meaning. Derrida makes it clear that the sign-object dichotomy is a woefully inadequate conception of the process of reference, which explains why such conceptions always lead to assumptions of essentialism or nominalism.

Derrida's phenomenological preoccupation leads him to conclude that the "historico-metaphysical epoch *must* finally determine as language the totality of its problematic horizon" (1976:6). As we have indicated, Derrida, like Peirce, highlights the fact that our access to the world is always mediated through the signs in terms of which our knowledge is structured. His argument is that the western philosophical tradition failed to acknowledge this state of affairs and incorporate in into its fundamental world-view. It is not the world as such that has a semiotic character (as is Peirce's eventual argument) but the phenomenological world. And Derrida envisages our phenomenological make-up as being structured according to the logic of writing, as he analyses it. However, it can be argued that Derrida remains firmly caught up within the metaphysics of his own tradition in this regard, in that he does not do much by way of establishing his phenomenological assumptions outside of the conventions of the tradition. One can agree with Derrida that one's metaphysical heritage can never be completely left behind, without agreeing that deconstructing metaphysical assumptions "from within" is the only way in which such assumptions can be challenged. Aside from quite restrictive metaphysical/philosophical commitments, there seems to be no reason why one should not look beyond the particular boundaries of the discipline in a given era for alternative ways in which to challenge an inherited metaphysical world-view.

In what manner can our cognitive structures be said to embody the structure of writing in the Derridean sense? Perhaps Derrida assumed that the cognitive and information sciences that were coming into their own by the mid twentieth century would provide the mechanical underpinnings for his abstract theories. In order to comprehensively address the assumptions of the logocentric tradition, a naturalistic/mechanistic conception of the mind is needed. Such a conception would serve to undermine the pervasive idea that the mind somehow transcends the physical world

and hence eludes scientific scrutiny. In the following chapter we will examine such attempts at mechanising the mind.

Derrida's most pertinent insights with regard to our topic are his adaptation and extension of semiology to structuralism and philosophy in general, and his incorporation of temporisation, spacing, and context into language structure. As was argued in the introduction to this chapter, his work was indicative of broader trend within the sciences in general – a trend which he argued was inevitable. Chapter 5 will explore the application of a Derridean take on Peirce's semiology in order to account for *Homo sapiens's* apparently unique cognitive abilities. This attempt by Terrence Deacon moves well beyond structuralist and phenomenological conventions and is a radical attempt at integrating philosophical speculation with accepted scientific fact in order to address puzzles still presented by questions regarding mind and language. First, however, we will examine the semiological turn as it manifested itself in the mid- to late twentieth century, and which was the result of the insight that thought can be modelled and hence studied. As will be argued, it soon became apparent that the new information sciences, systems sciences, cybernetics, and life sciences needed to incorporate temporisation, spacing, and context into their models of complexly structured systems in order to create models that even approximate the abilities of their real-life counterparts.

CHAPTER 4

THE CENTURY OF THE MECHANISED MIND

Minds are not bits of clockwork, they are just bits of not-clockwork.
Ryle (2002[1949]:36).

Darwin insisted that his theory explained not just the complexity of an animal's body but the complexity of its mind. "Psychology will be based on a new foundation," he famously predicted at the end of The Origin of Species. But Darwin's prophecy has not yet been fulfilled. More than a century after he wrote those words, the study of the mind is still mostly Darwin-free, often defiantly so…The allergy to evolution in the social and cognitive sciences has been, I think, a barrier to understanding
(Pinker 1997:23).
…the design of our…minds gives rise to paradoxes, brain-teasers, myopias, illusions, irrationalities, and self-defeating strategies that prevent, rather than guarantee, the meeting of our everyday needs
(Pinker 1997:207).

1. INTRODUCTION

In the previous chapter the focus fell on Derrida's amendments to structuralism, through taking temporisation, spacing, and context into account when modelling structural phenomena such as language. The result was that he could account for meaning in terms of the structure of language – something that Chomsky suspected might not be possible because of the complexity of the problem. Derrida argued that his theory was part of a greater theoretical trend that had become an inevitability within the western intellectual tradition. Derrida's deterministic take on theoretical developments are an exaggeration, however, a conceptual shift does seem to have taken place by the mid- to late-twentieth century that his conception of "writing" exemplifies in various ways. This shift is particularly evident in the information, life, and cognitive sciences. Many of these disciplines where directly or indirectly concerned with modelling human thought and cognition, and this chapter will track some of the developments in this regard.

When it comes to studying structure in terms of temporality, spacing, and context a seemingly pertinent theoretical approach that comes to mind is that of evolution by means of natural selection as formulated by Charles Darwin in the mid-nineteenth century. Natural selection seems an especially pertinent theory when trying to explain two phenomena that seem fundamentally related to human biology (or psychology at least), namely language and

the mind. However, as will become apparent in the course of this chapter, there existed, and still exists, a deep-rooted resistance to evolutionary theory in various academic disciplines, as was exemplified by Chomsky's scepticism toward the theory (see Chapter 2). Whereas evolution by means of natural selection has met with great success in the biological sciences, it has been curiously absent in the social and cognitive sciences. As Pinker (1997:23) puts it: "…the study of the mind is still mostly Darwin-free, often defiantly so… The allergy to evolution in the social and cognitive sciences has been, I think, a barrier to understanding." This chapter will give a brief overview of Darwin's theory and of the major developments within the cognitive sciences. As will become apparent, it is only by the late-twentieth century that the evolutionary perspective becomes established as a legitimate academic position. The "evolutionary turn" in theories of mind will be discussed in terms of Pinker and Bloom's seminal 1990 paper "Natural Language and Natural Selection". Up until this point many firmly entrenched philosophical assumptions caused evolutionary theory to be regarded as a slightly disreputable secondary element, inferior to pure theory, to paraphrase Simon Blackburn (1995:456-459). It will also become apparent that Darwin's theory sparked a general conceptual revolution stretching far beyond biology proper, the extent of which has yet to be fully appreciated.

2. Darwin's Legacy (or the Absence Thereof)

Whereas prior to the Enlightenment the origins of the world and its inhabitants were generally thought of in terms of the Christian story of Creation and Greek essentialist thought (in the west, at least), the Enlightenment raised the possibility that these origins are non-miraculous and that they can be explained in terms of scientific methodology. As a result the theory of Evolution, in the sense of the idea that the world developed from primitive beginnings through natural processes, can be considered to be a product of the Enlightenment (Ruse 1995:254). It was in the middle of the nineteenth century that the theory of evolution was to enter into the realm of respectable science, with the comprehensive evolutionary theory put forward by Charles Darwin in 1859. In the twentieth century his theory was amended and expanded, when the science of genetics and evolutionary theory became central to biological thinking.

What is striking from our point of view is that, despite the tremendous impact that evolutionary theory has had on other aspects of biology in the twentieth century, it has been virtually absent when it comes to theories of mind during the same period. As we shall discover during the course of this chapter, it is only by the 1990's that an evolutionary approach to the mind has gained some acceptability within the scientific mainstream. Given that Darwinian evolutionary theory placed human beings on a par with other organisms in terms of their evolutionary origins, it seems curious that a central aspect of the human animal, its mind, has for so long remained impervious to evolutionary considerations.

It is sometimes difficult for a contemporary scholar to appreciate the revolutionary character of Darwin's theory of evolution through natural selection, and the impact that it had (and continues to have) on almost all aspects of scientific thinking. With the publication of *Origin of the Species* in 1859, the theory of evolution was to turn an entire world-view, or paradigm, on its head. Kuhn (1970:111) likens the impact of a change in scientific paradigm to a scenario where "the professional community had been suddenly transported to another planet where familiar objects are seen in a different light and are joined by unfamiliar ones as well." Changes in paradigm cause theorists to see the world of their "research-engagement" differently. Thus Kuhn's position is that scientists respond to a different world when they adopt a change in paradigm. The post-*Origin* world was indeed radically different from the one that preceded it, and the ramifications of the subsequent changes in world-view are still being felt today.

It should be emphasised that the theory of evolution was not a bolt from the blue. Kuhn (1970:7) argues that a revolutionary theory is "seldom completed by a single man and never overnight", and this was certainly the case with evolution. It is well-known that Darwin was spurred into publishing his theories on the evolution of species when it became known that Alfred Wallace was about to publish a similar theory, which he arrived at independently.[1] In *The Origin* Darwin is at pains to point out that many of his findings are based on the research and speculation of many other theorists from various disciplines, scattered across the globe. It is clear that evolution was an idea whose time had come.

Early in the nineteenth century most people would have believed that the earth was roughly six thousand years old and would have taken biblical narratives relatively literally. This world-view created a congenial environment for a school of religious apologetics, based on Natural Theology (Burrow 1968:20-21). In terms of biology, Natural Theology in the nineteenth century argued that the complexity and the intricacy of the various species were evidence of the existence of God as the creator of these species, since it was inconceivable that such structures could come into being as the result of the blind laws of nature. In this world-view, natural phenomena were made in a manner which was either pleasing to human sensibilities, or which met human beings' needs. Displeasing phenomena could be attributed to God's displeasure with his creation. Hence the natural world pointed to the existence and the goodness of God (Burrow 1968:21). Any suggestion that the world was not created by a divine creator would have been met with moral outrage. Furthermore, species were thought to be permanent, created for a specific purpose and fulfilling a specific role (Dennett 1995:36-39). Evolutionary theory was not entirely unknown at this point. According to Burrows (1986:27) in later editions of *The Origin* Darwin lists over thirty predecessors who put forward

[1] According to Pinker (2002:28) Wallace differed from Darwin when it came to the implications of natural selection for humans. Wallace held that the human mind could not have been created by natural selection and had to have been designed by a superior intelligence. Part of his reason for holding this belief was his insistence that the human mind could not be subject to the control of the deterministic world presumably implied by natural selection. His attitude was in keeping with those of his time and many of Darwin's contemporaries would have agreed with Wallace.

some form of evolutionary theory. Many scholars were sceptical, however. One of the main reasons that these theories were not taken seriously was the short period of geological time thought to have been available for such evolution to take place in. Furthermore, it was difficult to conceive of a process by means of which evolution could take place. It seemed impossible that the known natural laws could explain why and how "transmutations" are transmitted to offspring?[2]

However, by the nineteenth century geological evidence was pointing to longer and longer periods in which the evolutionary process could have taken place. Furthermore, fossil records showed that entire species had become extinct, which ran counter to the prevalent assumption that species were permanent and fulfilled specific roles.[3] Also, Darwin's grandfather, Erasmus Darwin, suggested that vestigial organs – which seemed to have served a purpose at some stage in the history of organisms, but had none now – indicated that modern species were different from their ancestors (Burrows 1968:28). Vestigial organs did not make sense in a world designed and executed by a divine creator. Erasmus Darwin also made the breakthrough suggestion that the evolutionary process was promoted by the struggle for existence among species and the competition for mates (28).

A further breakthrough allowed Charles Darwin to formulate a comprehensive and reasonably convincing version of evolutionary theory. He hit upon the idea that *randomly* occurring and favourable variations in offspring were exploited by the evolutionary process because of the advantages that such variations accorded the species in its struggle for existence with other species. He named this process Natural Selection (Darwin 1985:115). Darwin was inspired by Malthus's thesis that population growth continues until it outruns food supply, or until it is checked by war, disease, etc. (Darwin 1985:117; Dennett 1995:40-42). This principle serves as a foundation for the observation that species struggle for existence, rather than coexisting in specific niches (although Malthus meant to argue for equilibrium between the species (33)). Darwin formulates his principle as follows (Darwin 1985:115):

> All these results…follow inevitably from the struggle for life. Owing to this struggle for life, any variation, however slight and from whatever cause proceeding, if it be in any degree profitable to an individual of any species, in its infinitely complex relations to other organic beings and to external nature, will tend to the preservation of that individual, and will generally be inherited by its offspring. The offspring, also, will thus have a better chance of surviving, for,

---

[2] It should be pointed out that Darwin was also at a loss as to how to account for variation in offspring, which would then be transmitted to future generations (Darwin 1985:76).

[3] Burrows (1968:31) places great emphasis on the impact that this new conception of geological time had on thought in general:

> There are few greater revelations, if any, in the history of thought, than the change within a few decades from a prevailing view of geological time that could be grasped imaginatively to one which, like astronomical distances, could be grasped only mathematically.

> of the many individuals of any species which are periodically born, but a small number can survive. I have called this principle, by which each slight variation, if useful, is preserved, by the term of Natural Selection, in order to mark its relation to man's power of selection.

It is important to emphasise that Darwin was not concerned with progress or perfection in an absolute sense. "Fit" in his sense is always in relation to a particular environmental niche. Furthermore, natural selection is not a law for necessary development; it only takes place when a variation in the species is beneficial to it, within a given environment with given challenges (Darwin 1985:348). The same modification within another context may be inconsequential and hence will not be selected for. Darwin's conception of the "struggle for existence" is a nuanced view and the term should not be understood merely in terms of ruthless competition and rivalry between species. Furthermore, the emphasis of the struggle for existence not only entails the survival of the individual, but more importantly, is aimed at leaving (successful) progeny (Darwin 1985:116)[4]:

> I should premise that I use the term Struggle for Existence in a large and metaphorical sense, including dependence of one being on another, and including (which is more important) not only the life of the individual, but success in leaving progeny. Two canine animals in a time of dearth may be truly said to struggle with each other which shall get food and live. But a plant on the edge of a desert is said to struggle for life against the drought, though more properly it should be said to be dependent on the moisture.

The distinction between struggling for an individual life as opposed to struggling on behalf of one's progeny would lead to many of the misunderstandings and distortions that would plague the theory of Natural Selection in the years to come. One of Darwin's seminal insights, then, was the realisation that the "economy of nature," thus "every fact on distribution, rarity, abundance, extinction, and variation" can be accounted for by means of the struggle for existence and the way in which this struggle can be aided by utilising favourable, though random variations in organisms (Darwin 1985:116). If there were no struggle for existence, there would be no check on the reproduction of organisms which would very quickly result in there existing many more organisms than can be supported by the finite resources available in the world.

Darwin conceived of the interaction between individuals, species, the environment, and climate as an extremely complex web of interactions, which he demonstrates with a series of examples (e.g. 1985:117-129):

---

[4] Richard Dawkins (1976) emphasises this aspect of Darwin's theory, and refines it with his proposal that organisms are survival machines, constructed by replicators (genes) in order to increase the likelihood that they will be copied (21).

> A corollary of the highest importance may be deduced from the foregoing remarks, namely, that the structure of every organic being is related, in the most essential yet often hidden manner, to that of all other organic beings, with which it comes into competition for food or residence, or from which it has to escape, or on which it preys (Darwin 1985:127).

We have seen that the general conception of the age of the earth had changed by Darwin's generation, and that the new geological timeframe allowed him to propose that despite the slowness of the process of Natural Selection, the earth was indeed old enough to have allowed for the evolution of species by means of natural selection (153). Darwin also argued that the greatest amount of life can be supported by the greatest diversification of structure in species (157). The more intense the contest between species for resources in a given area, the more diverse the species dependent on that area will become. In this manner, specialisation will ensure that a particular species has access to a niche in that environment that others will not be able to exploit, thus ensuring its survival until circumstances change. Increased diversification leads to the development of new, distinct species (155ff). Darwin speculates that the individuals with the most (beneficial) divergent variations within a species would have the greatest chance of survival and reproduction because of their ability to exploit novel niches. Hence, such divergent variations will have a greater chance of being preserved by Natural Selection. Darwin argues that this is the reason that intermediate, or "missing link", species are not readily found (206ff).

Darwin was convinced that natural selection worked through heritable variation in offspring and not (primarily) in terms of the use or disuse of organs as proposed by Lamarck (1985:100; 173-204).[5] However, Darwin did not have a conception of genetics, of course, and was uncertain as to how such a process of heritable variation would work.[6] Nevertheless, he drew upon what he knew about heritable variation – partly from his experience as an avid breeder of pigeons – and sketched a scenario of variation through slight heritable mutations that would be born out by the subsequent science of genetics. He made a convincing case, for example, for how an organ as complex as the eye – an organ of "extreme perfection and complication" – could have evolved from slight, random variations across an adequate time span (1985:217-224). Darwin also indicated how characteristics such as instinct can be brought about by his principle (1985: 236-263).

---

[5] Darwin does allow for the possibility that use and disuse of organs and the effect of climate could have some effect on variation in species (Darwin 1985:100).

[6] For instance (Darwin 1985:76):

> The laws of inheritance are quite unknown; no one can say why the same peculiarity in different individuals of the same species, and in individuals of different species, is sometimes inherited and sometimes not so; why the child often reverts in certain characters to its grandfather or grandmother or other much more remote ancestor; why a peculiarity is often transmitted from one sex to both sexes, or to one sex alone, more commonly but not exclusively to the like sex.

According to Burrow (1985:40) *The Origin* had some notoriety upon its publication, mainly due to two things: Firstly, it "destroyed at one blow" the most influential tradition of Protestant apologetics at that time – Natural Theology. The intricacy in the natural world which Rational Christianity attributed to a Divine Designer could now be explained in terms of Darwin's natural law of natural selection, where the struggle for life resulted in the preservation of random variations in offspring. Of course, although he never explicitly made the claim, Darwin's argument implied that man was "first cousin to the apes"; a claim that was met with derision and outrage (41). However, (in accordance with Kuhn's description of a paradigm shift) Darwinism became scientifically respectable as younger theorists adopted the theory (42). The scientific acceptance of Darwinism resulted in a widespread "intellectual crisis" (43). Not only did it transpire that the world was not designed according to the purposes of a superintending Being, it now seemed that the world was not designed according to any purposes at all, but was the result of "blind chance and struggle, and man [was] a lonely, intelligent mutation, scrambling among the brutes for his sustenance" (43). Whereas the Greeks, the Enlightenment and the rationalist Christian tradition saw answers to human dilemmas in nature, a purposeless Darwinian world could presumably not offer such answers.

Burrow (1985:43) sees the naturalistic ethical tradition in Western philosophy as the root of subsequent tendencies to read moral prescriptions into nature as conceived of in terms of Darwinian theory. This led to the practice in the latter half of the nineteenth century and in the early twentieth century to justify states of affairs in terms of "the survival of the fittest" which guaranteed "progress" as supposedly demonstrated by the law of natural selection (Pinker 2002:14-16). This, despite the fact that Darwin explicitly described his theory as one of adaptation and not of progress (e.g. Darwin 1985:348). Furthermore, he did not endorse the application of his theory to social contexts (Burrow 1985:45). Paradoxically, such "Social Darwinist" theories, even though not Darwinian, adopted "Darwinian" and "Evolutionary" as "loose and honorific terms" with which to lend legitimacy to their theories (45).

Surprisingly, despite its popularity in the sphere of the social sciences, the theory of natural selection came to be challenged in biological circles, mainly because it could not account for the mechanism that allowed for variations to be inherited by the offspring of species. It was only in the nineteen thirties, with the incorporation of Gregor Mendel's theory of genetics (and later of course that of Watson and Crick) into the theory of natural selection that Darwin's theory would be reinstated within mainstream biological thinking (Burrow 1985:46-48; Dennett 1995:20ff; 58-60).[7] It subsequently became firmly established as a new paradigm, which would influence thinking far beyond the scope of the biological sciences.

---

[7] When Gregor Mendel's 1866 work on inheritance was rediscovered in 1900, it first supplanted the theory of natural selection. It is only in the 1930's with the work of theorists such as R. A. Fisher, J. B. S. Haldane, and Sewall Wright, which indicated that genetic theory and natural selection were fully compatible that evolutionary research was rekindled as a legitimate scientific concern (Patterson 1998:234). During the late 1930's and early 1940's genetic theory was integrated with more traditional

In recent years the perception has been created (particularly in the popular press) that "Darwinism"[8] is a theory under siege. Theorists like Gould (e.g. 1979; 1980; 1982; 1989; 1991; 1992a; 1992b; 2000), Kaufmann (1995), an even Chomsky (e.g. 1972; 1975; 1994; 2000a; 2000b; 2004) conjure up visions of an incomplete and outdated theory, fundamentally incapable of addressing certain puzzles within the life-sciences, such as the development of language or complex structure. However, the much-touted demise of neo-Darwinian theory has not come to pass; nor has significant revisions of its basic principles as set out by Darwin. As Dennett (e.g. 1995) and Dawkins (e.g. 1999) illustrate, much of the hype is generated by the popular press, presumable due to the controversy that evolutionary theory elicits in some religious circles. Dawkins sums up much current biological opinion best (with regard to the characterisation of contemporary Darwinism as "neo-Darwinism"): "I think the need for the 'neo' is fading, and Darwin's own approach to 'the economy of nature' now looks very modern" (1999:298).

One would have thought that if an entire world-view regarding human beings, their genealogy, and their place in the universe has come under pressure from evolutionary theory, language and mind, two of the quintessential characteristics attributed to human beings under that world-view, cannot but be affected as well.[9] However, nearly 150 years after the appearance of *The Origin*, there remain surprisingly many aspects of theories of mind and language that do not seem to take evolutionary theory into account at all. We know very little about the origin and evolution of language, even though many theorists list language as the one "peculiarly human" trait (Christiansen and Kirby 2003b:2). The same applies to the mind, which has only recently explicitly been subjected to evolutionary considerations.

Christiansen and Kirby (2003a) point to a possible reason for this state of affairs (with regard to language, at least). According to them (2003a:300) interest in the origin and evolution of language was already extensive by the time that Darwin published *The Origin of the Species*. With few credible theories, minimal consensus, and hence little theoretical constraint, conjecture and speculation burgeoned, leading to "outlandish speculations," which led the influential *Société de Linguistique de Paris*, in 1866, to ban the topic from scientific discussion for more than a century (Pinker and Bloom 1990:31; Christiansen and Kirby 2003a:300). This "ban" would effectively last until the last decade of the twentieth century, when advances in various fields, especially in the brain and cognitive sciences, would provide

---

fields of evolutionary interest such as anatomy, palaeontology, and systematics. In 1953 interest in evolution shifted focus to the molecular level with Francis Crick and James Watson's model of the structure of DNA, which indicated the material basis of heredity (234).

[8] Generally, "Darwinism" is taken to include "neo-Darwinism", which is the synthesis of Darwin's theory of evolution by means of natural selection and contemporary gene theory.

[9] Evolution by means of natural selection does not go completely unchallenged in contemporary orthodoxy and there exists an extensive literature on objections to aspects of contemporary evolutionary theory (e.g. Gould and Lewontin 1979; Gould and Vrba 1981, among many others) and counter-endorsements of the theory (cf. Dennett 1995). Dennett (1995:95) argues that the implications of "the Darwinian Revolution" are still not fully appreciated, however, he insists (and goes on to argue convincingly) that the idea that the central tenets of Darwinian theory will be refuted "is about as reasonable as the hope that we will return to a geocentric vision and discard Copernicus."

theoretical constraints in terms of which the evolution of language could be analysed in a more restrained manner. Currently, the field can again be said to be burgeoning, and is established as a field of legitimate scientific enquiry, rather than a source of outlandish speculation.

A further inhibiting factor in studying the origin and evolution of language is its inherently interdisciplinary character. Despite its scientific delimitation, language evolution is still a challenging and extremely complex problem. Given its complex subject matter, the study of the evolution of language is a broadly interdisciplinary field, incorporating to a greater or a lesser extent fields such as linguistics, the cognitive sciences, neurology, evolutionary biology, archaeology, psychology, and philosophy among others (Christiansen and Kirby 2003b:2). Taking this into account, it becomes obvious that it is next to impossible to specialise in all fields relevant to the object of study, and it seems an equally impossible task to master all the relevant literature on the subject, despite the fact that language evolution is in many respects a relatively young scientific discipline. Furthermore, seeing that language is invariably considered to be a function of the mind, it is virtually impossible to separate theories on the evolution of language from theories on the evolution of mind. Perceptions of what the mind entails have radically influenced subsequent theories on language. The converse is also true; Love (2004) argues that the classical language myth – that language is a fixed code with an inventory of determinately identifiable linguistics units that correlate form with meaning – lies at the root of the classical view of mind. The aim in the remainder of this chapter will be to highlight some developments within both theories as they pertain to the theories on language and mind/cognition that we have discussed thus far and to indicate how these theories may benefit (or may have benefited) from an evolutionary perspective. The argument will be made that any theory on language and mind that does not take the evolutionary perspective into account will necessarily be limited and even defective in its explanatory power.

3. Approaching Mind in the Twentieth Century

3.1 Philosophical Approaches to the Mind

Whereas extensive attention was given in the previous two chapters to prevailing approaches to the relation between language and mind in the twentieth century, relatively little has been said about mind as such. This section will sketch some of the more prominent approaches to mind in the twentieth century, with the focus on certain assumptions regarding mind that characterised such theories during this period, rather than on a comprehensive and chronological history of the subject. There are many excellent works available that do chronicle the minutiae of theoretical speculation on mind during this period and the reader's attention will be directed to some of the works pertinent to the aspects of the subject under discussion.

As we shall see, despite the impact of Darwinian theory on biology, much of the theory on mind in the twentieth century remained impervious to his work. While "the problem of mind" has been around as long as philosophy itself, "philosophy of mind" as a distinctive field within philosophy is a relatively recent development.[10] The emergence of philosophy of mind as a distinctive subject cannot be precisely dated, but Flanagan (1995:570) argues that it emerged in the late nineteenth century and during the first half of the twentieth century. It developed mainly as a result of the founding of scientific psychology (according to the lore by Wilhelm Wundt in 1879), and also from Franz Brentano's 1874 publication of his *Psychology from an Empirical Standpoint* (571). Brentano resurrected Aristotle's and Aquinas' notions of *intentionality*, which held that paradigm-case mental states such as beliefs, desires, hopes etc. have *intentional objects*, in other words, they are of or about something (571).[11] For Brentano, intentionality differentiated the mental from the non-mental and psychology, as the study of the mental, would need to give an explanation for it (Brentano 2002:479).[12]

Another contributing factor to the development of philosophy of mind was the publication in 1890 of William James's *Principles of Psychology* (Flanagan 1991:23-53). With this compendium of psychological knowledge gleaned from the new science of psychology, and from traditional philosophical thinking about mind, James recognised that there was a deep tension between a scientific theory of mind and traditional philosophical ways of thinking about the subject. Whereas Descartes considered the "higher mental processes" to be impervious to scientific study[13], James attempted to formulate a naturalistic position with regard to mind.[14] He was inspired by Darwinian theory, and approached the mind as a

---

[10] It is interesting to note that in *The Problems of Philosophy*, Russell, writing in 1911, did not consider the study of the human mind as being part of philosophy any longer, and describes this subject as having "become the science of psychology" (Russell 1962:155).

[11] As will be discussed in section 3.4, Dupuy (2000) argues that this standard interpretation of Brentano's thesis within analytical philosophy is, in fact, erroneous. Dupuy's contention is that Husserl's interpretation of what Brentano means with intentionality is the correct one.

[12] Flanagan (1995:571) notes that Freud studied under Brentano when he was at medical school and argues that it may be fruitful to think of psychoanalysis as involving an extension of Brentano's basic insight. We have also seen in Chapter 1 that Peirce was influenced by Brentano's concept of intentionality.

[13] One of the main reasons for the inference from Descartes' dualism that the mind is impervious to scientific study, is the contention that the mind ("soul") is "of such a nature that it has no relation to extension, or to the dimension or other properties of matter of which the body is composed" (Descartes 2002:22). Furthermore, the mind/soul was seen as being divisible from the body, when we "break up the assemblage of the body's organs" (22). In a mechanistic understanding of the universe, disembodied, invisible, and immaterial entities cannot be conceived of as in any way akin to visible, tangible "matter". As we have seen in Chapter 2, this consideration even led to discomfort with Newton's theory of gravity. With the development of the information sciences, however, it has become easier to conceive of matter (such as radio waves) as invisible, intangible, apparently immaterial, and yet in any event "material".

[14] As Flanagan (1991:24) explains:

> On the naturalist view, mentality has no metaphysically odd properties… The naturalist sees the world as comprised of physical objects, their properties, and their relations. The naturalist departs from traditional materialism however, by denying that mental phenomena, naturalistically interpreted, require a simple mechanical analysis, for example, a reflex analysis. The naturalist parts company with the reflex mechanist because he reads revolutionary theory as pointing not merely to an increasing complexity

product of natural selection. He could not, however, reconcile free will with his science of mind[15], and his theory reached an impasse that seemed insurmountable. The objection that a naturalistic approach to mind necessarily entails determinism and undermines free will still haunts philosophy of mind and causes resistance to such projects.[16] Together, Wundt, Brentano, and James laid some of the most significant foundations that would underlie the philosophy of mind in the coming years. They would also leave a powerful legacy of philosophical assumptions about the nature of mind, about various models of mind, and recommendations about the proper way in which mind should be conceived of and studied.

Flanagan (1995:571) identifies the next phase in the development of philosophy of mind as stretching from 1900-1950, when scientific psychology emerged as a collection of competing theories with radically different methodologies. In the first half of the twentieth century philosophy of mind was dominated by a dualistic approach, informed to a large extent by Descartes' legacy. His fundamental distinction between, and division of, the mental and the physical proved to be a persistent and pervasive idea. Its tenacity is most probably due to the resonance that a mind and body with distinctive characteristics has with western philosophy's Greco-Christian heritage and with our common-sense understanding of something like a mind or soul residing "within" our bodies. Flanagan (571) succinctly describes twentieth century philosophy of mind as follows: "There were introspectionists, and anti-introspectionists, behaviourists and functionalists, depth psychologists and their opponents". Fundamentally, debate arose about the assumptions and methodologies proper to psychology and, in many ways, philosophy and psychology were still intertwined. Dewey and Carnap were to become representative of the "two sides" of the philosophy of mind, the one side primarily concerned with the metaphysics of mind, and the other with the methodological foundations of psychology and the epistemological status of psychological reports (572). Dewey, inspired by Darwin and James, aimed to develop a conception of mind that was naturalistic, without being mechanistic. He conceived of the mind as an adaptive organism, shaped by the forces of natural selection. Carnap, on the other hand, was motivated by the positivist's appeal to observational reports (or perceptual reports) as the foundation for all science, and concerned himself with the epistemological status of first- and third-person psychological reports (572).

Much of the subject matter that would characterise twentieth century philosophy of mind derives from attempts to clarify our knowledge of the "objective world" in terms of our immediate, and presumably inherently subjective experience. Livingston (2002) argues that many prominent research programmes around the turn of the twentieth century were

---

at the level of explanation as we ascend the phylogenic scale: reflexes require reflex analysis; full-blown mental phenomena require mentalistic analyses.

[15] As became apparent in the first chapter, Peirce's grappling with free will was one of the reasons that he moved toward a position of "objective idealism".

[16] Refer to Chapter 2 where Chomsky's contention, that the primary obstacle in overcoming this apparent impasse is a somewhat naïve mechanistic view of the natural world, is discussed. See Dennett (2004) for a proposed (naturalist) way out of this apparent deadlock.

concerned with this project. He also argues that the inauguratory move of twentieth century analytical philosophy (and ultimately philosophy of mind) was to integrate this prevailing concern with "a programme of linguistic and logical analysis" (17). The intention with this move was to specify the relationship between immediate, subjective experience and objective knowledge through analysing the logical relationships between propositions that describe immediate experience and other, more "objective" propositions (17). Livingston further argues that subjective experience was equated with "consciousness", while objective knowledge was equated with "scientific explanation" – a very specific understanding of scientific explanation, at that. Scientific explanation in this context was (and often still is) understood in structural or functional terms. Here, scientific explanation (primarily understood as something akin to the methodology of classical physics) is seen as being geared towards explaining the structure and function of physical entities. The assumption is that consciousness entails something over and above "structure and function". Hence, explaining the structure and function of consciousness would ultimately not explain why or how such material structures and functions give rise to consciousness (e.g. Chalmers 2002b:247ff). Livingston gives a useful description of this common conception of physicalist or scientific explanation. His characterisation mirrors Peirce's mechanistic conception of the universe. We have seen in Chapter 3 that Chomsky argues that such mechanistic conceptions of the universe is outdated and somewhat naïve. This picture assumes that (Livingston 2002:23):

> …reality consists ultimately in elementary particle, or whatever basic units of matter our best physics tells us everything else is composed from, in causal relationships to one another. Accordingly, higher-level entities like molecules and cells are arrangements of the underlying units, and their properties can be deduced (at least in an idealised sense) from the properties of the underlying units. This makes for a unified logical structure of explanation in which all of the causally relevant properties of entities described by specialised sciences, including psychology, can, in principle, be explained in terms of, or reduced to, properties of the underlying units.

In such a structuralist understanding, scientific explanation would be comprised of a totality of propositions that refer to the structural and functional characteristics of the entities to which they refer. Thus, science describes the totality of relationships between physical entities, which forms a "logical network of explanation" (Livingstone 2002:23).[17] The logical

[17] What such a mechanistic understanding of the physical universe lacks, is the concept of feedback. As will become clear in the discussion on connectionism, the concept of feedback becomes extremely important in accounting for information processing in neural networks. Furthermore, it should be kept in mind that Darwin's theory of evolution also allows for feedback through the principle of natural selection. Metaphorically speaking, one can describe the weeding out of un-advantageous mutations through the death of the carriers of such mutations as nature's way of providing feedback in the evolutionary process.

unity of scientific description means that meaningful scientific propositions can be deduced from the structure of scientific propositions.

The structuralist/functionalist conception of scientific understanding underscored Carnap's (1928) distinction between the structural nature of objective scientific statements and the subjective nature of the experiences of individuals. His contention was that if we wanted to establish (objective) consensus on the knowledge that we form in terms of these experiences, we have to resort to the formal description of the structure of these entities of experience. Hence, his argument is essentially structuralist in that he believed that the way in which our individual experiences are structured, is informed by the immutable structure of the physical universe, despite possible idiosyncrasies in the content of those experiences.[18] Carnap argued that knowledge that is objective is such that when we formulate the experience on which that knowledge is based in language (*protocol language*) and compare it to sentences of the *physical language* (the language in which the unified system of science should formulated) and these systems are "inter-translatable", that instance of personal experience is confirmed (Carnap 2002 [1933]:39ff). The physical language in this instance would consist of a sentence about the "physical state" of the person. Carnap advocated the adoption of this "physical language" as the system language of all science, which would ostensibly lead to all science becoming "physics" and would render metaphysics obsolete. Our language would then exclusively about physical occurrences and its "law" would be all-encompassing. This move by Carnap allows the possibility that the analysis of the logical relationships among the concepts of language could yield epistemological insight (Livingston 2002:24). Subsequently, analysis of the language that we use to describe the world would allow us to differentiate between the logical structure of a proposition (the objective and scientific aspect of such propositions) and its individual empirical content. Much of the research project of twentieth century analytical philosophy, including matters pertaining to mind and consciousness, was to be founded on this assumption.

Peirce's semiotic preoccupation thus found a new life within twentieth century analytic philosophy – although in a physicalist, rather than a panphysicalist guise – in its belief that language is fundamental to our understanding of the world. What would follow were numerous attempts at reconciling subjective experience with objective knowledge in physicalist terms, including behaviourism (e.g. Ryle), Identity Theory (e.g. Place, Smart and Fiegl) and functionalism (e.g. Putnam and Armstrong). Livingston (2002:27) formulates an elegant and concise description of the physicalist project:

> Physicalism itself, begun as the semantic doctrine that all meaningful sentences can be expressed in a single, unified language, which is also the language of physics, soon became an overarching ontological view about the

[18] Carnap's structuralist understanding of our epistemology and the fact that he relates that structure to language echoes the structuralism of Sausurre and Derrida in many ways (see Chapter 3).

nature of entities and their relations of reduction. This ontological view developed into a sophisticated and self-admittedly metaphysical picture of causation, function, and reductive explanation. Throughout this development, philosophers defended physicalism and its offspring partly for its explanatory advantages in the philosophy of science; but more often, it was the problems of mind, intentionality, and consciousness that provided the context in which physicalism was articulated and defended and employed as an explanatory resource.

3.1.1 PHYSICALISM/ MATERIALISM: VARIATIONS ON A THEME.

Many consider Ryle's *The Concept of Mind* (2002 [1949]) to be the founding document for contemporary philosophy of mind. In it, Cartesian dualism is rejected, the doctrine of privileged access to the mind through introspection is criticised, and it is proposed that "mind" in its Cartesian conception is a needlessly mystifying way of speaking about certain behavioural dispositions of the organism. According to Ryle, "mental states" such as "beliefs" and "desires" are nothing other than the disposition of the organism to behave in certain ways. He (2002 [1949]:32) rejected dualism as the "official theory" of "the Ghost in the Machine" (34) and argued that the mind is an aspect of human behaviour. Hence, Ryle can plausibly be characterised as a behaviourist (although he denied that his position was a behaviourist one).

Ryle (2002 [1949]:33) explicitly identified the philosophical assumption underlying dualism, namely the assumption that that there are two different types of existence or status, the mental and the physical. A necessary feature of physical existence is that it is in space and time, and is made up of matter. Mental existence, on the other hand, exists in time, but not in space and "consists" of consciousness, or is a function of consciousness. Furthermore, whereas matter can be causally connected and can thus be causally acted upon, the mind is insular and cannot be acted upon causally. Mental states and processes are usually thought of as *conscious* states and processes, to which the person who is experiencing them has direct and certain knowledge. Because of its "inexistence" in space, mental states can only be commented upon with authority from the first-person perspective, hence the "problems" of other minds and the apparent impossibility of objectively studying mental entities.

Ryle (2002 [1949]:34ff) argued that the dogma of the Ghost in the Machine is false in principle and based on a category-mistake and is thus a philosophers' myth. He, like Chomsky (see Chapter 3), traces this myth back to Descartes' reaction to Galileo's scientific methodology and its ability to provide a mechanical theory of matter, as well as the Greco-Christian tradition of thinking about the soul. Ryle argues that Descartes (partly due to his religious convictions) could not conceive of the mind (which he often equated with the soul) in mechanical terms, which seemingly rendered human beings as differing "only in degree of complexity from clockwork" (36). Conceiving of the mind as being of a "non-mechanical"

nature allowed an escape-route from this conclusion, in that the mind would not then be subject to seemingly incontrovertible mechanical laws. Ryle calls this the "para-mechanical" hypothesis, where "minds are not bits of clockwork, they are just bits of not-clockwork" (36).

Ryle's own position is that everyone already knows how to apply mental-conduct concepts to other people's behaviour. Furthermore, the introduction of the mechanical causal hypothesis does not change our ability to ascribe mental-conduct terms to others. Hence, he concludes, Descartes "had mistaken the logic of his problem" (2002 [1949]:37). He argues that there is no necessary reason to conjoin mental and physical processes and expect them to be logically equivalent categories. His contention is not that mind should be reduced to matter (materialism) or matter reduced to mind, (idealism) because, he argues, both of these positions are "answers to an improper question" (38). Instead, he proposes a "logical behaviourism" where speaking meaningfully about an aspect of mind is to say something about the disposition of a person to behave in a certain way.

After Ryle, philosophy of mind proper is characterised by two kinds of preoccupations. Flanagan (1995:572) divides them into firstly, "works of analysis" which is work devoted to the philosophical analysis of mental concepts such as sensation and perception, intentionality, free action, the emotions, reasons and causes, the possibility of private language, and of knowledge of one's own mind and other minds. Secondly, we have a body of work devoted to developing materialistic alternatives to Cartesian dualism and (and to each other).The three main theories in this materialistic approach are type-identity theory (where every type of mental state is identical to an as-yet-undiscovered neural state), eliminative materialism (which raises the objection that our everyday mental terms might not map onto neural states at all, and might be wholly fictional, such as is the case with phlogiston), and functionalism (where mental events are physical events, but are not reducible by definition or natural law; a belief can have multiple physical realisations) Materialism was less concerned with the methodology of linguistic analysis of Ryle and the analytic tradition and more with the possibility of formulaitng empirically-based theories of mind. However, it seems that despite Ryle's critique of the "official theory" a latent mind-body dualism would resurface throughout the philosophy of mind of this period.

Identity theory or (type-identity theory), as developed by Place, Smart, and Fiegl in the late 1950's, held that mental states are identical to brain states. This move was an attempt to move away from the analysis of concepts to an empirical methodology. Hence, as Place argues, the hypothesis that consciousness is a process in the brain should be treated as a reasonable scientific hypothesis, and cannot be refuted by logical argument alone (2002 [1956]:55). He goes on to argue that the "phenomenological fallacy" which is perpetrated in discussions about the mind, causes the problem of "providing a physiological of introspective observations" to be made more difficult than it is. The "phenomenological fallacy" is the fallacy of mistaking descriptions of the appearances of things in the mind as descriptions of the actual processes taking place "in the mind" (55, 59). Hence, he concludes that accepting that we have inner processes does not entail accepting mind-body dualism.

Similarly, Smart (2002 [1959]:60ff) argued that, because science increasingly showed that organisms "are able to be seen as physico-chemical mechanisms", it is extremely plausible that human behaviour could also be explained in such terms, consciousness included. Smart held that there were no philosophical arguments that compelled one to be a dualist. He also argued that equating sensations with brain-processes does not imply that "sensation statements" can be translated into statements about brain-processes or that these two sets of statements have the same logic; it only entails that sensations statements are statements about brain-processes. In many respects, he argues, a dualistic position rests on an article of faith, rather than logic or empirical proof (67).

Identity Theory, although making a compelling case for the identity of consciousness or mental-states, and brain-states on logical grounds, could not give an indication of what a physicalist description of mental-states would look like. The best that Identity theorists could do was make confident predictions that science would one day be in a position to give some sort of physiological description for mental processes. This inability allowed theorists like Putnam, Fodor, Armstrong and Lewis to formulate Functionalist objections to Identity Theory.

Consequently, Putnam argued that Identity Theory cannot preclude the possibility that mental states are "multiply realisable". He argues (2002 [1973]: 75) that pain is not a brain-state in "the sense of a physical-chemical state of the brain" but that it is an entirely different kind of state. Putnam characterises this different kind of state as a functional state of a whole organism. Putnam's theory is more empirical than the theories of mind against which he reacts, in that he makes use of developments in the computational approach to mind that were serving to revolutionise theorising on mind (see section 3.2 below). Making use of the relatively recently developed notions of the computer (Putnam explicitly makes use of the notion of the Turing Machine[19]), Putnam and other functionalists could conceive of mental states as akin to computational states (75ff). The organism is equated with a Turing Machine and its mental states are seen as functional or formal states that result from the working of the organism, similar to the computational states of the system. Mental states are the result of the functional organisation of the organism. Putnam holds that Identity theorists conceive of brain-states as physical-chemical states of the brain, rather than functional or organisational states. He believes that functional states of whole systems are "something quite different" (2002 [1973]:76). His argument is that conceiving of mental states as functional states of an organism as a whole, allows for the possibility that any organism (not just human or mammal) could possess a functional state that corresponds to a mental state such as pain, even if that organism were to be made of "stuff" entirely different from the (physical-chemical) "stuff" that the human brain is made of. His contention then is that a state such as pain cannot be identical to a specific brain state, because a creature different from humans (a Martian, for example) could have a pain, without having the brain state that accompanies pain in

[19] Cf. section 3.2 below.

humans.[20] Thus, we cannot identify a mental state type (e.g. pain) with a physical state type (the brain state that underlies pain). However, it might be possible to identify mental state tokens (the specific pain of a subject) with physical state tokens (the specific brain state of that subject) (Chalmers 2002a:4-5).

In a similar vein, Armstrong (2002 [1981]:80ff) detects a general move in philosophy away from a preoccupation with conceptual analysis and back to, what he terms, first-order questions. He believes that developments within philosophy of science, such as the work of Karl Popper for example, had led to the realisation that science in itself is rife with speculation, assumption, theory and reasoning. Along with this insight came the further realisation that philosophy had a greater role to play within the scientific enterprise than "mere" conceptual analysis. With regard to the mind-body problem, Armstrong argues that philosophy can make a useful contribution with regard to our concept of mental states and with determining whether the concept is such that it can intelligibly be hypothesised that mental states are physical states of the brain. Showing that a scientific proposition is intelligible is an essential step in validating that proposition, and philosophy is in a unique position to provide such validation.

Armstrong (2002[1981]:82) proposes that mental concepts be subject to "causal analysis" where mental states are conceived of as "apt to be the cause of certain effects or apt to be the effects of certain causes". Thus, the effects caused by mental states will be certain patterns of behaviour (such as seeking out food because of hunger). The causes of mental states will be objects and events in the person's environment (a sensation of green is caused by a green surface in the environment) (82). Armstrong argues that it makes logical sense to explain a mental state, such as "purpose", as a causal concept. To then hypothesise that a purpose is a physical state or pattern of the central nervous system is, if not logically inevitable, at least plausible. It is also plausible that behaviour directed towards fulfilling purposes would utilise perceptions (about the person's current situation) and beliefs (about how the world works) in order ensure that it acts in the way that is most likely to bring about the desired end (83). Hence, Armstrong notes, a causal analysis of the concept of purpose leads to a causal analysis of perceptions and beliefs, although such an analysis would assign different causal roles to different mental states. Perceptions and beliefs can be characterised as structures with which we map or model the world for ourselves, while purposes are the causes that drive us to utilise these mappings.

Armstrong's analysis is functional in that he argues that corresponding mental states must be introduced together in order to make sense. Their function within the organism (person) as a whole has to be taken into account in order for them to make causal sense. Armstrong's theory is much more complex than the outline given here; however, this outline is sufficient to demonstrate the value of his Causal theory. His theory retains elements of both

[20] As we shall see in the rest of this chapter and in the following chapter, the move towards conceiving of "the mental" in terms of the functional organisation of the nervous system was crucial, in that it allowed for a way to conceive of mental states as physical in that they are the result of certain processes, rather than properties.

behaviourism and introspectionism, in that he argues that we are conscious of mental states as they are (directly or indirectly) linked to behaviour. Hence, we are aware of mental states because of the fact that they drive us to given behaviour, rather than in terms of their "intrinsic nature". Furthermore, he supports identity theory, in that, given that a brain state plays a causal role in producing a pain, that particular brain state can be equated with the pain that is felt. Armstrong (2002 [1981]:84ff) also argues that he gives an explanation for *intentionality*, a conundrum that some philosophers believe undermine all attempts at materialistic explanation of the mental (e.g. Searle 1981, 1983; 1992).[21] Armstrong (2002 [1981]:84-85) understands the intentionality of mental states as "the mapping" or model in the brain in terms of which a subject acts. Despite Armstrong's move away from conceptual analysis, and his proposal that his causal functionalism can account for "mental states" such as intentionality in materialistic terms there subsequently remained a strong commitment to preserving mental concepts in philosophy of mind, even where the ostensible project is to account for mental entities in physicalist terms (e.g. emergentism, anomalous monism, and supervenience). Furthermore, it is often objected that functionalism cannot account for the "qualitative" aspect of mental states (qualia) (e.g. Block 2002 [1978], Nagel (1982), Chalmers (1995), Jackson (1982) and Searle 1992) and much literature is devoted to proving or disproving the ability of functionalism to account for qualitative states.

Another line of objection to the project of Identity Theory was that of "eliminative materialism" as developed by Rorty and Feyerabend and later amended by Patricia Churchland and Paul Churchland. Rorty and Feyerabend object that ordinary ways of talking about mental events was never meant to refer to neural events. Furthermore, there is no case to be made for the possibility that mental concepts inadvertently refer to neural states. Hence, eliminative materialism challenges the attempt to preserve some of our common-sense conception of psychological phenomena – in other words "mental" phenomena – in a materialist theory of mind. The argument is that the phenomena that we "perceive" upon introspection and attribute to other minds, such as beliefs, desires, intentionality, etc. need not be irreducible to a materialist framework, if we do not persist in insisting that a materialist theory account for mental states *in terms of our commons-sense understanding of them*.

Churchland (2002 [1981]:568ff) argues that the view usually adopted in philosophy of mind about what mental states are, is narrowly linked to what he terms a "folk-psychology". Thus, the conceptual framework in terms of which we understand mental states is based on a common-sense network of principles, concepts, and laws based on our need (and ability) to explain and anticipate the behaviour of others. As he puts it:

> Each of us understands others, as well as we do, because we share a tacit command of an integrated body or lore concerning the law-like relations holding among external circumstances, internal states, and overt behaviour.

[21] Intentionality will be discussed in greater detail in section 3.4 below.

> Given its nature and functions, this body of lore may quite aptly be called "Folk Psychology" (2002 [1981]:568).

Folk Psychology is extremely effective in predicting and explaining the behaviour of others, as well as in explaining one's own desire, wishes, and intentions. Our current conception of "mental states" stems from this system of accounting for the mind. Furthermore, such apparent "mental states" are articulated and communicated in terms of language. Thus, Churchland argues, the ostensible "intentionality" ("propositional attitudes") of mental states can be understood as being a structural feature of the concepts of folk psychology (2002 [1981]:569).

Churchland's argument is that Identity Theory expects the folk psychological concept of the mental to be reduced to a physical (neuronal?) level, *as is*. On the other hand, dualists scoff at this idea, because they believe that mental states are autonomous descriptions of *non-redundant* descriptions of a unique and non-physical domain of natural phenomena. Functionalists are sceptical about the possibility of reducing the "mental" to the physical, because the internal structure characterised by folk psychology is something quiet different in principle from the law-governed structures that we find in the physical world. In answer to all of this, eliminative materialism raises the possibility that it is our folk-psychology (our conception of what constitutes the "mental") that is inadequate to account for the actual complexity of what takes place physiologically. Eliminative materialism thus holds that our folk-psychological theory of the mental will ultimately be replaced by a better theory, most probably a theory from the perspective of natural history and the physical sciences (Churchland 2002 [1981]:572).

One of the most prominent functionalist objections against eliminative materialism is the insistence that folk psychology is *abstract* in that it is characterised by the network of causal relation in which mental phenomena stand to one another. Hence, mental categories make no reference to their own physical constitution. The assumption is that this mental network may be realised in a variety of physical systems – an argument strongly influenced by the notion of the Turing machine. However, as will be discussed in section 3.2 below, this assumption is most likely based on a misunderstanding of what the Turing machine implies.

Theorists like Fodor reject eliminative materialism on the grounds that the main objective of psychology is to find the best functionalist description of mental phenomena (Churchland (2002[1981]:573]). He insists that some form of "mental vocabulary" (some abstract functional characterisation of mental phenomena) will be ineliminable from psychology, even if it does not consist of the current folk psychology. Thus Fodor formulates a conception of cognitive activity as consisting of the manipulation of propositional attitudes (the "language of thought") (cf. Fodor 1975). As Churchland (573) sums up Fodor's position:

> …whatever tidying up FP [folk psychology] may or may not require, it cannot be displaced by any naturalistic theory of our psychological substrate since it is

on the abstract functional features of his internal states that makes the person, not the chemistry of his substrate.

Fodor's objection does not manage to defend folk psychology, per se, of course, but only some form of abstract description of the "internal states" that make the person. Furthermore, it is not clear why such "abstract internal states" are more suited to describing a person than "chemistry". This opposition is a false one in that it commits eliminative materialism, and any other materialist position, to reducing "personhood" or mental experience to chemistry. A materialist description of the mental need not imply that the mental be reduced to chemical workings at a neuronal level. Eliminative materialism's primary position is that folk psychology is ill-suited for developing a coherent theory about mental phenomena, while a general materialist position implies that mental phenomena, just like any other natural phenomenon, can be more accurately described in terms of some form of natural science, rather than in terms of an abstract theory derived from introspection, intuition, common-sense, and a very limited understanding of what rational thought entails.

This short overview of some of the major theoretical positions in twentieth century philosophy of mind does not begin to reflect the intricacy and sophistication of the many arguments for and against the various positions highlighted here. Nor does it do justice to the all of the theories that have been proposed in order to solve the various issues related to the mind. The discussion is meant as a brief guideline and overview of the positions salient to this work on the one hand, and also as an illustration of some of the philosophical commitments that underlie the current resistance to an evolutionary theory of mind. On the whole Livingstone (2002:31) gives an illuminating description of the foremost of these commitments:

Churchland (2002 [1981]:571) explicitly argues that folk psychology be evaluated in terms of its coherence with well-established and relevant theories, such as evolutionary theory, biology, and neuroscience. His position seems to be in the minority among philosophers of mind. What is striking about the emergence of the philosophy of mind in the twentieth century is the absence of evolutionary theory. James's and Dewey's early forays into evolutionary approaches to mind seem to have fallen by the way side in terms of philosophy proper. With regard to the other dominant approach to mind in the twentieth century, cognitive science, evolutionary theory is also conspicuously absent.[22] The reason for the virtual absence of evolutionary mind theory lies in a few factors. Firstly, as we have seen, Darwin's theory was subjected to severe criticism and was only redeemed with the development of the science of genetics.[23] Secondly, philosophy remained under the influence

[22] However, Dennett (1995:50) argues that Darwin's theory reflects insight into the "power of the algorithm" and in this manner links Darwinian theory with the computational approach that dominated the scene in twentieth century philosophy of mind.

[23] Currently, Darwinism (or neo-Darwinism) is also strongly (and very publicly) opposed by a band of vociferous detractors (see section 1). In many respects their concerns mirror those of earlier detractors of the theory. However, Darwinism has emerged from this "debate" quite unscathed in terms of the validity of the theory.

of Cartesian dualism on the one hand, and an a priori methodology on the other hand, well into the twentieth century.[24] These influences led to assumptions about mind and the methods through which it should be studied, which were fundamentally incompatible with the view that the mind could be an organism that evolved in increments to its present state in order to facilitate the reproduction of genes.[25] Finally, within the science of mind itself, despite the interdisciplinary approach in the emerging cognitive sciences, the mind was seen primarily as a functional/computational system rather than as a biological system, and this approach had fundamental implications for assumptions about the mind's structure, function, and status.

A very important factor in the resistance to the "reduction" of the mental to the physical in philosophy of mind is the early commitment within the discipline to studying the mental in terms of linguistic analysis. Despite the subsequent move towards materialism within the discipline, the assumption remains that a first-person linguistic account of "metal events" is a legitimate, and indeed the only legitimate, way of describing the mind. Mind thus becomes an irreducibly phenomenological entity, impervious to the scrutiny of science. In this manner, the intuitive (and presumably correct intuition) that language is somehow fundamentally intertwined with the mind has served to an excessive and unwarranted role being given to linguistic analysis in the study of mind. As will become clear in the following sections and the next chapter, language probably is fundamental to the mind. However, this does not logically lead to the conclusion that "mental events" mirror linguistic propositions, or that they are best studied in terms of first-person reports.

Dennett (1995:402) believes that the focus on language in twentieth century philosophy, and especially on meaning, has had a "debilitating effect" of theories of mind. He argues:

> …by concentrating first on linguistic meaning, philosophers have distorted their vision of the minds these words depend on, treating them as somehow *sui generis*, rather than as themselves evolved products of the natural world. This is manifest especially in the resistance philosophers have shown to evolutionary theories of meaning, theories that purport to discern that the

[24] In the introduction to a contemporary compendium of philosophy of mind, the editors note that, despite the rapid growth of the science of mind since the mid-twentieth century and the subsequent blurring of the boundary lines between sciences of mind and philosophy of mind, "the philosophy of mind at its core remains a branch of metaphysics, traditionally conceived. Philosophers continue to debate foundational issues in terms not radically different from those in vogue in previous eras" (O'Conner and Rob 2003a:xi).

The question arises as to whether this traditional "metaphysical orientation" can continue without being influenced by developments in the sciences of mind.

[25] Such assumption include: i) the states of the mind have "phenomenal feel", while the physical is "phenomenally lifeless"; ii) the contents of the mind lack spatial location, while the contents of the physical world are necessarily located in space; iii) some mental states are representational in that they *have* (original) intentionality, while the physical can be representational only when we bestow meaning on it, hence it has derived intentionality; iv) we discover truths about the physical world through perception, while we discover truths about our mental states through introspection; v) our knowledge about our mental states is more secure than our knowledge about the physical world; vi) the mental is private, while the physical is accessible to everyone (O'Conner and Robb 2003b:3-4).

> meaning of words, and all the mental states that somehow lie behind them, is ground ultimately in the rich earth of biological function.

Dennett (1995:402) holds that while few philosophers have wanted to deny the obvious fact that human beings and their capacity to speak and to "mean anything" is due to specific adaptations, they have been reluctant to accept that evolutionary thinking may contribute to the puzzle of how words, and their manifestations in minds or brains, have meaning. Prominent among those philosophers resistant to the idea he lists Jerry Fodor, Hilary Putnam, John Searle, Saul Kripke, Tyler Burge, and Fred Dretske. Their opposition is often cast in the form of opposing "real" meaning to "ersatz" meaning, or "real" intentionality, as opposed to "derived" intentionality.[26] According to these views, artefacts cannot have "real" or "original" intentionality or meaning, domains which are reserved for human beings (and animals) (cf. Searle 1981).[27] Dennett rejects these positions as "essentialism applied to meaning" (1995:408). He proposes that the "meaning" or "intentionality" associated with artefacts is no more derived or pseudo than those associated with human beings. Cast in these terms, there seems to be no way to distinguish between an instance of "real" intentionality or "pseudo" intentionality without appealing to a form of essentialism. However, conflating "real" intentionality or meaning with simulated intentionality or meaning seems like an equally disagreeable alternative.[28]

Outside of the developments in philosophy of mind proper during the 1960's, 1970's, and 1980's a new kind of approach to mind developed in the form of an amalgamation of subjects known as cognitive science.[29] As we shall see, the broad field of cognitive science made it clear that there are many aspects of mind that cannot be known on *a priori* grounds. It

---

[26] See Dennett's (1995:407ff) (entertaining and persuasive) thought experiments that point out the errors in these assumptions.

[27] Searle (1981:305) argues, for example, that only machines with the *same causal powers* as brains can think:

> Whatever else intentionality is, it is a biological phenomenon and it is as likely to be as causally dependent on the specific biochemistry of its origins as lactation, photosynthesis, or any other biological phenomena. No one would suppose that we could produce milk and sugar by running a computer simulation of the formal sequences in lactation and photosynthesis; but where the mind is concerned, many people are willing to believe in such a miracle, because of a deep abiding dualism: the mind, they suppose is a matter of formal processes and is independent of specific material causes in a way that milk and sugar are not.

As we shall see, Hayles (1999) and other theorists offer similar criticisms of the project of artificial intelligence and other disciplines that treat the mind as computation. We will argue that approaching the mind as a product of evolution through natural selection may provide us with a way to account for intentionality, without having to postulate that it is somehow "essentially" biological.

[28] Unless, perhaps, we attempt to explain intentionality in terms of the "goals" of the system/organism that exhibits intentional behaviour. We will argue that an evolutionary approach may provide a way out of this quandary. Perhaps the causal powers that Searle (1981:305) speaks of should be replaced with "design-goals" whether evolutionary or otherwise.

[29] Haugeland (1981a:1) describes the term *cognitive science* as covering enterprises with the common conceptual foundation that "reason is but reckoning" in Hobbes's formulation. He argues that the term does not apply to every scientific theory of cognition, but only to those sharing the "information processing" or "symbolic manipulation" approach.

has even raised the possibility that there is no subfield of philosophy proper that can deepen our understanding of the mind, at least not on its own. As Flanagan describes this realisation: "Mind will be understood, if it is understood, by our best science" (574). With this in mind, Flanagan poses the question, which we will come back to: In what sense, one might ask, is the question of the nature of mind a philosophical question rather than a foundational question within the science of mind?

3.2 Mechanising the Mind

The beginning of the 1950's brought on the "cognitive revolution", in the form of the development of computationalist and connectionist theories of mind, which led away from abstract speculation about the mind and to the idea that mental processes could be studied just like any other physical process. From this realisation the science and philosophy of Artificial intelligence was born. The idea arose that the mental world can be rooted in the physical world through the concepts of information, computation, and feedback, and that the mind can be modelled in terms of information processing machines. In this view mental phenomena such as beliefs, intentions, and memories are composed of information; information which is embodied in patterns of activity within the brain (Pinker 2002:32). Pinker condenses this broad approach as follows (2002:32):

> Thinking and planning are systematic transformations of these patterns, like the operations of a computer programme. Wanting and trying are feedback loops, like the principle behind a thermostat: they receive information about the discrepancy between a goal and the current state of the world, and then they execute operations that tend to reduce the difference. The mind is connected to the world by the sense organs, which transduce physical energy into data structures in the brain, and by motor programs, by which the brain controls the muscles.

Dupuy (2000:30) argues that this change in outlook was the result of developments within mathematics in the 1930's. Until then the Hobbesian notion of genetic definition was thought to hold in mathematics, namely that all properties of an object can be derived from its definition. The work of Gödel and Turing, however, caused this notion to be called into question. As Dupuy describes it, their work led to the conclusion that "while we are the source of mathematical entities, some of these remain forever beyond our reach" (32) and he argues that this conceptual revolution was the point of departure for the new science of mind.

In 1931 Kurt Gödel published his paper, "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems". In it he established an "incompleteness

theorem", namely that any formal system in the logical sense[30] that is rich enough to accommodate arithmetic has one of the following properties: either the system is inconsistent in that it generates contradictory theorems, or there exists at least one true proposition that cannot be proved within that system (Hofstadter 1984:15-19; Dupuy 2000:33). Whereas it had once seemed obvious (to theorists such as Frege and Russell, for example) that every mathematical truth is provable, it now seemed that all properties of mathematical entities were not derivable from their definition.[31] Mathematics had been considered to be derived from logic, and its certainty lay in its foundation in the human mind. It was thought that the logical products of mind, by virtue of its property of reason, could be known by the mind and be true. Thus, ironically, when it transpired, thanks to Gödel's theorem, that mathematics, even though it is a product of mind, may not be demonstrably true, the possibility arose that the mind may in fact be amenable to study. Dupuy (2000:31-35) argues that, whereas mind had once been thought of as irreducibly opaque, because (unlike mathematical entities) it is not a product of mind (itself), it now seemed that even entities that are products of mind could be opaque. Yet, such products of mind could still be studied. The possibility arose that the mind may be coded or modelled, and hence, studied. In constructing his theorem, Gödel demonstrated that it was possible to use integers to code formulas and sequences of formulas. A proposition that asserts that a given formula is provable can be expressed in terms of a mathematical proposition – hence, logic could be arithmeticised. Thus, Gödel indicated that logic can be arithmeticised, which gave a foundation to the maxim that reasoning can be characterised as computing with integers (Dupuy 2000:34). A classic paradigm had been shaken and this would form the point of departure for the new science of mind.

In tandem with Gödel, Alan Turing had a great impact on this shift beginning with his seminal 1936-1937 paper, "On Computable Numbers, with an Application to the *Entscheidungsproblem*", which heralded the beginning of a new approach to the science of mind.[32] His aim with this paper was to solve a logical problem posed by Hilbert, the *Entsheidungsproblem*, namely: "given any formula of the predicate calculus, does there exist an effective, general, systematic procedure for deciding whether or not the formula is provable?" (Dupuy 2000:33). Gödel's theorem went some way towards addressing Hilbert's problem, however, it lacked a rigorous, mathematical definition of the *algorithm* as the mechanical operational procedure within formal systems (Penrose 1989:45; Dennett 1995:50-60).[33] As Penrose (1989:45) puts it, the *Entscheidungsproblem* that concerned Turing was

[30] Dupuy (2000:33) describes such a formal system as one that provides a formal language, thus one with a set of formulas that function as axioms, and with rules of inference.

[31] See Hofstadter (1984:19ff) for a short discussion of the context in which this conviction arose.

[32] See Pylyshyn (1981:68).

[33] As Dupuy (2000:34) puts it, what was needed was "the formalization of the intuitive notion of an automatic procedure, governed by fixed rules and devoid of anything akin to meaning, interpretation, or creativity."

whether there exists "some general procedure [*algorithm*] which could, in principle, solve all the problems of mathematics (belonging to some suitable well-defined class)".

The concept of "mechanical procedure" was alien to normal mathematical theory at that time and it was thus necessary to clarify the notion in order to address the *Entscheidungsproblem*. In this regard Alonso Church suggested that every computable function is a recursive function (in the mathematical sense),[34] although his thesis could not be proved.[35] Turing attempted to formalise the concept of a "machine" and its operation, breaking it down into elementary terms (Penrose 1989:45). He presented a mathematical formulation for the abstract notion of a machine as a devise for carrying out a finitely definable calculational procedure, or an algorithm (46). The model subsequently became known as the Turing machine.[36] Turing believed that he had succeeded in penetrating the essence of mechanical procedure with his model. His (Turing 1948:7) thesis was that every mechanically computable function is computable by a Turing machine (the first conception of the computer), although this claim was not provable.[37] Turing managed to reveal a kinship between the notion of effective computability and that of mechanical procedure. In terms of Hilbert's problem, Turing could rephrase it to: "Does there exist a (Turing) machine capable of deciding whether or not such a formula is provable?" (Dupuy 2000:36). He was able to answer this question in the negative.

Turing also developed a Turing machine which he called the universal Turing machine. The universal Turing machine was capable of simulating the behaviour of any other Turing machine. Whereas other Turing machines were envisaged as models of thinking[38], the universal Turing machine constituted the model of models. As Hopcroft and Ullman (1969:102) put it: "One can think of a universal Turing machine as a general purpose computer which is powerful enough to simulate any computer, including itself."[39] Modelling through simulation, as happens in the case of the universal Turing machine, involves

---

[34] See Penrose (1989:86-92).

[35] His theorem could not be proved, as Dupuy puts it, "any more than the mathematical definition of a sphere could be proved to be identical with our intuitive concept of a round object" (2000:34). However, it did provide a notion of effective computability.

[36] The Turing machine is an idealisation of an ideal, finite calculating agent and hence an abstract mathematical entity (Penrose 1989:45-56; Morten 1995:883-884). It consists of an infinite tape (which represents the machines memory and is divided into squares), a head which reads symbols from a finite list and writes or erases symbols accordingly, and a finite number of states. A machine table with instructions that specify in every given state what could be overwritten at that point, as well as the next state, in accordance with the mark that is written in the square (or not). The head can move one square at a time to the left or the right, or stand still (cf. Putnam 1980:365; Copeland 2000a:1-5). The output of the Turing machine is completely determined by the input, which is made up of the configuration of the machine at a given moment, i.e. the machine's internal state and the content of the scanned square.

[37] Pylyshyn (1981:68) declares that this thesis cannot be proved, but that it "has withstood all attempts to fund exceptions."

[38] Copeland (2000b:8ff) is at pains to point out that Turing did not intend to provide a model of human thinking in general, but as an idealised description of a specific human thinking activity, namely *numerical computation*. Thus, the Turing machine "is a model, idealised in certain respects, of a human being calculating in accordance with an effective [mechanical] procedure" (8).

[39] See Hopcroft and Ullman (1969:102-107) for a description of universal Turing machines.

reproducing the functioning of a system. In this sense the universal Turing machine can be seen as a possible model for a certain understanding of mind (conceived of as human beings engaged in computation). Copeland (2000b:8) quotes Turing (1950a:436) as follows in this regard:

> The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer.

Turing's approach was to treat a formal system in the logical sense as a mechanical procedure capable of producing new formulas called theorems (Dupuy 2000:36). Hence, the Turing machine is an arithmetic computer, and since Gödel showed that logic can be arithmetised, the Turing machine is also a symbolic computer; Turing thus formulated a model of symbolic thought (37). Turing's work held implications for the relation between thought and machine, and his ideas led to the development of cybernetics in the 1940's and shortly thereafter cognitive science.[40] Eventually his ideas would serve as a basis for a functionalist approach to the mind-body problem within philosophy. Dupuy (2000:38) argues that the doctrine of computational-representational functionalism, as it took shape within artificial intelligence and in the philosophical projects of Pylyshyn and Fodor, for example, was the result of taking the metaphor of a computer consisting of a Turing machine embedded in electronic circuits literally (see Putnam [1980:370-373] where he discusses possible implications if the abstract Turing machine is physically realised). The mind was conceived of as a Turing machine operating in terms of a private, internal language. In this understanding the symbols of the mental Turing machine enjoy a triple mode of existence: i) they are physical objects embodied in the neurophysiological system, and hence are subject to the laws of physics; ii) they have form because they are governed by syntactic rules (analogous to laws of inference in a formal system); and iii) they are meaningful and have a semantic value. A parallel is drawn between physical processes subject to causal laws and mechanical processes carrying out computations. A parallel is also drawn between syntax and semantics. The computer model has trouble formulating an adequate explanation for how symbols acquire meaning, however, as it completely disregards the meaning-giving programmer who is the origin of the meaning of the symbols employed by the system (Also see Haugeland 1981:1-34).

The Turing machine is abstract, in that its description defines all possible mechanical computational operations, *computable by human beings* (Copeland 2000b:9). The physical construction of a Turing machine is unimportant, provided that it could carry out the instructions. The pioneers of cognitive science worked with the notion that the activity of thought (the faculty of mind that had knowledge as its object) is in essence a rule-governed

[40] See Newell and Simon (1981:35-66).

mechanical process, which required no insight (Dupuy 2000:39). With time Turing's thesis was (erroneously) taken as proving that a thinking machine in human terms is possible in practice, and it was accepted that all activity of the human mind can be accurately described and is executable by a suitably programmed computer.[41] These unquestioned notions would form the heart of much of the theory on mind in subsequent years.[42] The idea arose that the "construction" of the mind is not important for intelligence. Thus, Turing's work spawned a conception of human beings as, in Hayles's (1999:7) phrase, "information processing entities who are *essentially* similar to intelligent machines."[43] The "embodiment" of the mind was considered to be immaterial to research into the mind and its activities.[44] The result was a cognitive revolution, where the mind came to be regarded as an entity that could be studied, as could any other object of scientific interest. As Dupuy writes (2000:40):

> The Turing thesis, in spite (or rather because) of the ideological distortions to which it so readily lent itself, was what it took to rally the resources of energy and intelligence needed to bring about the birth of a mechanistic and materialistic science of mind.

One of the results of the revolution in logic in the 1930's was that its implications inspired the move towards founding a science of mind in the 1940's; a movement which would come to be known as "cybernetics". Furthermore, cognitive science as we know it

---

[41] It is important to note at this stage that Turing had not, as is commonly asserted in philosophy of mind and computational theory of mind, established that the universal Turing machine can simulate the behaviour of *any* machine (Copeland 2000b:4). What he proved is that his universal Turing machine can compute any function that any *Turing* machine can compute. Hence, assertions that psychology must be capable of ultimately being expressed in terms of the Turing machine are erroneous. It is entirely possible that mechanical models of human cognition not equivalent to Turing machines are possible (13). Turing's thesis does not entail that the brain can be simulated by a Turing machine (14). In terms of Turing's thesis human brains are Turing machines in so far as they engage in the activity of numerical computation.

[42] See, for example, Haugeland (1981:13):

> But the fundamental importance of Turing machines rests on yet another truly amazing fact – *a theorem first proved by Turing* [TdV] – which has, perhaps more than any other single result, shaped modern computer science. It is that there are special Turing machines, called universal Turing machines, which can be programmed to imitate any other Turing machine. In particular, one could imitate a Turing machine that was itself imitating some other automatic formal system – which means that, indirectly, the universal machine is also imitating some other automatic formal system. So, combining Church's thesis and Turing's theorem, a universal Turing machine can (suitably programmed) imitate any automatic formal system whatsoever!

[43] Both Hayles (1999) and Dupuy (2000) devote extensive space to the role that the Macy Conferences on Cybernetics, which ran from 1943 to 1954, played in establishing this new paradigm. To paraphrase Hayles (1999:7) the major contributions to this new paradigm were Claude Shannon's theory of information, Warren McCulloch's model of neural functioning, John von Neumann's computers that processed binary code and could conceivably reproduce themselves (and hence be considered analogous to biological systems), and a visionary that could appreciate the larger implications of the cybernetic paradigm, which they had in Norbert Wiener.

[44] See Searle (1990) for a critique of this assumption. He argues that proponents of the view that the brain is a digital computer ignore the important role that syntax plays in the computational process and the fact that syntax is an observer-relative notion. Hence, he concludes that the homunculus fallacy is endemic to cognitivism, viewing the brain as if there is an agent within it, using it to compute.

today had its origin in cybernetics.[45] However, the mind as an "unembodied" information processing entity was studied in terms of models of information-processing and scant attention was given to the origins and possible functions of mind. Certainly, the notion of mind as a biological entity did not feature in this research programme. Furthermore, the implications of the developments in logic in the 1930's can be interpreted in an entirely different way, as was indeed done by John von Neumann.[46] Turing's theorems indicate that thinking is not the same as computation (Copeland 2000b:8ff). Dupuy (2000:40) even argues that "Turing's mechanical model was well suited to deepening the fundamental distinction between thinking and knowing developed by Kantian philosophy"

3.3 THE SCIENCE OF MIND

In general 1943 can be seen as the year in which a science of mind was founded with the publication of two articles by five authors who were to form the nucleus of the cybernetic movement. With the work of Rosenblueth, Wiener, and Bigelow (who together wrote "Behaviour, Purpose, and Teleology") two concepts central to the first phase of cybernetics were developed, namely information theory and the concept of the feedback loop.[47] The work of McCulloch and Pitts (who together wrote "A logical calculus of the ideas imminent in nervous circuitry" 1943) was inspired by the Turing thesis and was aimed at giving a neuroanatomical and neurophysiological basis for synthetic a priori judgements. In this way they hoped to found a biologically natural neurological model of mind (a project that came to be known as connectionism) (Hayles 1999:57-63; Dupuy 2000:45). Their proposed model of the mind was a network of idealised interconnected elements that were thought to work in a way similar to biological neurons. The connectionist approach is based on the insight that functionally the brain only consists of interconnected nerve cells or neurons.

Neurons are single cells, and consist of a cell body, the dendrites that convey impulses to the cell body, and the axon that relays impulses from the cell body (cf. Nathan

---

[45] Dupuy (2000:43-44) argues that cybernetics shaped our era in terms of technical, ideological, and theoretic points of view to an unimaginable degree. As part of its legacy can be listed: the logico-mathematical style of formalism and conceptualisation applied to the sciences of the brain and the nervous system; conceiving the design of information processing machines and laying the foundation for artificial intelligence; producing the metascience that came to be known as systems theory; inspiring conceptual innovations in economics, operations research, game theory, decision and rational choice theory, political science, sociology and other disciplines; and providing the metaphors for various scientific revolutions in the twentieth century in terms of which they could describe their break with the prevailing paradigm.

[46] Cf. Dupuy (2000:67-69).

[47] The concept of the feedback loop was conceived of as being intrinsic to systems regulation, based on the gap observed between the behaviour of a system (its output) and the projected result (input) (Dupuy 2000:45). Hence, where objects or systems were conceived of as devices transforming input messages into output messages, the feedback loop ensured that the system was capable of changing the relationship established between the input and the output. To the observer the system seems able to modify its response to a given input (stimulus) with the view to achieving a goal. It appears as if the system is able to learn from feedback and adjust its behaviour, and hence systems are not conceived of as operating according to simple stimulus-response rules. In the first phase of the cybernetic movement, account was not taken of the internal organisation of the object or system.

1998). The axon with its surrounding membrane is called the nerve fibre and the point where it ends on another neuron is called the nerve end. The nerve fibre connects to other neurons in the synapses. Receptors are connected to the central nervous system and convert the energy that they receive (from the senses and the rest of the body) into an electric current, which is then passed on to the neurons. The nerve fibres that connect neurons pass nerve impulses between various neurons. Each neuron can be likened to a “processor” that calculates the sum of its inputs, and then, if this sum succeeds a certain threshold, generates an output (Cilliers 1989:61). The output of this particular neuron, in its turn, becomes the input of all the neurons that are connected to it. Not only can neurons be connected to a vast number of other neurons at the same time (whether directly or indirectly), they can also be connected to themselves, usually with other neurons as intermediaries.[48]

The connections between neurons are mediated by synapses (a minute gap between the nerve endings of one neuron and the cell body of the next), which regulate the strength of incoming signals and can determine whether the signal would excite or inhibit the neuron. This arrangement allows for flexibility and is one of the keys behind the plasticity of the brain. A chemical substance – a neurotransmitter – is put out into the synaptic gap when an impulse reaches the end of a nerve fibre, and allows the impulse to pass the gap (if this sum exceeds a certain threshold) (Nathan 1998:518). The connection between any two neurons has a certain “weight” which determines the strength of the influence of these neurons on one another (Cilliers 1998:16). The characteristics of the neuronal network seem to be determined by the values of these weights.

The connectionist approach created idealised models of the brain, attempting to emulate its functioning. Nodes, which are analogous to neurons, are connected in a manner similar to the interconnection of neurons by means of dendrites and axons. The operation of a neural network is determined by a node’s reaction to the input that it receives from the nodes that it is connected to, by means of an input-output transformation rule (Cilliers 1998:25ff; Picton: 1994; and Deacon: 1997:129ff). Large numbers of nodes are connected together in a highly interconnected manner. The functioning of the network is thus determined by the global patterning of signals and not by the state or activity of any one node (Deacon 1997a:131; Cilliers 1998:27ff). As with neural pathways, the connections between the nodes of neural

[48] According to Hebb’s principle the connection strength of a synapse between two neurons should increase proportionally to how often it is used (Cilliers 1998:25ff). The stronger pathways’ synapses become more effective and are used more often, while unused pathways wither away (Young 1998:455). In this way the mass of largely undifferentiated neurons can develop a structure that is based on the information available to each neuron locally. In other words, the networks of neurons learn from inputs available to them and accordingly develop a (genetically constrained) structure, which means that networks can be "trained" to solve given computational problems through manipulating the connections strengths between given neurons.

networks can be modified so that the connection strengths between given nodes can be altered. In this manner, the structure and functioning of the network can be adapted to embody a given rule linking input function and output functions. In this manner, a neural network can be "trained", and mirrors (to some extent) the learning ability of the brain.

Another characteristic of neural networks that distinguishes them from their serial predecessors (digital or symbolic computers) is the distributed manner in which they embody information. Information is decomposed and embodied in global manner throughout the network. Information is "represented" in the logical connection between the nodes, rather than being "mirrored" in symbols as is the case with the Turing machine. Thus, whereas computational theories of mind had up to then conceived of the mind as a symbol manipulator[49], connectionism conceives of the mind as employing parallel distributed processing (PDP) to process information where much information is distributed throughout a network of interconnected neurons and is processed simultaneously. In this manner it was hoped that the brain (and mind) could be modelled more accurately than was possible with models that viewed the mind as a symbol manipulator.

Thus, whereas Wiener, Bigelow, and Rosenblueth denied reality to the mind, McCulloch aimed to identify the logical and material mechanisms in which the mind was embodied.[50] The aims of the two groups of theorists may seem at odds, but Dupuy (2000:50), for example, argues that what McCulloch was in fact doing was relocating the behaviouristic approach of Wiener *et al* inside the brain.[51] The contents of mind were now thought to be amenable to scientific study, but were in themselves thought to be constituted of the behaviour of smaller units inside the impenetrable brain. In the absence of the concept of feedback from the neural network itself in the early neural network models, the contents of mind came to be regarded as operations taking place in neurons, which transformed inputs into outputs. The behaviouristic and communicational approach of Wiener *et al* was retained and applied to a "logically lower level" (50). McCulloch helped promote the conception of brains as "computing machines". He held that not only the function of the brain (mind) could now be regarded as a Turing machine, but that its structure – the physical brain – could as well. In McCulloch's view both the mind and the brain could be regarded as equivalent machines, and hence mind and brain were, in fact, equivalent.

McCulloch, taking Turing's thesis to be proven fact and misinterpreting its implications,[52] remained convinced that his logic machine could model all the faculties

---

[49] In general, functional position in philosophy of mind that conceive of the mind as computational have this view of the mind as "symbol manipulator" in mind.

[50] Interesting in terms of our discussion is that McCulloch was an admirer of Peirce (Dupuy 2000:105).

[51] This view is strengthened by Picton's (1994:1) description of what a neural network does:

> "The function of a neural network is to produce an output pattern when presented with an input pattern."

[52] McCulloch made the same mistake that Copeland (2000b:4ff) accuses various theorists on mind of making, which is, as we have seen, interpreting Turing as proving that his machine can solve any problem that can be solved "by instructions, explicitly stated rules, or procedures" (Gregory 1998:784)

attributed to mind, namely perception, thought, memory, concept, formation, knowledge and the recognition of universals, will, and consciousness (Dupuy 2000:55). His ultimate aim was to model the capacity of the mind to form and know universals. He came to regard the 1943 model as inadequate for achieving this goal, which led him to add a *learning capacity* to his networks which allowed for traces of past events to be retained in the network and to influence the functioning of the network through *feedback*.[53] In this manner, it was hoped, the network would be able to generalise from examples that it had already learned [cf. Picton (1994:4) and Cilliers (1998:28-30)]. Furthermore, McCulloch and Pitts added randomness to their network, to allow the networks to function in the presence of errors and noise that may be introduced to the networks in the process of their functioning.

McCulloch and Pitts demonstrated, in principle, that their neural network could be equivalent to a Turing machine in that everything that can be done by the one can be done by the other. This strengthened McCulloch's position that brains "compute thoughts the way electronic computers calculate numbers" (quoted in Hayles 1999:58). However, an actual McCulloch-Pitts style neural network is *not* equivalent to a Turing machine; their networks did not have the same computational capacities as a Turing machine, which is, after all, idealised. The brain, as a finite organ, cannot compute everything that a Turing machine, with its potentially infinite tape or memory, can. Dupuy (2000:58) argues that McCulloch and Pitts and the other cyberneticians were the first to forget this limitation[54], and what had initially been conceived as a theory of real automata (actual objects that could embody the faculties of mind) became automata theory, an abstract, disembodied mathematical theory (Hayles 1999:57-63).[55] In subsequent years, research in cognitive science would be dominated by representational and computational functionalism of the mind as a Turing machine.[56] Various disciplines would take their cue from this cybernetic model, including systems theory, operations research, optimal control theory and decision theory (61).

---

and that universal Turing machines can "compute any function that any computer, with any architecture, can compute" (Dennett 1991:215).

[53] *Feedback* refers to the existence of causal relations between the network, its environment, and the previous states of the network.

[54] As Dupuy (2000:58) puts it:

> …throughout the Macy Conferences they went on repeating in assured tones the magic formula "Any behaviour that can be logically, precisely, completely, and unambiguously described, in a finite number of symbols, is computable by a neural network".

[55] Hayles (1999:60) argues that McCulloch did not entirely equate his logical model with the brain, because he remained dedicated to finding an "empirical epistemology", where he could combine embodied actuality with "the force of logical propositions". Hence, he was not in favour of disembodying information or neurons, which lead to a permanent tension within his work (62).

[56] Haugeland (1981:31) gives the following definition of cognitive science:

> The basic idea of cognitive science is that intelligent beings are semantic engines – in other words, automatic formal systems with interpretations under which they consistently make sense. We can now see why this includes psychology and artificial intelligence on a more or less equal footing: people and intelligent computers (if and when there are any) turn out to be merely different manifestations of the same underlying phenomenon. Moreover, with universal hardware, any semantic engine can in principle be formally imitated by a computer if only the right program could be found.

A theorist who was very aware of the idealised nature of the McCulloch-Pitts model was von Neumann. Dupuy (2000:60) quotes von Neumann on the subject:

> They wanted to discuss neurons. They took the position that they did not want to get tied up in the physiological and chemical complexities of what a neuron really is. They used what is known in mathematics as the axiomatic method, stating a few simple postulates and not being concerned with how nature manages to achieve such a gadget. They went one step further…They said that they did not want to axiomatise an idealised neuron, which is much simpler than the real one. They believed that the extremely amputated, simplified, idealised object which they axiomatised possessed the *essential* traits of the neuron, and that all else are incidental complications, which in a first analysis are better forgotten. Now, I am quite sure that it will be a long time before this point is generally agreed to by everybody, if ever; namely, whether or not what one overlooks in this simplification had really better be forgotten or not.

Von Neumann was inspired by the 1943 McCulloch and Pitts paper in his insight that the logical conception of a calculating machine could be separated from the design of its circuitry, i.e. the hardware/software distinction. Von Neumann abstracted the same logic machine from the brain that McCulloch and Pitts did. However, he placed the program that governed the operation of the computer on the same level as that data on which it operates. Furthermore, he conceived of a serial calculator, which made the computer capable of very high speeds. With regard to the relationship between logic and the brain von Neumann would come to the conclusion that the number of formal neurons (in the tradition of the McCulloch-Pitts model) necessary to produce properties of the mind would be substantially greater than the number of neurons actually in the human brain. He concluded that logic was too rigid and combinatorial to be considered as the only logic of either natural or artificial automata. He proposed to formulate a more complex logic, "whose variables could take continuous values, drawing upon the methods of probability theory, thermodynamics, information theory, and the calculus" (Dupuy 2000:68). Dupuy (68) argues that von Neumann thus posed the question of complexity, and foresaw that it was to become the greatest question for science in the future. In his view, McCulloch and Pitts reduced function to structure, hence not addressing the question of what a complex structure could be capable of.

By the middle of the 1950's Artificial Intelligence (AI) set off on its own as an infant discipline (Newell and Simon 1981:46), while cybernetics entered its second phase in the 1960's and 1970's. Whereas cybernetics persisted with the attempt to model natural intelligence, artificial intelligence freed itself from the constraints posed by natural intelligence and incorporated developments in the field of computers.[57] Both traditions continued to treat

[57] Cf. Haugeland (1981:1-34).

thought as computation, however, cybernetics located the "agents" of computation at the neuronal level, whereas artificial intelligence located computation to a psychological level of mental representations, creating cognitivism (63).[58] A key issue in the second wave of cybernetics would be to account for the observer both as part of the system to be studied, and as a system in himself, addressed by Heinz von Foerster' *Observing Systems* (Hayles 1999:10, 133-135). A programme of research into self-organising systems in the form of systems theory and various theories of self-organisation was developed. It culminated Maturana and Varela's (and later Luhmann's) view of the world as a set of informationally closed systems, where autopoietic organisms respond to the world in terms of their internal self-organisation (10; 131-159). In "What the Frog's Eye Tells the Frog's Brain" (1959) McCulloch, Pitts, and Jerry Lettvin demonstrated that the frog's visual system actively constructs its perceptions from the visual stimuli that it receives, rather than passively representing the input that it receives.[59] Henceforth, the structure of the organism itself and the role that it plays in processing information and the resultant unpredictability of outputs became the object of study. For these theorists it became clear that cognitive reality is inevitably relative to the organism and its organisation.[60]

By the 1980's connectionism would be revitalised by the discovery that networks with layers of hidden neurons could master learning techniques which could be interpreted in

---

Newell and Simon (1981:65) characterise AI as follows:

> Research on artificial intelligence is concerned with how symbol systems must be organised in order to behave intelligently.

And Pylyshyn (1981:68) declares:

> …artificial intelligence (AI)…attempts to understand the nature of intelligence by designing mechanisms, in the form of computational systems, which exhibit it.

[58] For an example of such an AI approach, see Marvin Minsky's (1981) "A Framework for representing Knowledge." See McDermott (1981:143-160) for a critique of the tendency to use "mnemonics like 'UNDERSTAND' or 'GOAL' to refer to programs and data structures" (144) in AI. McDermott argues that the use of borrowing concepts from human language often begs the question in AI modelling, and is misleading in that it assumes a result that has yet to be proven. He asserts (149):

> …I wish to rail against a pervasive sloppiness in our thinking: the tendency to see in natural language a natural source of problems and solutions. Many researchers tend to talk as if an internal knowledge representation ought to be closely related to the "corresponding" English sentences; and that operations on the structure should resemble human conversation or "word problems"… The "natural-language fallacy" appears here in the urge to identify the naming problem with the problem of resolving references in English-language discourse. Although the two problems must at some remote intersection meet, it seems to me to be a waste of time to focus on their similarities… The uses of reference in discourse is not the same as those of naming in internal representations.

[59] The argument is that frogs do not seem to see the stationary parts of the world around them (Lettvin, Maturana, McCulloch, and Pitts 1968:234). Their food is determined by size and movements, their sex lives are conducted through sound and touch, and fleeing from danger seems to consist of "leaping to where it is darker" (234). The authors speculate that the "nervous apparatus in the eye" is devoted to detecting certain patterns of light and their changes, which correspond to "particular relations in the visible world" (239). Simply put, their finding is that "the eye speaks to the brain in a language already highly organised and interpreted, instead of transmitting some more or less accurate copy of the distribution of light on the receptors (255).

[60] Maturana and Varela were of the opinion that there was an overemphasis on evolution and reproduction as the defining characteristics of life, and autopoietic theory was seen as a corrective for this perceived tendency (Hayles 1999:149-154).

cognitive terms.[61] It transpired that networks were, among other things, capable of learning, recognising patterns, and memorising by association, and stimulus generalisation through the process of feedback.[62] These capabilities posed particular difficulty for programmed, digital computers. Furthermore, it began to be understood that self-organisation could lead to new unexpected emergent behaviour within a system.

Making use of these new insights, Jeff Elman created neural networks called recurrent networks, where output nodes are connected to input nodes, thus presenting the output of a previous state of the system as input for various following states. As a result, the network's behaviour can be modified to conform to a given rule linking input and output patterns in that the network can adjust the "weights" between nodes in terms of input previously presented to it. In this manner the network build up a rudimentary "memory" and can, in effect, learn from experience. It transpired that these modified neural networks (recurrent networks) could be trained master the syntax-learning problem (Deacon 1997a:132ff). By means of a sort of stimulus generalisation a trained network could predict the structure of novel English sentences from the inputs presented to it. (Some of the implications of Elman's work with regard to our understanding of human cognition will be discussed in Chapter 4).

Important in terms of our discussion is the role that a conception of *evolution* came to play in the third wave of cybernetics. This approach attempted to get systems to "evolve" in new directions (Hayles 1999:222). Cognitive science was to become locked in a debate between the traditional artificial intelligence stance that thinking is computation as executed by a serial computer[63] (cognitivism)[64], and the neural network approach where thinking is computation that operates in a massively parallel fashion with interesting behaviours emerging from the system of interactions among elementary calculators (connectionism) and where the structure of systems change and adjust in accordance with the experience that it gains through its functioning (cf. DyPuy 2000:64-65; Deacon 1997a: 131ff).[65]

---

[61] See Pollack (1989) for a concise overview of the major developments in connectionism, beginning with the McCulloch-Pitts model.

[62] A neural network can a real electronic circuit made up of physical nodes and connections, but theorists more often make use of computer simulations of the behaviour of a neural network (Deacon 1997:129ff). A basic neural network with feedback capability consists of three layers of nodes – input, and output nodes and hidden layers – and the connections between nodes. Here, output nodes are connected to input nodes, thus presenting the output of a previous state of the system as input for the following one. As a result, the network's behaviour can be modified to conform to a given rule linking input and output patterns in that the network can adjust the "weights" between nodes in terms of input previously presented to it. In this manner the network builds up a rudimentary "memory" and can, in effect, learn from experience. Hence, it transpired that these modified neural networks (recurrent networks) could be trained.

[63] Dupuy (2000:64) gives the following useful definition of serial computation: "sequential calculation using symbols that have both a physical reality and a semantic, representational value".

[64] Haugeland (1981b:243) defines "cognitivism" in philosophy and psychology as "roughly the position that intelligent behaviour can (only) be explained by appeal to internal 'cognitive processes', that is, rational thought in a very broad sense."

[65] See also Clark (2001:122-126).

Despite the changes and the apparent developments that were taking place in the computational sciences, their ultimate goal of modelling the human mind in an artificial form remained elusive. None of the new developments came close to imitating the complexity and the extent of the computational abilities of the human mind. Whereas theorists were enthusiastic about the possibility of creating artificial intelligence in the 1950's through to the 1980's[66], by then the possibility had already come to seem remote for some.[67] The entire research programme came under scrutiny. Clark (2001:124) notes that recent work in cognitive science finds itself in a "process of inner-symbol flight" moving away from the classical approaches to modelling rationality as, what he terms, "linguaform reason". He argues that theorists in this field have begun to doubt the centrality exemplars of thought and reason in the discipline, the "Turing-Fodor vision" (of the mind as a symbol processor) and even the "connectionist" vision (126). He identifies a trend to re-invent rationality as "an active, distributed, environment-exploiting achievement". Hence, a scepticism has developed with regard to old notions such as representation and the role that symbols and internal representations play an important role in human intelligence. Clark is sympathetic to such scepticism as directed at traditional computational conceptions of reason, however, he warns against taking such scepticism too far and denying that human-reason-governed rational behaviour takes place in the virtual absence of any form of internal representation (129). (We will come back to this position in the following chapter.)

Clark thus illuminates a trend towards an "embodied embedded" approach in the cognitive sciences (2001:122). As we shall see, part of this trend is the "rediscovery" of Darwin and his theory of natural selection by theorists of mind. It seems that the computational approach has come up against an insurmountable hurdle: the human mind cannot properly be conceived of as a Turing machine. It seems that thought constitutes more than computation. What remains critically absent in traditional computational approaches to mind are concepts of the environment in which the mind operates (receives its input) and a notion of the purpose of mind (what its input is used for).[68] It seems highly improbable that the human mind can successfully be conceived of as an abstract, disembodied logic-machine.

---

[66] See Haugeland (1981); Newell and Simon (1981).

[67] See Minsky (1981) where he criticises the "logistic approach" in simulating world-knowledge, arguing that it had been tried failed "since Aristotle" with little success (123).

[68] See Dreyfus (1981) where he presents a similar criticism of the AI assumption that the "know-how that enables human beings constantly to sense what specific situation they are in" (198) is the kind of knowledge that can be represented in a "knowledge representation language" of some sort. He argues that "there is no reason to suppose that moods, mattering, and embodied skills can be captured in any formal web of belief." He argues that most AI workers and cognitive scientists are committed to the belief that such "noncognitive" aspects of the mind can simply be ignored. Dreyfus's own position is that "knowledge" of human interests practises need not be represented at all (202) and furthermore, that it may be possible that human beings make use of nonformal representations, "more like images, by means of which I explore what I *am*, not what I *know*" (203).

See also Putnam (1981:205-219).

Dreyfus (1981:203-204):

> Looking back over the last ten years of AI research we might say that the basic point which has emerged is that *since intelligence must be situated it cannot be separated from*

3.4 INTENTIONALITY: WHERE PHILOSOPHY OF MIND GOT IT WRONG

It is essential to our argument to consider the role that philosophy played (and plays) in theories of cognitive science. Dupuy (2000:90) for example argues that coherence was given to the various research programmes associated with cognitive science by the philosophical work which was done in connection with them, i.e. "cognitive philosophy":[69]

> It is philosophers (and not psychologists, or linguists, or neurobiologists, or computer scientists) who have reflected upon and systematised the basic attitudes shared by various workers in the various disciplines that make up the field.

He considers this situation to be ironic, given that the cognitive sciences purport to have claimed the ancient philosophical questions of mind for science. Dupuy's position is that the "science" that connects the various disciplines that make up the field (neuroscience, artificial intelligence, cognitive psychology, and linguistics) is in reality philosophy (2000:91). More accurately it is a particular strain of philosophy, namely philosophy of mind. An important result of the dominance of philosophy of mind in the cognitive sciences is its exclusion of "rival philosophies" such as philosophies of consciousness, phenomenology, and existentialism, as well as rival psychologies and sciences.

Dupuy describes cognitive philosophy/philosophy of mind as "the most active and flourishing branch of analytical philosophy" (2000:91); a statement that he describes as "paradoxical". As we have seen, analytical philosophy was fundamentally characterised by its break with "psychologism". Dupuy suggests that another great tradition of contemporary philosophy, phenomenology, came into existence by means of the same break (see also Ricoeur 1979). For him, this parallel is not accidental, given that philosophy of mind and phenomenology share a fundamental concept. Dupuy argues that the formalism of Hilbert, the logicism of Frege, Russell, and Carnap, or of Wittgenstein and the Oxford school of ordinary language philosophy, all share the refusal to base philosophy on psychology, and the

---

the rest of human life. The persistent denial of this seemingly obvious point cannot, however, be laid at the door of AI. It starts with Plato's separation of the intellect or rational soul from the body with its skills emotions and appetites. Aristotle continued this unlikely dichotomy when he separated the theoretical from the practical, and defined man as a rational animal – as if one can separate man's rationality from his animal needs and desires. If one thinks of the importance of the sensory-motor skills in the development of our ability to recognize and cope with objects, or of the role of needs and in structuring all social situations, or finally of the whole cultural background of human self-interpretation involved in our simply knowing how to pick out and use chairs, the idea that we can simply ignore this know-how while formalising our intellectual understanding as a complex system of facts and rules is highly implausible.

[69] In a similar vein, Flanagan (1991:181-185) points out that Kant laid both the substantive and methodological foundations for modern cognitive science. He considered the mind to be active in the construction of language, and he prescribed the logical structure of the canonical method for "inferring hidden mental processes", namely transcendental deduction.

insistence on the priority of the study of language in their approach. Dupuy goes on to wonder at the "standard account" which holds that this "linguistic turn" at the beginning of the twentieth century was to be followed by a "cognitive turn". It seems paradoxical that the rejection of psychologism would eventually result in cognitive philosophy, or philosophical psychology.

Reasons given for this reversal in subject matter generally fall into two categories: Firstly, there is the type that relates to the "internal evolution of philosophy of language" (Dupuy 2000:92). In this view, theories such as Chomskyan generative grammar indicate that as speakers and interlocutors we can make certain assumptions about our cognitive capacities. Without appeal to these assumed cognitive capacities there would be no way in which we can explain the nature and properties of language. The second type of reason is rooted in the work done in cognitive science itself. It is argued that research in this field has made it possible to carry out Quine's programme of the "naturalisation of epistemology". Hence, traditionally philosophical epistemological questions can be answered by means of the empirical sciences.

Dupuy argues that these reasons "explain nothing" (2000:92):

> Philosophy moved to cut itself off from psychology, as from the natural sciences as a whole, in order to conduct its own inquiry into the objective validity of knowledge and the very legitimacy of a science of nature … for reasons of principle that were perfectly independent of the progress of the empirical sciences.

He argues that this stance was the source of both Frege's and Husserl's "anti-psychologism". Dupuy holds that the cognitive turn in analytical philosophy makes much more sense if one considers that the change brought about by it is much less radical than it may seem at first. Firstly, he argues, the cognitive branch of analytical philosophy still retains some of its Kantian heritage and keeps its distance from psychologism. Secondly, it has not managed to free itself from the priority initially given to language in analytic philosophy (cf. Dreyfus 1981:161-204).[70] Thus, whereas artificial intelligence presents itself as a science that

---

[70] McDermott makes a similar point in his critique of AI, much of which, he argues, makes the unwarranted assumption that internal knowledge representation in AI systems (and humans) ought to closely relate to the "corresponding" English sentences, and that their operations should resemble human conversation (1981:149). He states:

> It is hard to say where they have gone wronger [sic], in underestimating language or overestimating computer programs…(151) The obsession with natural language seems to have caused the feeling that the human use of language is a royal road to the cognitive psyche. I find this analogous to preoccupation with imagery of the study of vision. Most AI researchers react with amusement to proposals to explain vision in terms of stored images, reducing the physical to the mind's eye. But many of the same people notice themselves talking to themselves in English, and conclude that English is close to the language of thought. Clearly, there must be some other notation, different in principle from natural language… As of now, we have no idea at all why people experience their thoughts the way they do, in pictures and words…(153) Linguists, I think, suffered from this self-misdirection for years. The standard experimental tool of modern linguistics is the eliciting of judgements of grammaticality from native speakers. Although anyone can learn

contributes to the empirical understanding of nature, it can be argued that the very concern for universality in artificial intelligence makes it a philosophy, and even a transcendental philosophy. Artificial intelligence is universal in that it aims to account for the formal conditions of cognition that is common to all systems that are capable of cognition apart from its particular manifestation in organisms or machines.[71] In terms of his contention that the "cognitive turn" in analytical philosophy remains profoundly influenced by the philosophy of language, Dupuy (2000:94-99) argues that the question of representation plays a prominent role. In this regard, he highlights Daniel Andler's thesis that contemporary connectionism can link up with the cognitivist mainstream by borrowing the notion of representation from it. In this view, McCulloch's model fails to be philosophically sound because it fails to show how the signal transmitted by a neuron is able to represent a proposition about the world; or, in Dupuy's (95) words: to "show how the coding that allows an equivalence to be established between some thought content and a computation effected by the network can be assigned a meaning in terms of representation."

Andler's argument takes two things for granted: i) that knowing subjects have representations in the first place, and ii) that the form of such representations corresponds to representation as conceived of in cognitive philosophy, i.e. a "mental state" that has a content that concerns the world. In cognitive philosophy the representative capacity of mental states are referred to as "intentional" states (see Searle 1981, for example). If mental states are intentional they are about objects that are external to the mind. The question arises as to the content of such an intentional mental state. Dupuy (2000:95) argues that cognitive philosophy posits that the object of a representation is the thing-in-itself (a noumenon in Kantian terms), as opposed to the object in terms of the system, or a simulation of the object (phenomenon).[72] In these terms, what does the content of such a mental state consist of?

Dupuy's (2000:95-96) position is that the answer that philosophy of mind gives to this question continues to make it a philosophy of language in that it invariably gives a linguistic answer at two levels: Firstly, he argues that Russell's "propositional attitudes" – which are supposed to connect a psychological attitude of the type "to believe that", "to desire that", "to

---

> how to make such judgements fairly quickly, it is plainly not a skill that has anything to do with speaking English. The real parser in your head is not supposed to report on its inputs' degree of grammaticality; indeed, normally it doesn't "report" at all in a way accessible to verbalisation. It just tries to aid understanding of what it hears as best it can. So the grammaticality judgement task is completely artificial. It doesn't correspond to something people normally do (153-154).

Also see Dreyfus (1981:161-204) for similar criticism.

[71] In Dupuy's (2000:93) words:

> The purpose of this inquiry was ultimately to discover the a priori, necessary and sufficient conditions that both made language possible and grounded the objectivity, which is to say the universality, of this knowledge.

In Kantian terms, the transcendental subject was replaced by the "physical symbol system" and synthetic a priori by the Turing machine.

[72] Essentially his objection is similar to those of Hayles and Searle, in that the argument is that a system's "representation" of its input (information) cannot be abstracted from the configuration of that particular system.

fear that", etc. with a proposition about the world – have long captured the attention of cognitive science. In terms of current philosophy of mind, in its functionalist, computational, and representational versions as defended by Fodor[73] and Pylyshyn, it is argued that such propositions are expressed in the sentences of "the language of thought". The language of thought is a private language, the symbols of which are somehow inscribed in the material substrate of the brain. Secondly, Dupuy holds that even those philosophers who do not share the hypothesis of a language of thought, generally accept that the criterion of mental "intentionality" is linguistic on a second level, namely at the level of the sentences of the public languages that we use to attribute mental states to others. Such sentences are *intensional* in that they violate the rules of logical extensionality, namely the terms of existential generalisation and the substitutability of terms with the same reference.

Dupuy's (2000:97) position is that many of the difficulties that philosophy of mind encounters today can be attributed to its initial commitment to the linguistic interpretation of intentionality. He argues that many of its practitioners accept this interpretation of intentionality while trying to "naturalise" it, by which he means that they try to formulate an analysis of intentionality founded on the laws of physics. Whereas psychology would take it for granted that the contents of mental states are causally relevant for explaining our behaviour, philosophy of mind seeks for a "physicalist" explanation for this causality. Dupuy (2000:97) argues that this requirement becomes an insurmountable obstacle:

> …philosophy of mind had convinced itself that the semantic content of a mental state, as described by its conditions of truth and reference, depends on the entire physical and social environment of the subject; but if this content is supposed to have causal power in the physicist's sense, it can only be conceived in terms of the intrinsic properties of the mental state. It therefore appeared that the theory of knowledge could be naturalised only at the cost of depriving mental properties and mental states of all causal efficacy insofar as they are mental – thus making them pure "epiphenomena".

Hence, such materialist theories of mind seem to be caught up in a permanent dilemma: trying to reduce mental concepts conceived of in linguistic terms to physicalist terms.

For our purposes, Dupuy presents an interesting analysis of the origin of the concept of intentionality in philosophy of mind. He notes (99) that the concept plays a central role in another prominent philosophical movement of the twentieth century, that of phenomenology. Husserl attributes the property of being "intentional" to mental acts, where "intentionality" is the essence of the concept of consciousness. Both the "intentionality" of philosophy of mind, and Husserl's "intentionality" derive from the philosophical psychology of Franz Brentano.

[73] See, for example Fodor (1990:282-299).

Whereas Husserl studied under Brentano from 1884 to 1886, Roderick Chisholm (e.g. Chisholm 2002 [1957]) was instrumental in introducing Brentano's thought, and his conception of "intentionality" to America. The question arises: Which one of the two theorists, Husserl or Chisholm, accurately portrayed Brentano's original concept?

Dupuy (2000:100) bases his analysis on the following quotation from Brentano's (1874) *Psychology from an Empirical Standpoint*:

> Every mental phenomenon is characterised by what the mediaeval schoolmen called the intentional (or mental) inexistence of an object, and by what we, not without ambiguity, call the relation to a content, the direction to an object (by which reality is not to be understood) or an immanent objectivity (Brentano 2002 [1874]:481).

He interprets this passage as follows: "Intentional", here, is a scholastic term used in the sense of "mental" as opposed to "real". Furthermore, "form" is united with matter in the "real" object, outside the mind, which means that the "intentional" object is only present in its form. "Inexistence" (*Inexistenz*), in this instance, comes from the Latin *inesse* meaning "to be inside of" (*not* "nonexistent"). Hence, Dupuy interprets the passage as holding that "the object toward which the mind tends (i.e. its intention) is located *inside* the mind". In this understanding, Brentano's argument is that mental life is an activity, the activity of "presentation". The dynamic action of "presentation" (more accurately "presenting") has a content or an object, which is the very thing that the mind perceives, such as a sound, or a colour, etc, which Brentano calls "physical phenomena". Physical phenomena are also part of the data of consciousness (cf. Brentano 2002 [1874]:479ff).

In Dupuy's (2000:100) reading, Brentano's argument is that mental activity is intrinsically conscious of itself: "When we think, we have an immediate perception of the fact that we are thinking, and this perception of thinking activity is simultaneously a perception of the object of thought." Brentano does not take such "internal perception" to be observation; if it were, we would be left with an infinite regression of mental activities. Thus, thought as activity is a unique act of apprehension that links it both to itself and its intentional object. We *perceive* mental phenomena, but we do not *observe* them and Brentano concludes that there is a "special connection" between the object of inner presentation and the presentation: "In the same mental phenomenon in which the sound is present to our mind we simultaneously apprehend the mental phenomenon itself" (101).

Dupuy (2000:102) goes on to imagine what would happen if one were to misconstrue Brentano's argument by getting the meaning of his key terms wrong. In such a misreading one might interpret "inexistence" (*Inexistenz*) as meaning "non-existent". In such a misreading, the object towards which representation tends would be a "nonexistent physical object". Here, the object towards which representation tends may or may not exist outside of one's mind. Hence, intentionality is no longer a mental act that presents an object to the mind,

it becomes a mental state endowed with content, which is related to an object whose existence is not guaranteed by the fact that the mental state exists. In this context, such content is *intensional* and hence, linguistic.

It is thus Dupuy's contention that Chisholm, in his interpretation and translation of Brentano, made precisely this error, which was passed on to an entire generation of analytic philosophers. As an example he lists Quine's appropriation of Brentano's thesis as developed by Chisholm in *Word and Object* (1960). Dupuy (2000:102) concludes:

> "Brentano's thesis" is in fact none other than Chisholm's thesis. It asserts that mental states (and mental states alone) are endowed with the property of intentionality, understood as a linguistic relation to objects or states of affairs of the external world – external to the mind. Intentional expressions are irreducible to the terms that we use to describe "physical" phenomena.

Dupuy's (103-107) argument then is that if there exists a possibility of "naturalising" "intentionality", it would lay with the conception of intentionality as conceived of by Brentano and Husserl.[74] And the way that he suggests that it might be done is through the various schools that arose in the 1970's and 1980's that started to treat neural networks as dynamic systems, rather than "logic machines" in the tradition of McCulloch and Pitts. Under these he includes theories of self-organisation, such as that of Stuart Kauffman[75], for example, Maturana and Varela's school of autopoiesis[76], and the Attractor Neural Network school founded by John Hopfield[77] at Caltech. What these approaches have in common is that they

---

[74] A similar argument is made by Dreyfus (1981:180ff) who draws explicitly on Husserl's phenomenological analysis. He argues (180):

> Instead of modelling intelligence as a passive receiving of context-free facts into a structure of already stored data, Husserl thinks of intelligence as a context-determined, goal-directed activity – as a search for anticipated facts. For him the *noema*, or mental representation of any type of object, provides a context or "inner horizon" of expectations or "predelineations" for structuring the incoming data.

However, Dreyfus (182) cautions that Husserl's project ran into trouble in that, in trying to spell out the components of the noema of everyday objects, he found that he had to include more and more of the subject's total knowledge of the world.

[75] See Kauffman (1995) where he argues that complex systems tend towards self-organisation, which he believes may provide an additional organising principle to the natural world, over and above that of evolution by means of natural selection.

[76] Varela began to work in the new field that came to be known as Artificial Life (AL), and the autopoietic notion that systems are operationally closed made the transition to its successor (Hayles 1999:222). AL attempts to i) create artificial biological life through building components of unicellular organisms in test tubes, or ii) constructing robots, or iii) creating computer programmes that simulate emergent or evolutionary processes (225). In terms of the project of creating computer programmes that simulate life processes (such as in the case of Thompson's Tierra), the idea is that a few simple local rules coupled with highly recursive structures would allow complexity to emerge spontaneously (225). A striking aspect of this research programme, is its propensity to equate programmes that replicate inside a computer to living organisms and interpreting the process of replication in terms of Darwinian natural selection (see Hayles's discussion 1999:224-231).

[77] Hopfield introduced feedback to his networks, which meant that the networks are influenced by current and previous inputs. In this manner, memory is introduced into the system (Picton 1994:66-68)

do not treat neural networks as simple tranducers converting input messages into output messages, but treat them as autonomous in the sense that they are capable of *eigenbehaviours* ("self-behaviours"), which are generated by the network itself through feedback. External behaviours acquire meaning in terms of the network's activity. Any meaning that the network attributes to external entities thus cannot reflect "external" objectivity. Dupuy argues that such a network is intentional in Brentano's and Husserl's sense.

Against the background of his analysis of both the logical origins of the mechanised approach to mind and his analysis of the conceptual commitment made to an understanding of intentionality that may be interpreted in a drastically different way, Dupuy concludes that current philosophy of mind, which inspires mainstream research in cognitive science, risks losing touch with the world in which cognition takes place (2000:154-155):

> The abstraction of the forced and artificial thought experiments that almost exclusively provide it with material for analysis (along with the properties of ordinary language) strongly resembles that of the esoteric games to which the second generation of cyberneticians… devoted themselves. In both cases, philosophical reflection is fed more by science fiction than by science.

It aims to "physicalise" or "naturalise" intentionality; thus cognitive science considers itself to be a branch of physics. However, Dupuy argues that the physics that philosophers of mind have placed on a pedestal is a "very French image of physics as *reines des sciences*", or as the "queen of the sciences". In this "French view" physics is viewed as an *a priori* science. Physics in this understanding becomes a philosophical and abstract discipline instead of a field in which empirical research takes place to verify hypotheses. To quote Dupuy (2000:155):

> The physics of the laboratories bears more similarity to a program of perpetual tinkering than to a logic of concepts: faced with new problems to be solved, it invents new operations, new methods of inquiry, new ways of manipulating objects, without always worrying about the logical coherence of the new ensemble procedures and results… When philosophy of mind … stubbornly persists in conceiving of intentionality as a relation between a mental state and its object (or content), it cuts itself off from the new physics and the attempt to unite connectionist research with the physics of disordered systems and the thermodynamics of irreversible processes. Much of what presently goes on in the name of "naturalising intentionality" is missing this crucial opportunity.

Such sentiments are not idiosyncratic and have been voiced by a number of theorists, especially in recent years. There has been a marked trend in revisiting Darwinian theory in

relation to both mind and language, the two topics which seem to be hopelessly intertwined and virtually inseparable no matter the theoretical approach by means of which they are studied. One aspect of the Darwinian revival has seen theorists exploit the "crucial opportunity" that Dupuy speaks of, appropriating Darwinian theory for connectionist research into the dynamics of various types of neural networks.[78] Furthermore, the "physics of disordered systems and the thermodynamics of irreversible processes" have been explicitly incorporated into Darwinian evolutionary theory, with claims that such theories can supplement the theory of evolution through natural selection and strengthen its explanatory power, or, in some instances, replace it.[79] In terms of certain preoccupations of twentieth century philosophy, such as meaning and intentionality, contextualising mind and language in evolutionary terms may provide a way to distinguish between the "pseudo-understanding" or "derived intentionality" of an artefact, and "real" intentionality and understanding, without recourse to essentialism.

4. Evolution Revisited

We have seen that, despite the increasing appeal that evolutionary theory (in various forms) held for the biological sciences in the twentieth century, two biological human traits were not systematically subjected to evolutionary theory for most of that time. In the previous section we saw that one of the main reasons for this state of affairs with regard to mind was the assumption that reason can be likened to "computing with integers". The research programme that arose around this assumption dominated the science of mind in the twentieth century. We also saw that whereas interest in the origin and evolution of language was already extensive by the time of the publication of *The Origin* the lack of credible theories in terms of which to couch theories on the topic led to it being dropped from mainstream science for the good part of a century (Pinker and Bloom 1990:738; Christiansen and Kirby 2003a:300). Christiansen and Kirby (2003b:2-3) hold that interest in the evolution of language only began to be rekindled in 1975 with a conference entitled "Origins and Evolution of Language and Speech", sponsored by the New York Academy of sciences. However, many regard the landmark paper "Natural Language and Natural Selection" (1990) by Stephen Pinker and Paul Bloom to have been the catalyst that ensured that the question again entered the scientific mainstream.

In "Natural Language and Natural Selection", Pinker and Bloom (1990) argue that human language ability is a complex biological adaptation that evolved by means of natural selection. While their contention may seem self-evident in the light of the success of the theory of natural selection when applied to other biological phenomena, we have seen that various factors contributed to making this position quite contentious in some circles. Firstly,

[78] See Dennett (1991) for example.

[79] See for example Von Bertalanffy (1973); Prigogine and Strengers (1994); Kauffman (1993); Bak (1997); Holland (1998).

there are objections stemming from a strong tradition of dualism within Western philosophy, whose advocates hold that the human mind and everything associated with it (including language ability) are of a different order from other biological phenomena. Hence, despite the success of the theory of natural selection, it cannot be applied to the evolution of language and mind in any meaningful way. We have also seen that Chomsky (2000b:163; 2004:150) for example, is sceptical of the ability of the theory of natural selection to account for the evolution of the human mind, believing that other principles may be discovered that are just as relevant to the evolution of complex structures as the principle of natural selection. Newmeyer (2003:58-76) notes that this position is relatively prevalent among linguists, despite the evolutionary stance being enthusiastically adopted in various related fields.[80] Other objections that Pinker and Bloom (1990:707) note are: language may be a by-product of selection for other, adaptive abilities; biological specialisation for grammar is incompatible with Darwinian theory, because it does not show genetic variation, could not exist in an "intermediate form", does not confer a selective advantage, and would require more evolutionary time and genomic space than is available.[81] In fact, they point out that the "nonselectionist" view of language had become the consensus view in many theoretical circles (1990:709).

Pinker and Bloom argue that such objections generally rely on inaccurate assumptions about biology or language or both (1990:707) (also see Dennett 1995). They hold that evolutionary theory offers clear criteria for when a trait can be attributed to natural selection, namely that it exhibits complex design for a specific function and the absence of an alternative process capable of explaining the complexity. Language seems to meet these criteria, although there are those who would argue that it does not, either because it is not clear what the function of language is (e.g. Gould) or, as we have seen with Chomsky, because there may be other as-yet-unknown processes that are capable of explaining the complexity of language.[82]

Pinker and Bloom hold that language does indeed meet these criteria in that "grammar is a complex mechanism tailored to the transmission of propositional structures through a serial interface". Hence, their argument is that specialisation for grammar evolved by the process of natural selection. They approach language as Darwin approached other complex organs: They note that it is ubiquitous in human societies, and argue that it differs from cultural phenomena such as agriculture which was invented by some groups and spread to others. They note that all languages are "complex computational systems" that use the

---

[80] See also Bickerton (2003:77-93).

[81] It is interesting to note that many of these objections mirror objections to the possibility that other adaptations, such as the eye, for example, could have evolved by means of natural selection. Furthermore, many of these objections (such as those pertaining to the evolution of the eye) were already foreseen and addressed by Darwin (cf. 1985:217-224).

[82] It is interesting to note that Pinker and Bloom (1990:709) point out that the reverse strategy is adopted by a range of theorists in their critique of Chomsky's theories on language. These theorists argue that the apparent discrepancy between his theories and that of natural selection might indicate that it is Chomsky's position, and not that of natural selection, that needs to be revised.

same basic kinds of rules and representations (1990:708). Nearly all human beings are proficient in language use, and, like Chomsky, they argue that although there is a poverty of stimulus for children learning language, it need not inhibit their ability to acquire grammar. Furthermore, language functions can be influenced by injury or disease, and some language disorders can be genetically transmitted. Various aspects of human physiology, such as brain structure, the structure of the vocal tract, and the auditory system seem to be adapted to facilitating the production and perception of language (708). Pinker and Bloom take all of these facts to indicate that language is a biological rather than a cultural entity, and should be approached as such. Seeing that the theory of natural selection is the only successful account of the origin and development of complex biological structures, they contend that language should be studied in terms of natural selection. They go on to defend their position against possible objections, such as those of Chomsky and Gould. It should be noted that Pinker and Bloom are very much taken with Chomsky's theory of generative grammar, and try to incorporate in into the theory of the natural selection of language (709).

In their objection to this view, Gould and Lewontin (1979:147) argue that it is naïve to view all biological traits as adaptations by means of natural selection for specific functions. They argue that some traits may very well be "spandrels" – biological traits that are by-products of other adaptations that may have been appropriated for the function which it currently fulfils. In other words, not all biological traits are necessarily immediate adaptations to local conditions and may be explained in terms of other laws or mechanisms, such as genetic drift, laws of growth and form, induction of form by environmental forces, etc. Gould's position has been taken to imply that nonadaptationist processes are an alternative to Darwin (Pinker and Bloom 1990:711; Dennett 1995). Pinker and Bloom (1990:711) argue that this criticism of adaptationism does not take into account that "natural selection is the only scientific explanation of adaptive complexity". Adaptive complexity is defined as describing a system that is composed of many interacting parts where the structure and arrangement of the parts suggest a design to fulfil some function, such as the structure of the eye. We have seen that the Natural Theologians used the eye as an example of a structure that could not have developed by means of natural laws and hence pointed to the existence of a divine Creator. Darwin (1985:217-220) gave an extensive (and convincing) explanation of how the structure of the eye could have come about through the process of natural selection. Pinker and Bloom (711) extend the argument and contend that a structure as intricate and complex as the eye is an extremely low-probability arrangement of matter. It is very difficult to imagine how a nonadaptationist process such as genetic drift or a general law of growth could result (in more than one instance) in an arrangement of matter that would result in such a complex structure, which would then be fixated within a population. The advantage of Darwin's theory is it that allows for seemingly teleological adaptation, in the absence of some form of foresight or a designer. As Pinker and Bloom argue with regard to the evolution of the eye, natural selection "is the only physical process capable of creating a functioning eye, because it is the only physical process in which the criterion of being good at seeing can play a causal role. As

such it is the only process that can lead organisms along the path in the astronomically vast space of possible bodies leading from a body with no eye to a body with a functioning eye" (711).

A possible problem with this argument is the need for clearer definitions of "function" and "design". Notably, Pinker and Bloom (1990:712) argue that legitimate functions would invariably be linked to the overall function of survival and reproduction. Epiphenomenal spandrels are not involved in any function or behaviour of the species and would necessarily not be attributed to adaptation through natural selection. They point out that those spandrels which are put to use, have been modified for a specific purpose, which means that selection played a role in the process of modification. They argue that spandrels, exaptations, laws of growth, and so forth can explain the basic plans, parts, and materials that natural selection employs. There is no reason why a by-product of a particular adaptation could not be appropriated for another purpose.[83] Such a spandrel would in all probability need some modification. According to Pinker and Bloom (712) the best examples of structures that are produced by nonadaptationist processes usually have one-part or repetitive shapes or processes that correspond to simple physical or geometric laws, such as hexagonal honeycombs or spiral markings. The evolution of a useful feature without natural selection may have occurred in the case of unmodified spandrels, but the argument is that such spandrels would most probably only be capable of fulfilling simple functions with regard to the survival and reproduction of the species.

Specific adaptationist proposals should be testable in principle and practice. Given the criteria of complex design to enable reproductive success, one would determine whether a given adaptation is useful in terms of the ecological conditions in which it functions, and one would then have to measure the relative reproductive success of the individuals that posses them. Bloom and Pinker go so far as to argue that it is in fact *non*adaptationist accounts that are in danger of vacuity in that many such accounts merely suggest that there may be some hitherto-unknown law of physics or growth; a statement which is in itself unfalsifiable (Pinker and Bloom 1990:713):

> Specific adaptationist proposals may be unmotivated, but they are within the realm of biological and physical understanding, and often the problem is simply that we lack the evidence to determine which account within a set of alternative adaptive explanations is the correct one.

Hence, Pinker and Bloom (1990:715) conclude that objections, such as those by Gould and Chomsky, merely shift the emphasis within an orthodox neo-Darwinian theory[84]

---

[83] Dennett (1995:390) makes a similar argument.

[84] Neo-Darwinism refers to the synthesis between genetic theory as a theory of inheritance and natural selection as the mechanism that drives evolution, and is sometimes referred to as *synthetic theory* (Patterson 1998:234).

and cannot be taken to be proposing radical alternatives to natural selection as the driving forces behind the evolution of adaptations, including language. Nor can language be considered to be the result of a nonselectionist process, given that it exhibits evidence of complex design geared towards fulfilling a reproductively significant function. Of course, both of these assumed attributes need to be substantiated through further argument. Does language structure show signs of design and does language fulfil a significant function with regard to survival and reproduction? These are by no means straightforward questions, and many theorists who agree with Pinker and Bloom that language is an adaptive function, differ from them in terms of the reasons that they give for this belief.

4.1 The Argument From Design

As may be expected, Pinker and Bloom argue that language shows signs of design, namely the "design for the communication of propositional structures over a serial channel" (1990:715). As a point of departure, they assume that human knowledge and reasoning is couched in a language of thought (mentalese). The propositions of mentalese are relational structures "whose symbols pertain to people, objects, and events, the categories they belong to, their distribution in space and time, and their causal relations to one another" (715). Other people are understood as behaving in terms of their beliefs and desires (thus through the intentional stance), which can be understood as relations between the individual and the propositions that represent the content of that belief or desire. They conclude that certain kinds of contents are necessary and worthy of being communicated among humans, such as: the ability to refer to individuals and classes, to distinguish among ontological categories, the ability to talk about events and states, to distinguish between participants in such events according to their roles, the ability to talk about our intentional states and those of others, the ability to distinguish in truth values and modality, and to indicate the time of a state or event, among others. The vocal-auditory channel is a serial interface, and cannot make use of various typographical devices for clarification; its tools consist of distinguishable symbols and their structure (concatenation). Further constraints on this medium of communication are: limited short-term memory of its users, it needs to be utilised rapidly, and it needs to operate according to a code that can be shared by an entire community of participants (716).

Pinker and Bloom do not consider the view that language is a complex system to be contentious at all (1990:716):

> The fact that language is a complex system of many parts, each tailored to mapping a characteristic kind of semantic or pragmatic function onto a characteristic kind of symbol sequence, is so obvious in linguistic practise that it is not usually seen as worth mentioning.

It is impossible not to agree with them; no matter what theory of language structure one adheres to – language is a complex structure that fulfils a complex function. Human language can communicate incredibly intricate messages, which it communicates through a serial channel. For language to fulfil such complex functions without having been "designed" through natural selection for doing so seems highly unlikely. Pinker and Bloom (718) argue that the "special and unusual" properties needed for mapping complex propositional structures onto a serial channel are to be found in grammar. Furthermore, they hold that alternative explanations for the evolution of language – such as representing internal knowledge for internal computation and reasoning – are easily refuted when one considers the design of language. They argue that natural languages are hopeless for internal representation, in that they are needlessly serial, exhibit ambiguity, are complicated by alternations that only make sense in terms of discourse, and cluttered with devices that do not contribute to reasoning (but that facilitate discourse). Language seems to be designed for communication, with mechanisms that presuppose the existence of a listener and hi-fidelity distribution throughout a particular community. The theory relies, of course, on the assumption of the underlying unity of diverse natural languages.

If language were a biological adaptation within a relatively uniform species, the question may arise as to why there are diverse languages at all. Pinker and Bloom (1990:721) speculate firstly that learning mechanisms which allow for new, culture-specific cultural innovations that need to be named in language are necessary and this causes languages to diverge. Secondly, they hypothesise that it is difficult to evolve the huge, innate code that would be necessary for standardising language across all users over time. It is unlikely that the genome would be able to incorporate such an extensive code and maintain it in the face of perturbations from sexual recombination. The argument is that the *mechanisms* for acquiring language are genetic; particular language information is gained from particular communities.

It should be kept in mind that, like with all other adaptations, language would not be *the* perfect adaptation for the function to which it is adapted (in this case the communication of propositional structures over a serial channel). Inevitably, natural selection works within various constraints, such as the need for tradeoffs among adaptive goals. Pinker and Bloom (1990:722) point out that language adaptation is inevitably caught up in a conflict of interests between speakers and hearers, for example. Speakers would want to minimise the effort of articulation and would favour brevity, while hearers would want to minimise the effort associated with comprehension and would favour explicitness and clarity. There would also be conflicts between speakers and learners, where the learner would have a relatively limited vocabulary and comprehension rate which would inevitably influence the form the mature language would take. Language viewed from any one of these perspectives, to the exclusion of the others, would appear somewhat arbitrary. Hence, even if it is possible to conceive of natural language evolving in a different way, it does not follow that natural language as it stands is not an adaptation.

A common objection to the proposition that language evolved by means of natural selection is that language has arbitrary features that do not contribute to communication in any obvious way. Pinker and Bloom go on to argue that the arbitrariness of grammar is, in principle, part of the adaptive solution to effective communication. A coding protocol for any communicative system can be arbitrary as long as it is shared by the users of that system (1990:723). They suggest that arbitrariness is built into "the language acquisition device" at two levels: it forms only those rules that fall within the (arbitrary?) set that is defined by universal grammar, and it tries to choose rules within that set that match those used by the specific community within which the language acquisition device operates.[85] The benefits of such a system become apparent when one thinks of the enormity of the task of determining the rules of a particular language through learning and deduction. The argument is that arbitrariness is a necessary feature of natural language and that such arbitrariness is incredibly important when it comes to understanding the role of the communicative function of language in its acquisition and evolution. The conclusion that Pinker and Bloom (1990:725) draw from this aspect of language is that the processes of language evolution (in the species as a whole) and language acquisition (in the individual) must differ. Functionalist theories of language evolution can be true, while functionalist theories of language acquisition can be false (225):

> Evolution has had a wide variety of equivalent communicative standards to choose from; there is no reason for it to have favoured the class of language that includes Apache and Yiddish, but not Old High Martian or Early Vulcan. But this flexibility has been used up by the time the child is born; the species and the language community have already made their choices. The child cannot learn just any useful communicative system; nor can he or she learn just any natural language. He or she is stuck with having to learn the particular kind of language the species eventually converged upon and the particular variety the community has chosen. Whatever rationales may have influenced these choices are buried in history and cannot be recapitulated in development.

Hence, because language as a tool for communication needs to be standardised in terms of many often conflicting requirements, it makes more sense for evolution to build a language acquisition device that incorporates the code of a particular language that surrounds an organism, than one that invents a language that is useful from the point of view of the child. In terms of the poverty of stimulus principle, it is probable that many grammatical constraints and principles are hardwired into the "device", in whatever way it is conceived. As

[85] It should be kept in mind that Pinker and Bloom are very much under the influence of Chomsky's theories and hence retain concepts such as the language organ and aspects of his generative grammar, although there is no reason to retain these assumptions in an evolutionary approach to mind and language, as will become apparent in Chapter 5.

such, it would make more sense for the mind to have evolved to support language, than for language to be a "spandrel" piggybacking on a system that was designed for other functions, as Gould would have it (e.g. 1979). Hence, Pinker and Bloom (1990:728) conclude that "the most likely explanation for the complex structure of the language faculty is that it is a design imposed on neural circuitry as a response to evolutionary processes."

4.2 The Argument from the Adaptive Function of Language

As far as the criterion of the adaptive benefit of language goes, the question is not any less complicated than that of proving that language is indeed a designed entity. Pinker and Bloom (1990:728) argue that it needs to be shown that language has a beneficial function and that there was a series of steps from no language to language as we know it. Each intermediate step needs to have been small enough to have been produced by a random mutation, and in itself must have conferred a reproductive advantage to its owners, in order to have spread through the population. All of these questions are still being debated extensively within the literature, especially in the light of the fact that it is extremely difficult to gain data pertaining to the evolutionary history of language.

One objection to the possibility of the evolution of language is that the intermediate links would not have been viable communication systems (Pinker and Bloom 1990:730). The argument here is, given that a genetic mutation initially involves a single individual, how would such a mutation benefit communication with the other, non-mutated individuals in the population? One, rather unconvincing, possibility that Pinker and Bloom (1990:730) raise is that such a mutation would in all probability be shared by individuals who are genetically related, who would then be able to communicate. But a more convincing, and more general, possibility is their argument that comprehension and production abilities need not be perfectly synchronised throughout a community for them to be able to communicate relatively effectively. An intermediate mutation may still be beneficial to its possessor, without completely inhibiting communication with non-mutated individuals. Furthermore, the cognitive effort required to meet the challenges presented by the mutation within the population would provide pressure for the evolution of neural mechanisms that would facilitate the new decoding processes required. They also argue that not all linguistic innovations need to be due to linguistic change – useful linguistic innovations could set up pressures for speakers and hearers to grammaticise such innovations (730).

Another common objection is that a natural language either functions as a whole, or does not function at all; hence intermediate forms of grammar would be useless. Pinker and Bloom (1990:731) answer, much as Darwin (1985:217-224) did with regard to intermediate forms of the eye, that grammar did not have to have its current structure and capacity in order to be useful. They argue that pidgins, contact languages, the language of tourists, children, telegrams and headlines prove that there is a continuum of viable communicative systems, even if they differ in efficiency and expressive power. Also, language could, in part, have

evolved as the result of learning mechanisms that may have existed prior to mechanisms specifically dedicated to language, which may eventually have led to language-specific learning mechanisms that would allow more complex forms of language to be acquired.

In order to ascertain what reproductive advantages language may have bestowed on our ancestors, we need to have some conception of the kind of life-world they may have inhabited. Pinker and Bloom (1990:720) point out that modern hunter-gatherers, whose life-style very likely mirror that of our ancestors, have an intricate folk biology encompassing knowledge of life-cycles, ecology, and the behaviour and properties of plants and animals. Such evidence, as well as fossil evidence, indicates a lifestyle that was completely dependent on extensive and intricate knowledge acquired from the environment. Furthermore, pedagogy is a universal human trait, and it is reasonable to suppose that pedagogy relies heavily on language. There would be a selective advantage in being able to receive and communicate vital information more efficiently. Thus, they argue, humans are and always have been social animals and would always have relied on complex social interactions and cooperative efforts among non-kin for survival. Pinker and Bloom (1990:733) argue that language would always have been woven into such interactions, and in a manner that was qualitatively no different from that of current interactions, where socially-relevant information such as "time, possessions, beliefs, desires, tendencies, obligations, truth, probability, hypotheticals, and counterfactuals" can be exchanged.[86]

In terms of the dynamics of cooperative interactions among individuals, Pinker and Bloom (1990:734) argue that major opportunities arise for evolutionary gains and losses. Over and above the benefits gained from group living, such a set up is also vulnerable to invasion by cheaters, who could reap the benefits of group living, without paying the costs. In order to counter the dangers posed by cheaters, individuals need a cognitive apparatus that at the minimum allows them to remember individuals and the ability to enforce social contracts, in order to ensure that costs are paid. Such contracts would require quite subtle linguistic abilities. Of course, it would be to the benefit of cheaters if they develop the ability to convince others that they have paid costs that they had in fact not paid. In turn, the group would have to develop more accurate measures for detecting such cheating behaviour, and so on. Added to these pressures are conflicts of reproductive interests between males and females, siblings, parents and offspring, and so forth. The argument is that this "cognitive arms race" led to the rapid evolution of linguistic abilities, which would among other things, allow for intentions to be stated, and promises and threats to be made.

5. The Evolutionary Mind

Pinker (2003) updates the thesis put forward by him and Bloom and places much emphasis on what Tooby and DeVore have called "the cognitive niche". With their theory

[86] In his updated version of the classic 1990 paper, Pinker places emphasis on the suitability of the structure of grammar for conveying information about technology as well (2003:27).

Tooby and DeVore (1987) tried to develop a unified explanation of the human traits that are idiosyncratic compared to the rest of the natural world. These idiosyncrasies include the manufacture of and dependence on complex tools, a wide range of habits and diets, extended childhoods and long lives, hyper-sociality, complex patterns of mating and sexuality, and divisions into groups or cultures with distinctive patterns of behaviour (Pinker 2003:27). These characteristics point to the evolution of a mind with extensive computational capacities and the ability to generalise extensively. Pinker (1997:187) argues that most theories about the reason for the natural selection of the human mind are "badly underpowered" when we take the result into account. If particular, and rather narrow, problems such as making tools or forming social bonds in a large group were the cause of the selection for the mind, the result could be said to be overkill. He points out that comparable examples in other species point to the tendency of natural selection to form "idiot savants", where animals develop specific abilities to address specific problems. What problem could the mind with its computational capacities and ability to generalise be designed for? Tooby and DeVore (1987) argue that humans have entered a unique "cognitive niche." They work from the Darwinian observation that species evolve at the expense of one another in the struggle for survival (Darwin 1859:114-129). Most food is part of the body of another organism, which means that organisms need to develop defences against being eaten; eaters, in turn, need to overcome such defences. The modification of defences are genetically based, and hence remains relatively stable within the lifetime of an individual. Tooby and DeVore's suggestion is that humans entered into the cognitive niche because it enables them to overcome the genetically based pattern of defence and attack and hence tailor strategies to various situations within the lifespan of the individual (Pinker 1997:188; 2003:27). Through learning which manipulations will receive which goals, humans can develop novel courses of action to reach particular goals. Humans can create new knowledge by "mentally playing out" possible combinatorial interactions between forces in the world, and predicting an outcome (Pinker 1997:188).

> Humans analyse the world using intuitive theories of objects, forces, pasts, places, manners, states, substances, hidden biochemical essences, and for other animals and people, beliefs and desires… Living by their wits, human groups develop sophisticated technologies and bodies of folk science. All human cultures ever documented have words for the elements of space, time, motions, speed, mental states, tools, flora, fauna, and weather, and logical connectives (not, and, same, opposite, part-whole, and general-particular). They combine the words into grammatical sentences and use the underlying propositions to reason about invisible entities like diseases, meteorological forces, and absent animals. Mental maps represent the locations of thousands of noteworthy sites, and mental calendars represent nested cycles of weather, animal migrations, and the life histories of plants (Pinker 1997:188).

In this understanding, language is a means for exchanging such knowledge, hence increasing the benefit of knowledge, which embodies knowledge that might have been acquired by expensive trial-and-error for which someone else paid the cost. Knowledge can also be exchanged for other resources through language, at a negligible cost to oneself (Pinker 1997:191; 2003:27-29).

Hence, the argument is that language meshes with the other features of the cognitive niche. Humans rely on cause-and-effect reasoning about the world, and have evolved the ability to encode and share that information. Furthermore, we are an extremely social species because information is such an important commodity and can be gained from other's experience. The long childhoods and extensive parental investment from both parents form a sort of apprenticeship where we have time to learn knowledge that has been accumulated thus far. Investment in children results in the pay-off of another information-gathering individual, which explains why fathers also invest extensively in children. This explains the unusual sexual and social arrangements that humans have developed such as marriage and families. Such arrangements connect men to their children and the mothers of those children (Pinker 2003:28-29). Pinker describes culture as "part of the pool of local expertise", and variation in practises (and language) can be explained as a result of the fact that survival skills and local information are acquired through a local network of information sharing (29). This ability to formulate abstract knowledge allows for flexibility which allows humans to spread over such a wide range of habitats. On this view, then, three key features of the distinctively human lifestyle – know-how, sociality, and language – co-evolved, each constituting a selection pressure for others (Pinker 2003:29).[87]

---

[87] Pinker (1997:191) speculates that our ancestors possessed four traits that probably tipped the scales in favour of developing better powers of causal reasoning in the way in which it is embodied in the human mind: Firstly, primates are visual animals, with depth and colour perception, which Pinker speculates causes the brain to split the flow of visual information into two streams, a system for objects and their shapes and compositions, and one for their locations and motions. He argues that this is the reason that we conceive of the world (including abstract conceptions) as a space filled with movable objects. Abstract thought has co-opted the coordinate system and inventory of objects made available by the visual system (191). This allows the possibility of conceiving of mental models of three dimensional spatial and mechanical relationships that characterise our mental life.
A second possible factor that contributed to the development of the human mind is group living (Pinker 1997:192). Group living poses challenges that could give rise to certain cognitive developments. Apart from the advantages of living in a group, such as security, cooperative hunting, etc., there are also disadvantages. The danger of being exploited by other group members arises, whether it be through theft, cuckoldry, extortion, and so forth. (1997:193). Being smarter becomes a requirement for preventing such exploitation. Furthermore, cooperative behaviour requires relatively complex communication. Social animals exchange favours, enforce and repay debts, form coalitions, and punish cheaters. Pinker (193) does not believe that group living in itself accounts for human intelligence. In these terms, the cost of intelligence would have exceeded the benefit of the intelligence gained, especially considering that human intelligence is not limited to social intelligence, but extends to mechanical and biological niches as well.
A third factor contributing to human intelligence is the hand (Pinker 1997:194). Primates evolved to be able to grab branches and developed hands that are capable of manipulating objects with some precision. Hands have also allowed primates to manufacture simple tools, such as using stems and rocks to get at food that might not be obtained in any other way. The upright posture of primates compared to other mammals allowed hands to be freed and thus available for use in manipulating objects. Pinker argues that having trees as a habitat caused apes to evolve a body plan that would eventually allow for an upright posture. He speculates that being bipedal would have been advantages

It should be emphasised that Pinker and Bloom raised some of the significant issues with regard to an evolutionary theory of language and set the scene for a veritable explosion of theories on the topic. Theirs is by no means the last word on the subject and all aspects of their theory have consequently been debated, disputed, amended and reformulated. For our purposes, their paper is an example of what an evolutionary approach to mind and language would look like in that it raises some of the most significant issues in this regard. In many ways, their arguments in this early paper are still highly speculative. In chapter 5 we will examine an attempt to formulate a more comprehensive account of the evolution of mind and language, based on current theories in various disciplines related disciplines, such as psychology, neurology, archaeology, etc. One of the most significant developments after the "evolutionary turn" with regard to language and mind is the move towards incorporating evolution into computational theories of mind.

5. The Advent of Computational-Evolutionary Theories of Mind (and Language)

In the wake of the explicit incorporation of evolutionary theory into theories on the origin of language and into contemporary approaches to the computational theory of mind, the trend has arisen to combine computational theories of mind with evolutionary theory in order to account for the evolution of both language and mind. The work of Steven Pinker subsequent to "Natural Language and Natural Selection" (1990) is a good example of this development. Pinker (1997:21) argues that the mind is "what the brain does".[88] In this understanding the evolution of mind is inherently linked to the evolution of the brain. And one of the things that "the mind does" is language. Thus, Pinker combines the ideas of the computational theory of mind and the theory of the natural selection of replicators to explain the origins of both mind and language. The result is a comprehensive theory, incorporating a wide range of disciplines, including anthropology, palaeontology, psychology, philosophy, biology, cognitive science, neuroscience, and so forth. The interdependence of the brain/mind, language, the evolutionary environment, human physiology, the social

---

upon entering that savannah, both in terms of being able to cover a lot of ground and being able to see across the grass for quite some distance.
The final impetus to the development of human intelligence would have been hunting (195). Pinker does not concentrate so much on the cognitive abilities required for hunting and gathering. He adheres to the proposition of Tooby and DaVore that hunting provided sporadic episodes in which greater concentrations of nutrients were consumed. Meat made up a far greater proportion of the diet of foraging humans than of any other primates. Hence, meat obtained from hunting may explain how our ancestors could afford to evolve the expensive, energy-consuming brains that would allow for human intelligence to evolve. Hunters could also move into latitudes and altitudes that might not support sufficient plant life during the winter. Furthermore, meat would have been a major currency in social life, and would have influenced sexual politics, as well as the investment that males may have made in their offspring (196-197).

[88] The computational-evolutionary approach is also exemplified by the work of Dennett, hence his definition of mind: the mind is "not a miracle-machine, but a huge, semi-designed, self-redesigning amalgam of smaller machines, each with its own design history, each playing its own role in the "economy of the soul" (1995:206).

environment, language structures, learning abilities, the reproductive forces driving natural selection, and a range of other factors is taken into consideration in such accounts. In this approach, it is generally accepted that the mind needs to be understood in the context in which it is used and for which it was developed (note that the context can change over time). It is assumed that the structure and abilities of the mind is fundamentally linked to the role that it can play in terms of the "goals" of natural selection, namely survival and reproduction. It is generally accepted that a computational approach to the mind in this context cannot ignore the substrate in which such computation takes place, namely the brain. Hence, a computational "machine" in this sense does not equal a "Turing machine" or an "idealised neural network". Furthermore, it is generally accepted that structure of the nervous system and the body both enables and constrains computational abilities. Various roles are attributed to language, in this regard, both in terms of its role as a source of vital information which the mind "computes", and as an entity that can literally shape the structure of a brain (to a greater or a lesser extent, depending on the theory).

As an example of such a computational-evolutionary theory of mind and language a brief overview of Pinker's theory is useful. Pinker's argument is that the structure of the mind is best studied by approaching it with the assumption that it was designed by natural selection to solve the kinds of problems that the species has been faced with for most of its existence. In his view these problems include: "understanding and outmanoeuvring objects, animals, plants, and other people" (1997:21). He also defines the mind as "what the brain does", i.e. the processing of information, thus thinking is seen as "a kind of computation" (21). He proposes that the best way in which to study the mind is to reverse-engineer it as one would any machine that was designed for a specific purpose. The rationale for this approach is of course Darwin's theory that natural selection seizes upon random, advantageous mutations in species because such mutations enhances the organism's fitness in the struggle for existence and hence ensures that it increases its progeny. To current sensibilities such an approach might seem obvious, and even banal, but as we have seen there has been significant resistance to, and scepticism about, such an approach, for various reasons, well into the twentieth century.

In this understanding, the mind is designed to solve engineering problems, with various components dedicated to their own problems (Pinker 1997:4). Pinker argues that a system dedicated to solving particular engineering problems cannot accumulate information from its environment at random, but must be equipped with a set of "core truths" about its environment and a set of rules from which the implications of these truths can be deduced (14). He refers to these rules as the "rules of common sense". Hence, by means of the "rules of common sense" the mind can deduce *relevant* implications from what it knows. Pinker argues that generations of philosophers overlooked the implication of having to "frame" knowledge in terms of relevance to the organism because of the apparent effortlessness of the common-sense in terms of which they thought. It is only with the advent of artificial intelligence and its attempt to duplicate our common-sense in computers that the problem of

framing knowledge became a serious topic of study (15).[89] Our common-sense understanding of the world now seems to be a function of the structure of our minds rather than incontrovertible and universal laws.

Thus, in terms of this approach, the mind consists of modules, which are designed for specific interactions with the world. The basic structures of the modules are specified by our genetic programme and they were shaped by natural selection in accordance with the problems that the species had to solve in order to survive (Pinker 1997:21). Pinker incorporates into his thesis Dawkins's (1976:23) conception of organisms as "survival machines", struggling with one another and with the environment in order to ensure that the genes that they carry are replicated (we have seen that Darwin lacked the concept of genes and hence was at a loss to explain how characteristics are mutated and carried over into future generations). Hence, our minds are machines that have been designed to solve survival problems that we face as an organism in order to ensure the survival of the species through replicating its genes. With the problem of mind framed in this manner we have a point of departure in terms of which we can answer the question "What is mind?"

An important result of framing the problem of mind in this manner is the emphasis that is placed on the inherent structure of the mind and the role that it plays in structuring the information that we receive from the "external" world. The argument can be made that much of our common-sense understanding of the world and many of our goals had evolved as adaptations to the environment in which our ancestors lived. Such adaptations would generally be relatively universal across human cultures. The structure of the mind is in all probability universal, in a manner similar to the universal structure of the rest of human physiology.[90] Pinker argues that much of human behaviour that appears to be non-adaptive in terms of natural selection (such as celibacy, or adoption) can be accounted for in terms of the discrepancy between the environment that the species evolved in for the major part of its history, and the extensive changes that the particular environment underwent in our relatively recent history. Furthermore, the design of the mind should be taken into account in this regard. To a large extent human behaviour is also the outcome of the internal struggle between various mental modules and adaptations (Pinker 1997:42). The mind should not be viewed as a homogenous structure, harmoniously pursuing all goals in common; various adaptations to the ultimate goal of replicating genes could very well be at odds with one another, even within the same organism, especially if complex external factors come into play. Behaviour is an indirect product of our genes (Pinker 1997:42). Pinker uses the example of the indirect relation between behaviour and genes to emphasise Dawkins's argument that genes are "concerned" with their own replication. The argument is not that all human activity is aimed at spreading an individual human's genes. As Pinker puts it, genes "selfishly" spread

---

[89] Pinker (1997:83) suggests that the proper name for the study of mind informed by the computational theory of mind should not be Artificial Intelligence, but Natural Computation.

[90] Evolution is a homogenising force, in that it selects beneficial variations, which are then propagated through the species through successive generations. Inevitably, variants that are less beneficial or not improvements at all, disappear from the species (Darwin 1985:115ff.).

themselves by means of the way in which they build organisms, including our brains (44). It is important not to conflate the "goals" of genes and natural selection, with those of human beings in their everyday lives. His argument is that confusing these separate categories has lead to much of the scepticism about the value of evolutionary theory to the study of mind. Organisms and their genes may very well be in conflict, in a manner of speaking.

Pinker describes the "form of mental representations" in the brain as "the symbol inscriptions used by the mind" (1997:84). He argues that the way in which people generalise is the most significant indication that people make use of mental representations (85). Mental representations connected with a concept, forms something like a semantic network (87).[91] He speculates that we have several kinds of "data representations" in our heads; and that there are at least four major formats of representation that the brain uses (89-90): i) the visual image; ii) phonological representation, which is an important component of our short-term memory; iii) grammatical representations; and iv) mentalese, which he describes as the language of thought "in which our conceptual knowledge is couched". He argues that these different formats of representation evolved because they allow "simple programmes to compute useful things from them" (91). Presumably different mental modules make use of different representation formats in their mental computations.

The computational theory of mind has quietly infiltrated neuroscience; the entire field is informed by the idea that the fundamental activity of the brain is information processing (Pinker 1997:83). Pinker (1997:25) describes the computational theory of mind as holding that beliefs and desires are information, incarnated as configurations of symbols. The content of brain activity lies in the patterns of activity among the neurons. He (62-63) draws on the work of Allen Newel and Herbert Simon and formulates the following definition of intelligence: the ability to attain goals in the face of obstacles by means of decisions based on rational (truth-obeying) rules; intelligence entails specifying a goal, assessing the current situation in terms of that goal, and applying a set of operations that reduces the difference between the goal and the current situation. Anthropomorphically speaking, intelligence entails having desires, pursuing them, and using beliefs about the world that are effective in that they are at least approximately or probabilistically true. It is the structure of the mind and its resultant "common sense" orientation to the world that explain the existence of beliefs and desires within the organism. It is a function of our understanding of organisms that they act in order to accomplish some goal (food, shelter, movement, and the like) that causes us to view them as having beliefs and desires. Pinker refers to this orientation to organisms as "intuitive psychology" (Pinker 1997:63), and it is a powerful tool for predicting, controlling, and explaining behaviour. Hence, intelligence relies on information, and information is a correlation between factors that can be said to be produced according to a lawful process (as opposed to a correlation due to chance) (65). Information is found wherever causes leave effects (66).

---

[91] This conception of mental representations has much in common with Derrida's concept of "writing" as was examined in the previous chapter.

Pinker holds that the influence of language on both thought and feeling has been exaggerated in philosophies of mind and language (1997:366). He goes on to suggest that our language of thought may have its origin in our ancestor's ability to reason about space and force. The parts of their brains dedicated to this function could have served as "scaffolding" onto which symbols were added for more abstract concepts, such as states, possessions, ideas, and desires (355). The abstract domain of "mentalese" would then have a logical structure that mirrors objects in motion, and makes inferences in terms of the logic of motion, as "space- and force-simulators" (356).[92] Metaphors of space and motion symbolise possession, circumstance, time, and causation (356); hence his conclusion: "I suspect that parts of our mental equipment for time, animate beings, minds, and social relations were copied and modified in the course of our evolution from the module for intuitive physics that we partly share with chimpanzees" (1997:357).

It is in terms of such "intuitive physics" and "intuitive psychology" that the problem of intentionality can be approached. Searle is correct when he argues that AI does not present us with programmes that are intentional in the sense that human beings are intentional. The contentions here is not that intentionality depends on "biochemistry", which sounds needlessly mystical, but that it depends on the context and goals of the system that is said to possess it. Human beings act intentionally, and expect other human beings and animals to act intentionally, in terms of the evolutionary goals that led to the development of the brain/mind. Evolutionary speaking, humans (and all other organisms) act in terms of the "goal" of survival and reproduction. It is important to note that these goals are not necessarily the "conscious" goals of the organisms. As Dawkins's (1976) points out, the goals of survival and reproduction apply at a genetic level; these genetic goals do not translate directly into "human level" gaols.

"True" intentionality can then be seen as acting in terms of the goals that originate from the system itself, in terms of its evolutionary purpose (such a system would usually be able to access, navigate and represent its environment in terms of its own configuration), and "pseudo" intentionality can then be seen as goals that are programmed into the system/ or ascribed to the system by a (human) designer. Such systems would appear to act in terms of goals, but such goals would be "externally given" and the system would have no way in which to access, navigate and represent its environment without the intervention of a designer/programmer external to it. If it were to become possible to build robots that are able to internalise certain goals and pursue them through interaction with its environment (without

---

[92] Location in space is one of the two fundamental metaphors in language, used for thousands of meanings. The other is force, agency, and causation (Pinker 1997:354). Thus he concludes (Pinker 1997:355):

> …These concepts and relations appear to be the vocabulary and syntax of mentalese, the language of thought. Because the language of thought is combinatorial, these elementary concepts may be combined into more and more complex ideas…the discovery that the elements of mentalese are based on places and projectiles has implications for both where the language of thought came from and how we put in to use in modern times.

the intervention of a designer/programmer), such a system would for all intents and purposes have "real" intentionality.

In terms of Pinker's argument, the human mind is not adapted to think about arbitrary abstract entities like the rules of mathematical logic; we think in terms of our inherited framework made up of the key features of encounters among objects and forces, as well as other important themes of the human condition, including fighting, food, and health (Pinker 1997:358). However, since human thoughts are combinatorial and recursive, breathtaking expanses of knowledge can be explored with a finite inventory of mental tools (360).

As can be imagined, Pinker's theories and others of its kind are not without controversy. Over and above the objections from theorists that would conceive of the mind in more traditional, "non-naturalised" terms, there is much disagreement among various evolutionary theorists on mind and language about most of the particulars within the approach itself. Some of these disagreements include: what the adaptive function is that language fulfils (or originally evolved to fulfil)[93]; how language should be characterised[94]; which pre-adaptations in the hominid lineage allowed for the emergence of language[95]; how the study of the evolution of language should be approached[96]; the relation between human and non-human language[97]; whether grammar is innate or emergent through cultural transmission[98]; whether symbolic reference is biological or cultural, what the constraints on the development and transmission of language are and what their various effects are on how to interpret the fossil record[99]; whether language originated as a result of manual gestures, or was originally vocal[100]; the relation between language and mind/brain[101], and lastly the causal capacity of language (Christiansen and Kirby 2003:12-15). Such disputes do not concern us as much as the question as to whether such evolutionary approaches to language and mind could overcome some of the problems run into by traditional cognitive science and philosophy of mind. It is the contention of this paper that the evolutionary approach to mind does, indeed, seem more promising than the other approaches that we have touched upon.

There is, however, one criticism against Pinker's argument that is pertinent with regard to our discussion, namely his sustained adherence to certain Chomskyan conjectures,

---

[93] Some of the possible adaptive functions suggested include (Pinker 2003:26-31): in order to exchange vital information (Tooby and DaVore); in order to deceive and manipulate others (Dawkins); in order to facilitate thought (Chomsky); as a substitute for grooming in maintaining social relationships (Dunbar); or as a courtship devise to advertise fitness of the brain (Miller).

[94] Christiansen and Kirby (2003a:300).

[95] See, for example, Hurford (2003:57).

[96] See, for example, Botha (2004).

[97] See, for example, Hauser and Fitch (2003:158-181); and also Hauser, Chomsky, and Fitch (2002) and the reply by Pinker and Jackendoff (2005).

[98] See, for example, Tomasello (2003:94-110) and Deacon (2003: 111-139).

[99] See, for example, Davidson (2003:140-157).

[100] See, for example, Arbib (2003:182-200), Corballis (2003:201-218), and Dunbar (2003:219-234).

[101] See, for example, Lieberman (2003:255-271).

such as the existence of a language faculty and Universal Grammar. Pinker's (1995) contention is that innate grammatical knowledge evolved by means of natural selection, hence endowing Homo sapiens with a "language instinct". Deacon (1997), for example, argues that such a form of innate organisation is highly unlikely, and he argues that it offers little more than the "miraculous accident" theory of Chomsky, with some evolutionary embellishments. As we shall see in the following chapter, Deacon makes use of an extensively interdisciplinary approach to substantiate his contention and to formulate an alternative theory, which is evolutionary and overcomes the hold that Chomsky has on the "poverty of stimulus" problem.

6. CONCLUSION

The evolutionary approach to mind and language is not entirely new. We have seen that Peirce, for example, who strongly argued for collaboration between philosophy and the sciences, incorporated evolutionary theory into his semiotic theory.[102] However, this potentially fruitful approach fell by the way side for various reasons, most notably the pervasive effect of mind-body dualism in the Western philosophical tradition. Hence, cognition was considered to be of a different order from the physical world, and thought was conceived of as rational processes, something akin to logic. We have seen that with Gödel the possibility arose that thought might be modelled and studied and thus be amenable to scientific study, and the sciences of mind were born. Prior to this development, language was seen as the only access that we have to the mental realm, and hence was inextricably intertwined with theorising on aspects of the mind. This pairing survived well into the twentieth century, underpinning those philosophical traditions that considered philosophical problems to be fundamentally related to language.

Towards the end of the twentieth century the promise of the computational approach to mind had begun to wane. Various attempts at modelling mind failed to render models with

---

[102] We have seen that Peirce's theory of abduction leads him to a decidedly anthropomorphic take on the universe (1965 vol. I: 159 [1.316]):

> [E]very scientific explanation of a natural phenomenon is a hypothesis that there is something in nature to which human reason is analogous; and that it really is so all the successes of science in its application to human convenience are witnesses. They proclaim that truth over the length and breadth of the modern world. In the light of the successes of science to my mind there is a degree of baseness in denying our birthright as children of God and in shamefacedly slinking away from anthropomorphic conceptions of the universe.

Peirce argues that through evolution, the universe would become "an absolutely perfect, rational, and symmetrical system, in which mind is at last crystallised in the infinitely distant future" (Peirce 1965 Vol. VI:26 [6.33]). We have seen that Peirce takes this route, because he does not believe that a "mechanical" understanding of the universe could account for the mind as part of the physical world, where free will and consciousness are not determined and illusory. We have argued that Peirce still adheres to a mind-body dualism, where the body is physical and hence subject to mechanical laws, and the mind is essentially not-physical and hence subject to different laws. In this manner he aimed to guarantee human autonomy and free will. However, with his "law of mind" Peirce also seemed to deviate further and further from his intention of grounding philosophy in the scientific and the empirical and he constructs a metaphysics that seems extraordinary and peculiar to contemporary sensibilities.

the complex and extensive capabilities that the human mind exhibits. The dream of artificial intelligence, which seemed a certainty only a few decades earlier, remained elusive. Furthermore, the computational approach seemed unable to explain basic aspects of the human mind, such as how information is encoded in it, or how the information embodied in it is able to refer to the world. It seemed that the assumption that intelligence comprised of the computation of information regardless of the substrate in which such computation takes place, was flawed. Furthermore, a scepticism developed with regard to notions such as representation and the role that symbols and internal representations play in human intelligence. Despite some moves toward interdisciplinarity in the cognitive sciences, philosophy of mind remained on the whole impervious to developments within the sciences of mind, especially given the fact that they seemed unable to account for "mental" phenomena such as intentionality, qualia, the methodological validity of introspection and so forth.

We saw that as work in the field developed, the realisation began to take hold that computation cannot be considered to be a disembodied process. The substrate and context in which computation takes place plays a fundamental role in determining what is to be computed and how. It is in this milieu that evolutionary theory began to regain prominence on the intellectual scene. Pinker and Bloom (1990) indicated, with regard to language, that not only did language exhibit the characteristics that one would expect from an entity that had developed by means of natural selection (apparent design and adaptive benefits), but also that an evolutionary approach held much promise for solving questions that had consistently plagued language theorists, such as the apparently idiosyncratic (i.e. non-"logical") structure of grammar and Plato's problem. The promise of the evolutionary approach was also extended to theories of mind. Not only does the mind show apparent design and distinctive adaptive benefits, the possibility arose that computational approaches of mind could be combined with evolutionary theory. In this manner it became possible to "contextualise" or "embody" the computational processes of the mind. A problem horizon presented itself, namely that the mind computes information that is relevant to the organism of which it forms part, in order to facilitate the survival of the organism. Furthermore, the move to an evolutionary explanation for mind and language had the, perhaps unexpected, result of creating more distance between language and mind than had existed under the assumption that the mind "computes symbols". A mind that receives information from all of the senses of the organism that it inhabits surely does not "think" only in terms of language symbols. The argument was put forward that the human mind evolved to enable *Homo sapiens* to inhabit the "cognitive niche" and that language is one aspect of this niche.

The evolutionary approach to language and mind is still in its infancy, but is developing rapidly. Unfortunately the scope of this work does not allow for a comprehensive review of the developments within the field. In the following chapter we will examine one computational-evolutionary theory of mind, which is also a co-evolutionary theory of language and mind, by Terrence Deacon. His theory aptly illustrates how many of the theoretical questions regarding mind and language that have been raised in the course of this work can

be (convincingly) addressed by an evolutionary approach to both phenomena. Furthermore, Deacon manages to extricate his theory from the Chomskyan paradigm much better than Pinker does, to good effect. It also illustrates that both phenomena are amenable to scientific investigation. The argument will be that is impossible to study mind and language without taking their evolutionary origins into account.

CHAPTER 5

DEACON ON HOW EVOLUTION TRANSFORMS THE MIND (AND LANGUAGE)

The Darwinian Revolution is both a scientific and a philosophical revolution, and neither revolution could have occurred without the other (Dennett 1995:21).

Hundreds of millions of years of evolution have produced hundreds of thousands of species with brains, and tens of thousands with complex behavioural, perceptual, and learning abilities. Only one of these has ever wondered about its place in the world, because only one evolved the ability to do so
Deacon (1997a:21).

…assumptions about the nature of language and the differences between non-human and human minds are implicit in almost every philosophical and scientific theory concerned with cognition, knowledge, or human social behaviour. It is a truly multidisciplinary problem that defies analysis from any one perspective alone, and where the breadth of technical topics that must be mastered exceeds even the most erudite scholar's capabilities. So it is hard to overestimate the immensity of the task or the risks of superficial analysis, and it is unlikely that any one account can hope to achieve anything close to a comprehensive treatment of the problem
(Deacon 1997a:14).

1. INTRODUCTION

In the preface to his ambitious book, *The Symbolic Species: The co-evolution of language and the human brain* (1997a:15), Terrence Deacon declares his sympathy usually lies with "the cranks and the doubters and against well-established doctrines". He (15) gives the following reason for this tendency:

> …like Dewey, I believe that the search for knowledge is as often impeded by faulty assumptions and by a limited creative vision for alternatives as by a lack of necessary tools or critical evidence.

It is in this respect that Deacon's project in *Symbolic Species* can be said to be philosophical. He questions much of the accepted wisdom in many of the disciplines that are concerned with language and/or mind and explores the implications that a radically

interdisciplinary approach to language and the mind will have on such received wisdoms. The result is a vast and ambitious work that manages to raise as many questions as it purports to resolve. Philosophically speaking, one of the most important implications of Deacon's work is to highlight the irreducibly interdisciplinary requirements of subject matters as complex as language and mind. Another is the central role that is given to evolution, with respect to both the human mind/brain and language regarding their origins, functions, structures, and interaction. It becomes clear that evolution is the key to understanding human languages and the human mind/brain. Disregarding this foundational and essential element in both questions will inevitably lead to skewed and fanciful speculation.

As will become apparent, Deacon makes use of all of the primary theorists and their major insights that we have discussed in the foregoing chapters in constructing his theory of the co-evolution of language and the brain. He does not necessarily agree with all or any of them, but uses their insights as foundations in terms of which he can construct his analysis. True to the quotation above, Deacon attempts to rectify faulty assumptions and present alternative interpretations of the evidence that we already have at our disposal regarding languages, brains, and evolution. Inevitably, it will be impossible to cover all or even most of Deacon's arguments in the space available in a work of this nature. The following discussion will concentrate on those aspects of Deacon's work that are the most relevant to the topics that have been raised in the foregoing chapters. I refer the reader to Deacon's own work for an in-depth and comprehensive exploration of the implications of a co-evolutionary approach to language and the mind/brain.

As we shall see, Deacon sees the evolutionary origins of language as an entry point into the question of the logic linking our cognitive functions to brain organisation. Essentially, the question that Deacon addresses is "Where do human minds come from?" And essentially his answer is: From our ability to create and manipulate symbols, which stems from the novel form of mind that developed in the interface between the evolution of the human brain and human language. In effect he argues that: "Biologically, we are just another ape. Mentally, we are a new phylum of organism" (Deacon 1997a:23). Furthermore, he puts the development of the distinguishing human feature – the ability to think symbolically – to within the last 2 million years.

2. HUMAN LANGUAGE AS AN EVOLUTIONARY[1] ANOMALY

Deacon argues that the difficulty of the origins of the language question does not necessarily lie with what we do not know, but with what we think we know (1997a:25). There is a widespread assumption that language is limited to humans because of its complex structure and the demands that it puts on memory and learning. There is the further assumption that some sort of change in the human brain (usually an increase in size and intelligence or computing capacity) enabled the species to acquire language. Deacon argues that we can be sure of none of these "facts" and that they may even bias our enquiries into the subject matter, sending us on a wild goose chase. He attempts to question these basic assumptions in an effort to discover where the human ability to learn and use language comes from. Simply invoking "increased intelligence", "evolutionary progress", the existence of "mentalese", or a "special language faculty" does not offer anything by way of explanation.

An important aspect of Deacon's analysis is the distinction that he makes between human and animal languages. He examines how language differs from other forms of communication in the biological world and the result is a complete re-evaluation of the evolution of both human language and the human brain. He stresses that there are continuities between human and non-human brains,[2] but also insists that there is a singular discontinuity between human and non-human brains that accounts for the ability of human brains to learn language (Deacon 1997a:13).

Often, animal communication is viewed from an anthropocentric perspective, and is considered to be a simplified form of human language. Deacon does not consider animal languages to be a primitive variation of human language (Deacon 1997a:12). To him, the evidence suggests that animals are incapable of learning human language, or an equivalent thereof. As we shall see, he considers human language to be an evolutionary anomaly (34). Furthermore, he believes that it is this anomaly that ultimately accounts for the difference between human and non-human brains. It is human language that accounts for the qualities of the mind that we consider to be uniquely human.[3]

---

[1] When thinking about the evolution of language it is necessary to keep in mind that we have to do with evolution on three different time scales. Human language arises from three interacting adaptive systems: individual learning, cultural transmission, and biological evolution (Christiansen and Kirby 2003:302-303). All three of these systems (the period of language acquisition and lifetime of the individual, the lifespan of a language, and the lifespan of a species) are subject to evolution, although at very different rates, with the evolution of language and culture being much more rapid than biological evolution. Furthermore, these three systems are fundamentally interactive, with the learner's ability to acquire language determined in part by biological evolution and in part by cultural factors, while the structure of language is determined in part by cultural factors and in part by the genetic biases of learners.

As we shall see, much controversy still remains with regard to the structure, biological status, origins, and evolutionary history of these systems.

[2] See Deacon (1997a:145-320) for an extensive analysis of some of the continuities and discontinuities between human and non-human brains.

[3] Deacon (1997a:22) defines human language as, among other things: "a means for generating an infinite variety of novel representations and an inferential agent for predicting events, organizing memories and planning behaviour."

He goes on to suggest that we distinguish between human language and other forms of communication. One of the main differences between them is that human language is not just a mode of communication, "it is also the outward expression of an unusual mode of thought – *symbolic representation* [my emphasis]" (Deacon 1997a:22).[4] Deacon bases his assertion that other animals do not have access to symbolisation and the virtual world that comes with it, on his contention that the ability to think symbolically is not innate, but *develops with the internalisation of the symbolic process that underlies language*. Hence, only minds with the ability to communicate symbolically can think symbolically.

2.1 Language vs. Communication

One of the misleading evolutionary assumptions in this regard is that the emergence of language is an inevitable outcome of evolution because it is advantageous in terms of survival (Deacon 1997a:28).[5] In its most simple form, this view includes the assumption that human beings are at the top end of the evolutionary scale and that evolution embodies a progressive trend towards developing a sophisticated communication system. Species with smaller brains are often seen as "primitive" – leftover from an earlier evolutionary stage. Such assumptions rely on the faulty common-sense idea of progress in evolution, where evolution is seen as a cumulative process of design.[6] Deacon favours a conception of evolution as a process of diversification and distribution in all directions – a process of exploiting any possibilities that may present themselves (28-29). Constraints on viable "design" options do exist: for the smallest organisms there does seem to be a limit on the internal resources that can be devoted to representing the world beyond the system. But Deacon contends that "the upper end of the range of information-handling abilities was not similarly bounded, and so the difference between the low end and the upper end of this range has increased over the hundreds of millions of years of animal evolution as part of this diversification" (30). Deacon describes this as a "minor evolutionary niche", which was eventually filled and found helpful by human beings. He also believes that we are near the extreme for the information-handling distribution.

Not only do we often underestimate the complexity and subtlety of animal communication, we also tend to undervalue forms of human communication other than (symbolic) language (e.g. gestures, facial expressions, stance, crying, laughing, etc.). However, Deacon insists that even the most complex repertoire of communication in animal species do not seem to map onto any of the elements that make up (human) language

---

[4] In a similar vein Smith (2002:65) defines human language as "learned, symbolic, and exibit[ing] syntactic structure, a set of properties which make it unique among naturally occurring communication systems".

[5] See, for example, Komarova and Nowak (2001:1190) where such an assumption is made when attempting to model the most efficient language acquisition profile.

[6] Cf. Toulmin (1990) where he argues that viewing the world in terms of "progress" is a particularly modernistic phenomenon, stemming from the Enlightenment.

(Deacon 1997a:31). Furthermore, he does not believe that we might someday discover that at least some animal species possess a language equivalent to human language after all. Deacon argues that there are striking characteristics in human language that do not seem to be present in any form of animal communication studied thus far (32).[7] For animal language to exhibit human language-like characteristics it would have a combinatorial form with distinguishable elements that are able to recur in different combinations. A creative productivity of diverse outputs would be evident and there would be a limited amount of large-scale redundancy. Furthermore, there would be a high degree of variety in the possible combinations of the language elements; however, the majority of combinatorial possibilities would be systematically excluded. Finally, there would be a relatively *low* correlation between signals (calls, gestures, etc.) and contextual features and correlations between signs and their contexts should differ radically, yet systematically, on different occasions[8] (32). Deacon finds that animal languages do not exhibit these requirements. The communication systems and communicative behaviours of other species tend to exhibit isolated signals in fixed sequences, or in comparatively unorganised combinations (33). Furthermore, calls, gestures, etc. always seem to correspond in a one-on-one manner with events and behavioural outcomes. Hence, Deacon concludes that non-human forms of communication and human language are very different entities and should not be treated as different kinds of "language". Animal language therefore does not qualify as “language” in terms of Deacon’s definition.

Deacon identifies human counterparts to animal communication in the form of social communication through gestures and facial expressions, which seem to be universal to human beings. He insists that these human forms of communication are something other than language, and not a sort of "language without words" (34). In terms of this analysis, human language is not a more complicated version of a simpler system of communication; animal languages are not simple languages. Nor is human language the result of evolutionary progress, potentially accessible by other species "lower down" on the evolutionary ladder. In fact, Deacon argues that language is an evolutionary anomaly.

He does not take this to mean, however, that language is separate from the rest of our biological and neurological make-up. In other words, he does not see language as a "freak mutation" that happened to produce an organism capable of creating and manipulating language (Deacon 1997a:35). He rejects Chomsky’s postulation of the existence of a Universal Grammar, essential to all human brains as an example of recourse to a "freak mutation"-theory in order to account for our linguistic abilities. Deacon believes that the appeal of the Universal Grammar theory lays in its ability to account for a number of features of language that seem to be otherwise inexplicable: For instance, the apparently

[7] The necessary characteristics would be embodied in the surface features of such a non-human language, and would not require that we understand the meaning of what is being communicated.

[8] The final requirement may seem counterintuitive, but it represents perhaps the most salient feature that distinguishes (human) language from animal communication systems. More attention will be given to this aspect of language in the ensuing discussion.

insurmountable problem of acquiring a complex grammar, and at a very young age at that. If language were intrinsic to the brain, this difficulty is apparently overcome. Furthermore, a "language organ" unique to humans might account for the failure in animals to acquire language. Adding an extra part to the human brain for processing language seems to account for the relatively large size of the human brain. A common neurological language structure apparently accounts for the presumed features that all languages have in common, as well as the fact that human languages are inter-translatable. Finally, an innate language feature points to a definite and identifiable transition point in the evolutionary development of the human species: the point of mutation, where the language organ comes into existence. Such a dramatic event of transition would not require functional explanations, seeing that the accidental organ would have accidental utility (36).

Deacon rejects the accidental mutation argument because it enables us to ignore "the messy details of language origins". It also invites us to not look for precedents, causal mechanisms, or deep design logic to the structural and functional relationships of language grammars and syntax. Deacon's reservation with the theories of grammar instinct or universal grammar is that they fail to reduce the phenomenon of how language abilities are produced to basic causal mechanisms. In his words, these theories become placeholders for what ostensibly cannot be learned (37).[9]

According to Deacon the question as to the origins of language is dominated by two paradigms: the evolution of greater intelligence and the evolution of a specialised language organ (49). Both of these theories are aimed at resolving the problem of having to learn a large and complex set of rules and signs. In terms of these theories, other species do not possess language because it is too complex for them to learn and perform. Hence, human beings' language abilities either indicate that they are that much more intelligent, or that they have an innate language organ. Deacon argues that both these views only address one of the main problems with regard to language abilities, and it is not the most important one (39).

[9] To quote Deacon (1997a:38) in full:

> A grammar instinct or a universal grammar serve as placeholders for whatever could not be learned. The nature of this presumed innate knowledge of language is described only in terms of its consequences. Linguists have progressively redefined what supposedly cannot be learned in ever more formal and precise terms, and so we may get the feeling that these accounts are approaching closer and closer to an explanation. But although the description of what is missing has gotten more precise, ultimately it is only a more and more precise version of what is missing. These "explanations" of the nature of a language instinct are inevitably presented in the guise of elaborate definitions of grammatical principles or else as something akin to computer programs, and in this way they are only more formal statements of the problem of the missing information. Saying that the human brain alone produces grammar because it alone possesses a grammar factor ultimately passes the explanatory buck out of the hands of linguists and into the hands of neurobiologists.

2.2 LANGUAGE AS ITS OWN PRIME MOVER?

Deacon questions whether the complexity of language is the reason that it is limited to human beings. According to him, the most distinguishing feature of language, namely symbolic reference, cannot be accounted for by its complexity (Deacon 1997a:40). He identifies a minimal definition language in the generic sense as a mode of communication, involving combinatorial rules that comprise a system for representing synthetic logical relationships among the symbols – this definition includes music, etiquette, religious ceremonies, games, etc (41). In terms of this definition, one only needs a tiny vocabulary and two or three types of combinatorial rules in order to have a functional language. Such a language-like system would be even simpler than the communicative repertoire that many other species already possess. In other words, complexity does not need to be a prohibitive factor when it comes to constructing a language. Yet, we do not find any communicative systems among the vast majority of other species that fulfil these basic, and comparatively simple, criteria (barring perhaps those experiments in which primates were successfully taught some form of language-like communication). This, despite the apparent evolutionary advantages of mastering language. Hence, Deacon concludes, although an important factor, the complexity of language is a secondary consideration.

He argues that if complexity is not the prohibiting factor for language learning in other species, then theories that aim to explain language evolution in terms of overcoming such complexity lose their plausibility. He argues that even minimal powers of inductive learning would be sufficient for acquiring a simple language. Learning a simple grammar and syntax should be a trivial task for species that do possess inductive learning abilities. Although such minimal inductive abilities seem to be present in various species, there are no correlating simple languages present. Deacon's conclusion is that the (inductive) learning problems addressed by most theories on language only suffice to show that non-human languages should be simple in comparison to human language, not that they should be ubiquitously absent.

The more basic mystery to do with language seems to be the human/non-human difference. Why are there no simple human languages, or any non-human languages that we know of? It seems that complexity is not the most significant difference between human and non-human language. The only significant difference that remains is that of word meaning and reference, or, the problem of figuring out how combinations of words refer to things. In Deacon's words: "Neither grammar, nor syntax, nor articulate sound production, nor a huge vocabulary have kept other species from evolving languages. Just the simple problem of figuring out how combinations of words refer to things" (Deacon 1997a:42). In other words, the answer lies in the presence of symbolic reference in human language as opposed to non-symbolic reference in non-human species and some human forms of communication.[10] For

[10] Mortensen and Kirby (2003:301) give the following useful definition of symbol use:

> [S]ymbol use is typically construed as a capacity for linking sounds or gestures arbitrarily to specific concepts and/or precepts – in particular for the purpose of communication.

some reason, other species find it extremely difficult to grasp symbolic reference, and humans don't. The crux of Deacon's argument here is that, when it comes to theories of language and mind, the presence of symbolic reference in human language cannot be taken as a given. It is a language phenomenon that is central to the structure of human language and the human mind, and needs to be accounted for before other aspects of mind and language can be comprehensively addressed (43). Deacon takes note of the fact that there have been numerous attempts at explaining the basis for symbolic reference, but asserts that the issue remains "curiously unresolved"(44).

A further implication of the argument that language complexity is not the primary barrier against language acquisition in other species is that many theories on the evolution of the human brain have inverted the relationship between brain evolution and language evolution. It becomes possible that the supports that were thought to be prerequisites for the evolution of language – greater intelligence, articulatory abilities, etc. – may have been *consequences*, rather than causes of language acquisition. Deacon believes that language was "its own prime mover": "[i]t is the author of a co-evolved complex of adaptation arrayed around a single core semiotic innovation that was initially extremely difficult to acquire" (Deacon 1997a:44). Modern languages developed from simpler languages that must have existed in our prehistory, and as language became more complex, our brains followed suit.

If modern language evolved from simple languages and created a novel mode of information transmission (over and above the information transmission function of DNA), our ancestors were sent down a novel evolutionary path. Furthermore, symbolic language must have created selection pressure to reshape the brain, which means that the structure of the brain should reflect language structure in its architecture in some way. "That which is most peculiar about language processing should correspond to that which is most peculiar about human brains" (Deacon 1997a:45).[11] Hence, Deacon sets out to explore how human cognition differs from non-human cognition, which he believes is reflected in the difference between human and non-human languages.

2.3 On Reference

Deacon begins his discussion with the notion of being "pre-maladapted" (a notion akin to Peirce's *abduction* – see Chapter 1). The argument is that the members of a particular species would be naturally biased to attend to certain details of a problem and to ignore others, given their biological make-up, their habitat, their objectives, etc. These characteristics would preclude the members of a certain species from attending to the relevant details when it comes to solving a certain problem. In this regard, the species is pre-maladapted to a

[11] Deacon (1997a:145-320) devotes a third of *Symbolic Species* to analysing idiosyncrasies in human brain structure compared to a variety of other animals. Unfortunately spatial restrictions prohibit us from examining these structural differences. The reader is referred to Deacon's work for fascinating (and essential) reading.

certain problem and would in all likelihood not manage to overcome these innate biases.[12] The suggestion is that, because language is an unprecedented form of naturally evolved communication, it would need a different problem-solving orientation than other forms of communication in order to learn it. The fact that animals seem to find it extremely difficult to even grasp a rudimentary form of language, suggests that it is not just difficult to learn language, but that it goes against certain other predispositions – learning a language may run against other, more common, learning strategies. Deacon goes on to suggest that the most likely candidate for this "counterintuitive" element in language, to which most non-human minds might be pre-maladapted is representation, more specifically, symbolic representation. Very few non-human animals seem to exhibit symbolic capabilities (Deacon 1997a:50)

Deacon suggests that our continued inability to fully explain the problem of word meaning could indicate that "[i]t is not just a difficult puzzle; the concept seems to be as counterintuitive for us to understand as it is simple for us to use" (51). In Chapter 2 we saw that Chomsky (2000b:113-114) argues that common-sense conceptions of phenomena are responsible for many pseudo-problems within philosophy. However, he also speculates that the problem of reference may not be amenable to scientific study (1995a:43). Deacon sets himself the task of submitting the problem of reference to systematic, scientific analysis. What is more, he aims to do this by means of evolutionary theory – something that Chomsky is highly sceptical about. Deacon thus argues that it is essential to take the anomalous nature of the symbolic form of reference into account when analysing the problem, as it might be its rarity that is the key to understanding it. Theorists often assume that other species exhibit counterparts to words and sentences, and tend to treat forms of communication that animals exhibit as simple languages. As we have already seen, Deacon disagrees with this assumption and argues that it is important to separate language (meaning linguistic communication) and other forms of communication. The problem with treating language as a model for analysing other species of communication is that it treats human and animal understanding as equivalent. When trying to analyse animal communication as a form of language, all non-linguistic forms of communication comes to be seen as exceptions to the rule, whereas it is human language that is the most exceptional case of communication (Deacon 1997a:52). Deacon believes that overcoming this anthropomorphic bias could lead to many insights regarding all forms of communication, whether olfactory, vocal, or linguistic. In this analysis then, animal communication is not a more primitive version of human language, and the early stages of language evolution did not consist of a form of crippled language, equivalent to that of children or brain-damaged speakers (53).

[12] An example of where humans have exploited pre-maladaptedness is another species is the use of "baboon-proof" clamps on rubbish bins in areas where baboons tend to frequent. Whereas the baboons have no trouble in gaining access to normal rubbish bins and feasting on the contents, the addition of the specially designed clamps has stopped them from opening modified rubbish bins. Apparently the baboons (who can be innovative and persistent in gaining access to food) are unable to figure out how the mechanism works. Presumably, their innate make-up does not allow them to hit upon a strategy for manipulating the clamps – they are pre-maladapted to this problem.

Inverting the communication/language hierarchy also inverts traditional conceptions of their functional dependence.[13] Non-linguistic communication is self-sufficient, in that it does not need support from linguistic language to be acquired or understood. Linguistic communication needs various forms of non-linguistic communication (prosody, pointing, gesturing, objects, interaction with other people, etc.) for both of these purposes (Deacon 1997a:53). Deacon believes that writing has provided language with partial independence from such non-linguistic support. In this context it seems absurd to invert the communication/language relationship and treat language as the standard. It seems more plausible to argue that language evolved parallel to human calls and gestures (which are equivalent to non-human communication) and that it still compliments language (54).

Deacon argues that an important cause of the misconception that animal communication is somehow equivalent to human language (such as single word sentences, but without evident syntax) is rooted in misconceptions about the concept of reference (Deacon 1997a:54). The idea is that a repertoire of calls within a species constitutes the vocabulary of a sort of protolanguage. Once the possibility of a protolanguage has been established, it becomes tempting to infer a scenario where language evolves, with such a protolanguage of calls and gestures evolving into "language proper" – human language. This scenario depends on a consistent understanding of reference, meaning that the reference that applies to calls must be the same "kind" of reference as that which applies to words. It is precisely this assumption that Deacon tries to undermine. He does *not* argue that reference is unique to human language. He understands it to be "ubiquitous in animal communication" (Deacon 1997a:57). An example would be laughter, which Deacon describes as an innate human call. It need not be intentionally produced and it usually points to something, whether an internal state or the object that caused it. In Deacon's words, "it points to a definite class of experiences that are deemed funny" (57). In animal communication, alarm calls, etc. refer to their object in the way that laughter does, and not in the way that words do.

Deacon also distinguishes between intentional and unintentional or automatic forms of communication. Innate social signals are usually involuntary (laughter, crying, grimacing) and contagious (Deacon 1997a:58). We often find that, when confronted with another person laughing, people feel the compulsion to laugh as well, regardless of whether they are aware of the cause of the laughter. "Normal" language generally does not create the impulse to echo what is being said by the hearer. Social signals imply a redundancy, in that the repetition or echoing does not convey new information. The original laughter conveys something about the situation, or even the state of mind of the person producing the laughter, but in being echoed, no new information is produced. Conversations, on the other hand, require conscious effort and monitoring to be produced, and following speech requires attention and intentionally controlled analysis. Usually, conversation seems to be aimed at conveying information that the other party does not already possess. According to Deacon, in the case of human

[13] In a sense, Deacon can be said to be performing a Derridean deconstruction.

linguistic communication there is less redundancy in the information being conveyed, than there would be in the case with calls (59). The conclusion is that "reference" as such does not constitute the difference between calls and words – both can refer to things in the world and to internal states. But there is a difference between the *kind* of reference in linguistic and non-linguistic communication. Deacon sees the source of misunderstandings with regard to reference as a simplified understanding of reference (59).

Another dimension is added to the problem when the question of “meaning” is included to the enquiry. Traditionally “meaning” and “understanding” are conceptualised as the realm of linguistic communication and a human prerogative. Do calls (animal or human) have meaning? Deacon asserts that laughter and calls can indicate something (about the person’s “inner state” and the context), but that they do not have meaning in the same way that words have meaning (Deacon 1997a:61). He proceeds to distinguish between different modes of reference, which may be interdependent, in an attempt to overcome any misunderstandings that may arise because of this difference.

For Deacon, the primary difference between the way in which a word refers to something and an animal call refers to something would be the seemingly arbitrary nature of the relationship between the signifier (word/concept) and the signified (object/state of affairs) in linguistic reference, as opposed to non-linguistic reference, which seems more straightforward with a one-to-one mapping of the signifier and the signified (Deacon 1997a:61). Referring to the interpretation of a call or gesture as interpreting its "meaning" reveals a linguistic bias on our part.[14] He aims to establish that reference is historically dependent and can change. Furthermore, he wants to apply this logic to reference in general in order to support his theory on the evolution of a link between alarm calls and predators, or laughing and humorous experiences. This would allow him to demonstrate that reference has a hierarchic character and that some forms of reference (such as those exhibited by a smile, for example) need not be determined by a conscious concept or meaning. However, this concept of reference does not seem sufficient to account for words, although words might display this aspect of reference. He speculates that the sort of reference that is applicable to words becomes a special case of reference, to which the interpretive process is crucial. This special case of reference would be the result of evolution, and particularly human evolution. Furthermore, it is not intrinsic to signs, but is something that is created in response to them–

[14] Deacon (1997a:61) makes use of Frege’s distinction between the sense and the reference of a term to develop his view. In Frege’s theory, sense is the idea in the mind, while reference is the thing in the world which corresponds with the term and its sense (morning star, evening star – has the same reference (the planet Venus), but different senses (which are the result of an historical context in which they were thought to be two different bodies). The common-senseidea here (as in much subsequent language philosophy) is that what the word refers to results from the idea that it invokes. The question that inevitably arises, of course, is what are "ideas"? And how do ideas and their "objects" correspond, if at all.

And, as Deacon points out (1997a:62), the referential power of words can be independent of their sense in many cases, and we may discover that the sense of a term does not refer to the accepted reference. Thus, falsifying the accepted sense of a term need not influence its reference.

the result of “some cognitive action”, which is interpretive and can determine reference in different ways, as well as "different references for the same sign" (62-63).

Deacon models this postulated *interpretive response* on the theory of Charles Sanders Peirce, and accordingly names the interpretive processes under discussion *interpretants.*[15] Peirce might strike one as an odd choice, but Deacon’s reason is that "Peirce recognised that *interpretants* cannot only be of different degrees of complexity, but they can also be of categorically different kinds as well" (Deacon 1997a:63).[16]

The argument is that words will have different interpretants than those of calls and gestures. Deacon argues that it is the symbolic nature of words that give them their meaning. In terms of Peirce’s theory, Deacon believes that this symbolic nature requires interpretants over and above those required by other forms of reference (interpreting the source of an odour, for example). These interpretants are the result of "a great deal of additional learning", which has allowed for humans being able to produce them (1997a:64). As with Derrida, Deacon argues that the symbolic basis of words is further mediated by their underlying systemic connections. In being interpreted, words are interpreted in terms of the many associations (traces) that they have with other words.[17] This can account for the fact that even words that refer to abstract concepts, and evoke no mental imagery (e.g. falsity, capability, etc.) still have meaning.[18]

Of course, forms of reference other than word-reference would lack the additional interpretants (other words) that Deacon envisages word-reference to have. Against this background, Deacon believes that it is possible to train almost any intelligent mammal to use a complicated sign system that is appropriate to its sensorimotor abilities. The individual would be trained to produce certain behaviours in response to certain stimuli and perhaps to use such stimuli reciprocally, something akin to Skinner’s trained pigeons. Deacon (66) believes that most animal communication is partially or wholly dependent upon the use of signs in this way, whether the signals used are learned and arbitrary (primates taught to manipulate signs), or innate and linked to a state of arousal ("natural" animal calls). He describes these kinds of communication as "mechanical" in that the suspicion always lingers that animals that use such forms of communication do not "understand" what they are "saying". The signs are highly effective in the right contexts, but the possibility of applying

[15] Such criticism is given by Lumsden (2002:158-160) for example.

[16] See the discussion on Peirce’s interpretants in Chapter 1. In short, an interpretant is the process that enables one to infer the reference from a sign and its context, whatever that process may be.

[17] Deacon does not hold that mental imagery does not exist. He argues (1997a:63) that mental imagery has been recognised as an experience with neural and behavioural correlates (also see Pinker 1997:84-93). He believes, however, that mental images are only one sort of interpretive response that can be elicited by words, and "it may not be the most important one" (63). He argues that words may also elicit something like a "dictionary definition", a word with a related meaning, behaviour, or even a feeling corresponding to past experience that the word evokes. Deacon (64) raises the possibility that mental images may be the primary interpretive action in many non-symbolic reference processes.

[18] Contrary to Pinker (1994: 55-82), for example, it does not become necessary to evoke "mentalese" in order to explain the referential capabilities of words that do not refer to concrete objects. The meaning of such words is constituted through their relation to other words within the language system, rather than through their correspondence to imagery or internal tokens of whatever kind.

them appropriately in a new context is lacking.[19] Deacon suggests that the ability to apply signs in new and appropriate contexts, without having to learn what these contexts are by rote beforehand, could be a good indication of understanding their *meaning* (1997a:67). Such an ability suggests a radical change in cognitive strategy.

Innate calls and gestures and learned associations seem to rely on a relatively fixed correlation between the sign and a particular state of affairs in the world – such as a specific call that indicates the presence of a predator. Deacon suggests that a degree of a necessary, immediate, and stable correlation between the sign and its object is always involved when it comes to instances of communication, or "non-language". The correlation can be due to an innate feature, or due to learned associations between external stimuli. It is the concept of "necessary association" that is crucial to Deacon's distinction between non-language communication and language. Instances of non-language communication rely on "a relatively stable correlation with what they refer, in order to refer" (67). If the sign and its referent were to consistently fail to occur simultaneously, the association between them would become weaker, and might eventually disappear. With words, the scenario is completely different. There may be some degree of correlation when it comes to word-reference, but the co-occurrence of a word and a certain state of affairs is not a prerequisite for a word to have meaning. In fact, co-occurrence is often the exception rather than the rule.

Deacon's insight is akin to Derrida's when he criticises the approach of much of western philosophy to language as being caught up in the "metaphysics of presence" where the disclosure of the truth is the presentation of the thing itself (e.g. Derrida 1986:6; 10).(see Chapter 3). The traditional conception of reference seems to have myopically been focussed on an object-reference relationship, where two phenomena are involved that have to be somehow present to one another (either physically or in terms of shared characteristics) in order for reference to be established. Derrida's inversion of the spoken vs. written language dichotomy serves to highlight the same point that Deacon is making – correlation is not a prerequisite of (symbolic) reference. In fact, the symbol must be able to function in the radical absence of its referent. As we shall see, Deacon reaches the same conclusion that Derrida does in another respect – symbolic reference is distributed across the systemic relationships of the symbols within the system. (Note that Deacon makes an explicit distinction between "symbol" and "sign", as does Peirce. Derrida does not consistently apply these terms distinctively).

Deacon wants to establish that the relationship between words and their referents is the reverse of that assumed by Saussure,[20] for example. He wants a theory of reference that

---

[19] Such a use of signs is not only applicable to animal behaviour. Deacon (1997a:67) gives the examples of applying a mathematical formula and coming up with the right answer, without understanding what one is doing, or using a word that one doesn't know in order to impress someone. In both these cases one can hit upon the right strategy or meaning, but without understanding the meaning behind the symbols, it remains a matter of chance.

[20] Deacon argues that the immediacy of Saussure's mapping of signifiers to signifieds is applicable to alarm calls and other forms of animal communication, but not to word-reference (1997a:69).

would not only account for the relatively low rate of correspondence between words and their referents, but which would also enable us to distinguish between the rote understanding of words that a trained animal, or uncomprehending child would have, and, what he calls, the "semantic" understanding of a competent human language user. In his words: "The correspondence between words and objects is a secondary relationship, subordinate to a web of associative relationships of a quite different sort, which even allows us reference to impossible things" (Deacon 1997a:70). Deacon finds that theory in the classification of possible representational relationships, put forward by Charles Sanders Peirce. Philosophically speaking, this is one of the most criticised aspects of Deacon's work, and it merits close scrutiny.[21]

3. PEIRCE'S CONTRIBUTION – ICONS, INDICES, AND SYMBOLS

Deacon's point of departure is the belief that semiotic (encoding) systems underlie biological *emergence*[22] (Goodenough and Deacon 2003:801-805). This position has some parallel with Peirce's semiotic conception of the world, although Deacon does not go to the metaphysical lengths that Peirce does. Thus, Deacon sees biology as a "semiotic science", where representation and significance are central. Furthermore, he argues that semiotics is emergent by definition. The reader will recall that Peirce regarded the ontological world as an essentially semiotic system. Deacon argues that DNA is a semiotic system, in that it carries interpretable information; so too are hormones and molecules diffusing from decaying food

---

[21] Lumsden (2002:155), for example, contends that Deacon's notion of symbols diverges from Peirce's notion, and as such "needs to be investigated on its own terms in order to evaluate the idea that the human species has crossed the symbolic threshold." Cowley (2002:78), on the other hand, accuses both Deacon and Peirce of using *symbol* with "systematic ambiguity" and "as determinate entities that lack iconic and/or indexical aspects", which, as we shall see, is untrue for both theorists.

[22] Goodenough and Deacon (2003:801) propose that emergence can take on the following three forms:

First-order emergence: Here properties emerge as a consequence of shape interactions, such as the interaction of water molecules that generate the property of surface tension.

Second-order emergence: Here properties emerge as a consequence of shape interactions played out over time, and future events are significantly influenced by past events in the system. He views the properties generated by complex, self-organising systems as this type of emergence. (In this regard one should add that Deacon's distinction between second and third-order emergence only holds if he views self-organising systems in this context as artificial, i.e. as having their operating parameters imposed by a designer. Third-order emergence would then occur "naturally" in that feedback from environmental factors and natural selection fulfils the function of a "designer" in this regard and hence set up the constraints within which systems operate).

Third-order emergence: Here properties emerge as the result of some form of memory in the system. an example would be genetic instructions in biology, which place constraints on second-order systems and hence specify outcomes that are usually called biological traits. Third-order emergent traits are substrates for natural selection because their instructions are encoded and they endow organisms with adaptive properties.

Goodenough and Deacon argue that it is with third-order emergence that *telos* enters the world (here, again, we can see the parallel with Peirce's conception of Thirdness). What they means with *telos* in this context is that biological traits are about something, or for something. Motility, for example, allows an organism to move towards food or a mate and away from danger. In his words "there is a point to a trait that we cannot ascribe to a snowflake" (2003:804). Organisms, by virtue of their evolution, appear to obey teleological principles (see Chapter 4).

(804). All of these are biological entities that carry information which is interpretable by, and conveys meaning to, other biological entities. The argument is that the biological traits of specific organisms evolve to "fit" the organism to a given niche, and many such biological traits endow organisms with receptors by means of which they can negotiate their environments. Such receptors detect relevant signs in the environment and convey their meaning to the organism ("food", "danger" etc.). Deacon believes that such semiotic processes are ubiquitous, operating from the level of single-cell organisms, right through to more complex central nervous systems and brains. The hierarchy descends from humans, through other primates, to pets, to species such as flies snails and fish. The descent ends with animals or species that are thought to be able to perform a minimalistic sort of reference. Brains operate semiotically in various ways – their receptors include sensory receptors, hormones, neurotransmitters, etc. and their apparently teleological behaviour could be said to be vastly more complex and multi-faceted than that of an amoeba, for example. Deacon calls this instance of biological semioticity *brain-based awareness*.

Deacon (1997a:73) links this picture of diminishing competences to what he calls one of Peirce's "most fundamental and original insights" about the interpretative process, where the difference between different modes of reference is understood in terms of levels of interpretation. Different interpretations of signs can be placed in a hierarchical order, where prior interpretative competence is a prerequisite for a following level of interpretation. In this understanding, reference has a hierarchical structure in itself, where more complex forms of reference are built up from the simpler ones. This hierarchical model of reference allows him to explain the similarities and differences between words and animal calls. His argument is that most forms of animal communication are indexical, while human language is symbolic.

Thus, for Deacon the difference between the various semiotic systems in nature lies in the kind of reference that such systems exploit. He bases his analysis of the various possible semiotic systems on Peirce's semiotic theory, which we discussed in Chapter 1. Deacon makes use of Peirce's analysis of the three possible sorts of reference – iconic, indexical, and symbolic – and applies it to both human and animal interpretive abilities. Deacon's appropriation of Peirce is a heuristic move, rather than a wholesale endorsement of Peirce's greater project.[23] Thus, his theory cannot be faulted simply by highlighting perceived inconsistencies with Peirce's broader concerns (cf. Lumsden's (2002) criticism). Deacon expands on Peirce's analysis, attempting to apply it to biological processes. As Short succinctly puts it: "To do Peirce justice, we may have to go beyond him" (1996:493).

The reader will recall form the discussion in Chapter 1 that Peirce distinguished between iconic, indexical, and symbolic signs. We saw that Peirce maintained that our knowledge, even knowledge that is apparently direct and intuitive, takes on the form of a hypothesis. The truth or falsity of such a hypothesis is something that is subsequently tested

[23] Deacon (1997a:467) tries to stay close to Peirce's original insights, specifically with regard to focussing on representation as a process and not a static relationship, and viewing sign types as hierarchical levels of representation and not as categorical alternatives.

through experience. If our underlying assumption is correct, the object of our knowledge should behave as we expect it to; if incorrect, it won't (Peirce 1965 Vol. V:135-147 [5.213-5.237]). Furthermore, he argued that whenever we know something, we primarily know it as something standing *in relation* to other things – we gain knowledge of things by relating and classifying them, something which cannot be done without acquiring signs of some sort. Hence, Peirce held that all knowledge relies on the ability to manipulate signs, and it is in the course of learning how to manipulate such signs that we are instilled with the "first assumptions" on which our subsequent assumptions are based. Hence, we gain all the knowledge that we have of ourselves and the world from *signs* (Peirce 1965 Vol. V:177-185 [5.294-5.309]). In this understanding, making or reacting to signs is being "engaged in being a mind" (Gallie 1966:81).

Given that Peirce defines thought as the process where signs develop in accordance with the laws of inference (Peirce 1965 Vol. V:169-170 [5.283-5.284]), he develops a semiotic (theory of signs) in an attempt to model the mechanisms behind thinking. It is his contention that the *mechanisms* (or logical structure) behind cognitive functions somehow mirror those that constitute signs, not that signs are represented by tokens in the brain. One of Peirce's most important insights is the realisation that signification (meaning) is not based on a relation between two phenomena, the sign and its object, but is always the result of a *triadic* relation. He believes all sign phenomena to have this general and distinctive characteristic (e.g. 1965 Vol. II:135 [2.229]; 1965 Vol. II:136 [2.230]. As we have seen, a sign is always something that stands for something *to somebody*; hence every sign has an *object* and an *interpretant*. A sign is not a sign by virtue of some intrinsic characteristic; it can only function as a sign when it is *interpreted* as such. Hence, all sign phenomena are irreducibly constituted by a sign, an object, and an interpretant and meaning can only come about within the object-sign-interpretant interrelation.[24]

Peirce distinguishes between three categories of possible referential associations that characterise signs – those of *icon*, *index*, and *symbol*, and he uses these three categories to describe the formal relationship that can exist *for the interpreter* between signs and their objects.[25]

For an object or a sign to be an *icon* of something else, it needs to resemble that entity in some way. According to Peirce, such a resembling characteristic is something that "naturally" belongs to the iconic sign, which it "would possess just the same though its object did not exist" (1965 Vol. V:50 [5.73]). Religious icons or statues of mythic creatures are examples that one can think of in this regard. The objects of these entities might or might not exist, but their existence does not influence the shape or the accepted resemblance of their signs.

---

[24] We have seen that the *object* is that which the sign refers to, while the *interpretant* is the process that enables one to infer reference from the sign and its context. Peirce emphasises that every sign is essentially incomplete, and our competent understanding or use of a given sign is always a matter of degree (Short 1996:127).

[25] cf. Peirce (1965 Vol. V:50-52 [5.73-5.76]) and (1965, Vol. III:211-214 [3.359 – 3.365]).

For an object or a sign to be an *index* of something else, it needs to be either causally related to, or spatio-temporally associated with, the thing of which it is an index. An index is relational by virtue of a characteristic that it would not possess if its object did not exist, but which it would continue to possess, whether it is interpreted or not. A thermometer is indexical – it indicates the temperature of its surroundings, to which it is related in time and space. It might cease conveying information if nobody where around to interpret its reading, but the thermometer would continue to function in the absence of interpretation.

For something to be a symbol of something else, the link between the two phenomena needs to be established by means of a convention, agreement, or code. A symbol does not need to be similar to its object, nor does it need any factual connection with it in order to function – it fulfils its function only when interpreted (1965 Vol. V:50-52 [5.73]). Thus, "[a]ll words, sentences, and other conventional signs are Symbols" (1965 Vol. II:165 [2.292]).

As we have seen, Peirce holds that symbols develop out of other signs (especially icons) and can function as icons, indices, or symbols, depending on the context and the way in which they are interpreted. This characteristic is possible because there is nothing intrinsic to the sign that determines its reference. What distinguishes symbols from icons and indices is that their referential ability, and in fact their very existence, is based on convention and on the existence of a "mind" (in whatever form), which is privy to that particular convention, and can thus interpret, create, and manipulate symbols. Furthermore, the meanings of symbols are constituted, and develop, through use and experience.

In keeping with Peirce's semiotic, Deacon is careful to emphasise that objects are not intrinsically icons, indices, or symbols, but are *interpreted* to be so. The interpretation depends on the response that they produce. "In simple terms, the differences between iconic, indexical, and symbolic relationships derive from regarding things either with respect to their form, their correlations with other things, or their involvement in systems of conventional relationships" (Deacon 1997a:71). Hence, things are referred to as icons, indices, or symbols, when they were *designed*[26] to be those things, or when they are most likely to be *interpreted* in such a manner. A further important qualification that he makes is that these three modes of reference are not mutually exclusive alternatives. It is quite possible that the same sign (object/event) can be icons, indices, and symbols *depending on the interpretive process* (72).

Deacon's hierarchy of diminishing competences is not simply based on an increase in complexity, but on the relationships between the different modes of reference.[27] Thus, while

---

[26] As we saw in Chapter 4, biological entities, as the products of evolution through natural selection, can be viewed as "designed" in that they have adapted to fulfil particular niches.

[27] In this regard it is useful to keep in mind Peirce's three kinds of interpretants, namely emotional, energetic, and logical. He proceeds to posit three stages in the logical interpretant: those of conjecture, analysing alternatives open to us, and finally the formation of habit (Peirce 1965:326 [5.475]; Peirce 1965: 326-330 [5.476 – 5.480]). According to Melrose (1995:502) the logical interpretant can be seen "in terms of a progression towards the symbol that must *pass through the icon and the index*, abduction and induction" (my emphasis). Melrose takes "the ambiguous status of the logical interpretant" in Peirce's analysis to indicate that there are two ways of thinking about language: i) language as cognition, "as the verbal counterpart of rational thoughts", and ii) seeing language as "subcognition", "a

an organism might be unable to grasp the symbolic reference of a sign, it might still be able to interpret the sign as an index, or failing that, as an icon. Learning to read and write can be described as ascending the referential competence hierarchy (cf. Deacon 1997a:74). Hence, according to Deacon's analysis, having the competence to interpret something symbolically depends upon having the competence to interpret a host of other relationships indexically, which in itself depends on the ability to interpret correlated relationships iconically. In other words, symbolic relationships are composed of the indexical relationships between sets of indices and indexical relationships are made up of the iconic relationships between sets of icons (Deacon 1997a:74). And being able to interpret a sign symbolically depends upon being able to master prior, and very different competences. In his words (74):

> What constitutes competence in this sense is the ability to produce an interpretive response that provides the necessary infrastructure of more basic iconic and/or indexical interpretations. To explain the basis of symbolic communication, then, we must describe what constitutes a symbolic interpretant, but to do this we need first to explain the production of iconic and indexical interpretants and then to explain how these are each recoded in turn to produce the higher-order forms.

In studying symbolic competence then, one would have to start with explaining how iconic and indexical interpretations are possible. Deacon (76) attempts to do this by asking what it is that makes iconic interpretation possible. He explains the iconic interpretive step in negative terms: it is what the interpreter (amoeba/bird/human) fails to do – the act of not making a distinction between the objects that it is interpreting. In this view, iconic reference is the *default interpretive position* – the interpretive process that takes place in all species with nervous systems when the production of new interpretants stops, whether due to lack of ability or lack of effort on the part of the interpreter. It is not necessarily on the basis of physical similarity that something iconically resembles something else, but on the basis of the interpretative process not differing from other interpretative processes, and hence attributing physical similarity to objects that are not similar.[28] A sign is interpreted, and thus seen to be a

---

statistically emergent phenomenon born of patterns of neural firings triggering other patterns of neural firings, sometimes in apparently random ways" (502). A second aspect of (ii) is the "obligation" to reconcile these two modes of thinking about language. In this way of thinking about language both cognition and subcognition are fundamental to understanding language (503). Subcognition is then compared to Derrida's views on language, specifically with his concepts of *différance* and *dehiscence*.

[28] Deacon (1997a:74) argues that camouflage is successful because of iconic interpretation as the basic interpretive step of organisms. Thus, insects or animals whose natural colouring enables them to blend into their surrounding environment manage to do so because of the interpretive response that is elicited in a potential predator. In effect, the predator interprets the colour of the insect or animal as indicating "more of the same"; more and undifferentiated environment. The predator fails to make an interpretive distinction when it "sees" the camouflaged animal because nothing overt triggers an alternative interpretation to the default one. Some of the features of the camouflaged animal are iconic of the background, from the perspective of the potential predator. In Deacon's (76) words:

representation, by being reduced (i.e., analysed into its component representations) to the point of no further reducibility (due to competence or time limitations or due to pragmatic constraints), and thus is ultimately translated into iconic relationships. This process does not necessarily require any effort; in many cases it is where the interpretive effort ceases. It can merely be the end of new interpretation, that boundary of consciousness where experience fades into redundancy (Deacon 1997a:77). In other words, the interpretive process that produces iconic reference is the process of the basic *recognition* of the external world by an organism with some form of nervous system. Iconic relationships are the most basic way in which things can be represented, and forms the basis for all other forms of representation.[29]

Interpreting indexical relationships follows on iconic interpretation. As mentioned earlier, indexicality requires connectedness (physical or temporal), or predictable co-occurrence between objects. As with icons, the indexical relationship rests on the interpretive response to it. But, whereas iconicity can be a function of interpretive incompetence, indexical interpretation requires a certain interpretive ability. In fact, Deacon describes icons as arising from "a failure to produce critical indices to distinguish things" (77). In other words, indexical association relies on a certain amount of learning. Deacon uses the example of the smell of smoke which indicates that something might be burning. Associating the smell of smoke and the likelihood of a fire is a learned association, arising from past experiences where the smell of smoke did indeed accompany a fire. The organism that makes this association should, on the basis of that association, be able to extrapolate that knowledge to novel situations. An important aspect of this argument is that such a basic indexical competence relies on the relationship between a set of *icons*. Having an indexical ability means being able to apply a given set of iconic relationships to a new situation. The (iconic) memories of the smell of smoke share similarities with both the current situation and with one another. Similar elements in these situations (smoke, fire, their co-occurrence)[30] should point to the likelihood that a similar co-occurrence is probable in this situation.[31]

> Iconic resemblance is not based on some prior ground of physical similarity, but on that aspect of the interpretation process that does not differ from other interpretive processes… Iconism is where the referential buck stops when nothing more is added. And at some level, due either to limitations in abilities to produce distinguishing responses or simply a lack of effort to produce them, the production of new interpretants stops. Whether because of boredom or limitations of a minimal nervous system, there are times when almost anything can be iconic of anything else.

[29] Deacon (2000:455-456) describes iconic interpretative ability as the basic requirement for sentience. Sentient organisms (organisms with some form of brain) must be able to represent the world to themselves in some way. In other words, a basic requirement is that they be capable of iconic and representational thought. Such organisms are also actively and spontaneously adaptive in response to their environment on the basis of information (e.g. aural; visual, etc.) received from the environment. Being able to form correlations between sets of iconic inputs constitutes learning. Learning is inherently an indexical process.

All nervous systems support iconic and indexical representational processes; this is a basic prerequisite for adaptation. Deacon suspects that each living nervous system exhibits consciousness with respect to the iconic and indexical representations it can support (Deacon 2000:499). He argues that a change in consciousness takes place when the transition is made from indexical to symbolic representation, and self-consciousness arises (see also Goodenough and Deacon, 2003).

[30] Deacon restricts himself to three icons that the situations could have in common, but it is important to emphasise that the icons common to a particular situation are potentially innumerable.

Hence, indexical interpretation depends on the conjunction of three iconic interpretations: i) the stimulus, which must be interpreted as an icon of similar occurrences, ii) the stimulus must also correlate with other stimuli in space or time, and iii) past correlations must be interpreted as iconic of one another. Deacon (1997a:78) explains his position as follows:

> The repeated correlation between the smelling of smoke and the presence of flames in each case adds a third higher-order level of iconicity. This is the key ingredient. Because of this I recognise the general similarity of the present situation to these past ones, not just the smoke and not just the fire but also their co-occurrence, and this is what brings to mind the missing element in the present case: the probability that something is burning. What I am suggesting, then, is that this is the result of a special relationship that develops among iconic interpretive processes.

Deacon's argument is that the relationships that develop between iconic interpretive processes form the basis for associative learning which in turn forms the basis for indexical interpretations. The relationship is necessarily hierarchic, with iconic relationships being a prerequisite for indexical reference, but not the other way around. The common-sense idea of forming an association when an organism learns to pair a sound, image, or sensation with something else is thus an indexical relationship.[32] Hence, the common-sense idea of learning a language by learning to pair words with immediately visible external objects is a process of learning to refer indexically, rather than symbolically. Such a process necessarily comes about through experience and learning and hence cannot be said to rely on any form of innate "neural tokens".

The iconic and indexical interpretive processes have implicit inferential possibilities by virtue of their relation with other icons and indices, past and present, real and imaginary.

---

[31] A distinctive characteristic of brain-based awareness is its *indexical* semiotic capacity (Goodenough and Deacon 2003:805). When the sensory systems are stimulated, they make synaptic connections in the brain with: i) neural pathways that encode memories of previous encounters with such stimuli, ii) pathways that encode its emotional and instinctual valences, and iii) learned associations between that stimulus and others relevant to its meaning (806). These reactions to stimuli are then converted into coherent behavioural responses. Deacon argues that brain-based behaviour is nothing but cellular awareness. Neurons are cells with the same kind of receptors, hormones, ion channels, etc. that single celled organisms and other cells such as fat cells have. However, he also argues that brain-based awareness is "something more". This "something more" lies, in the observation that, in theory, the indexical possibilities of learning are only limited by the kind of stimuli that the brain is equipped (and motivated) to perceive.

[32] Goodenough and Deacon (2003:805-806) argue that the brain (and all nervous systems) performs a "semiotic feat" in that it receives an array off information, or signs, about its environment from various neurons, as well as information from the rest of the body in the form through signs such as hormones and neurotransmitters. It integrates all of this information and oversees responses to it – a process which he terms brain-based awareness. Brain-based awareness is a primarily indexical process, in that a stimulus would trigger the synaptic connections in the brain with the neural pathways that encode previous memories of a similar stimulus, pathways that encode related emotional and instinctual valences, as well as learned association between the stimulus and ones related to its meaning. The triggered responses that are brought into indexical relation can be thought of as iconic.

According to Deacon, in the case of icons and indices, these are represented "by producing the perceptual and learned responses that would be produced if they were present" (Deacon 1997a:78). "In this sense mental processes are no less representational than external communicative processes, and communicative processes are no less mental in this regard" (78). Indexical relationships (physical and pragmatic relationships between objects, their signs, and their attributes) serve to ground symbolic reference in experience (Deacon 1997a:271). Semantic features are predicted on these potential indexical associations.

A similar logic applies to symbolic interpretation, and as with indices, symbols are constituted by relationships among indices, and therefore among icons as well. But Deacon does not simply believe this relationship to be correlated to the first pairing. In Deacon's (and Peirce's) hierarchy of reference, the referential relationships between words form a system of higher-order relationships that allows words to be *about* indexical relationships, rather than being indices themselves. In his words: "Their indexical power is distributed, so to speak, in the relationships between words" (Deacon 1997a:88). Symbolic reference comes from the combinatorial possibilities in the system of words. And words derive their meanings in context of other words, by systematically indicating other words. In other words, the human brain can use symbols (words) to refer to indexes and sets of indexes, and it can use syntax to indicate how words are related to one another (Goodenough and Deacon 2003:812).[33] This means that not only can words function as straightforward signs, but most of them can be used to point to complex sets of indices that are connected by means of complex syntactical relationships. Through the use of a system of symbols (words) with a logic that does not rely on spatial-temporal (indexical) correlation for reference, a virtual "reality" is created. Everyone who has access to the system of symbols and has the ability to think in symbolic terms can engage in semiotic activity, which is freed from iconic and indexical referential constraints. This allows for the conceptual or the abstract to have meaning, even in the absence of real-world antecedents (be it dragons or justice). On the basis of this analysis, Deacon makes a distinction between human and non-human language. An animal language typically consists of isolated "words" indexically paired to external objects or events, and lacks the regularities that govern possible combinations of those signs (812).

To Deacons' mind, Saussure's signifier/signified relationship is, in fact, an instance of indexical reference and thus no different from animal communication. Words in human language can serve an indexical function, but their reference would be derived from being in specific and unique contexts – an expletive might indicate being shocked at something that one sees. A toddler could use the same expletive, without any conception of its meaning, usually to the amusement of adult onlookers. How would "truly understanding" a word differ from the situation of the toddler? Or to put it differently, when would we know that the toddler understands the word he/she is using, and not just mimicking what he/she hears? A fairly accurate, if not sufficient, indication would be if the toddler uses the word in a new and

[33] Goodenough and Deacon (2003:812-813) suggest that the human sense of self, or self-awareness is made possible by symbolic language in that it also allows for self-reference.

appropriate context. But Deacon considers the possibility that it might be fairly easy to use a word in a situation that may appear to be novel, but which has features or relationships similar to the original context in which it was heard (80). He does not see such a situation as an instance of understanding the meaning of the word, but rather as an instance of iconic and indexical "understanding".[34]

As we have seen, for the indexical relationship to be possible there must be a spatiotemporal correlation between the word or sound and its object. When the correlation ceases to exist the association is forgotten and the word or sound loses its indexical power. Indexical associations also remain independent of one another. If one element in a set of associations is eliminated or changed it should have very little effect on the other indexical associations (Derrida 1997a:82). In the light of this, Deacon sees two crucial differences in the relationship between a word and its reference, which would indicate that the relationship is not indexical, but, in fact, symbolic. Firstly, there need not be any correlation in time and/or space for a word and its reference to remain associated. A word can be transferred to a new and completely different context with its reference relatively intact, and its symbolic power cannot be lost. Indeed, when it comes to abstract concepts such as "justice" or "incomprehensible", there doesn't seem to be any object for the word to be related to, certainly not a spatiotemporal one. Secondly, symbolic associations are not and cannot be independent. Deacon argues that where an animal that has learned to associate certain words with certain foods, these associations have little effect on one another (82). A certain association can be modified or dropped, without really influencing the other associations. The same cannot be said for word reference, because words essentially represent other words, "(i)n fact, they are incorporated into quite specific individual relationships to *all* other words of a language" (82). If the shared mapping between the words of a language breaks down, the reference of the word would also fail.[35] Deacon argues that it is this last characteristic of word reference that explains the first. A lack of physical referents (signifiers) does not cause the indexical associations of words to be lost, because the possibility of the indexical link is implicitly maintained in the relatively stable associations between the words themselves. Words have a dual reference – to objects and to other words. He concludes that the distinction between sense and reference (as already discussed) is a result of this duality of reference in words (83).

4. LANGUAGE LEARNING (AND UNLEARNING)

If Deacon is right in distinguishing between animal communication and human language in terms of this hierarchy of reference, and the ability to ascend it or not, his theory

[34] Deacon (1997a:443455) argues that Searle's (1981) Chinese Box thought experiment demonstrates (perhaps unwittingly) the difference that we intuitively feel exists between symbolic understanding as opposed to indexical "understanding".

[35] Deacon's analysis essentially mirror's Derrida's concept of *différance* (see Chapter 3).

has implications for learning, in particular. Typical human and typical animal learning would have to be different kinds of learning. What is it that allows human beings to be able to learn and implement a system of symbolic associations, while animals seem unable to do so (except, perhaps under certain experimental conditions)? How is this shift from "conditioned associations" to "symbolic associations" made (Derrida 1997a:84)? It is important to emphasise that Deacon does not see such learning abilities as a matter of degree – humans are not simply further along a learning continuum than their primate cousins are, for example. He argues for a fundamental shift in cognitive ability; a shift which is something of an evolutionary anomaly.

Deacon starts off by describing learning as essentially being a function of the probability of *correlations* between things, from the synaptic level to the behavioural level (Derrida 1997a:83). Furthermore, past correlations tend to be predicative of future correlations – in other words, the basis of learning corresponds to the basis for indexical associations. But, as we have seen, with symbolic relationships the emphasis shifts from such correlations to the relationships between symbols, which are not highly correlated.[36] Language does not lend itself to indexical learning. What Deacon suggests is that symbolic reference is a shift from associative reference and that learning it begins as a change in mnemonic strategy. "It is a way of offloading redundant details from working memory, by recognising a higher-order regularity in the mess of associations, a trick that can accomplish the same task [as indexical association] without having to hold all the details in mind" (89). Instead of stimulus generalisation or learning-set generalisation, a new kind of generalisation is employed, namely, logical or categorical generalisation. The referential relationship at issue here is not one of co-occurrence of stimuli, but one where new words are incorporated into the system in terms of their potential combinatorial roles. Being able to make such generalisations requires that the subject spots the regularities in the relationships between the words in the system. These regularities form patterns that serve as mnemonics to simplify the memory load (89). In other words, as a learner acquires language through individual rote association, he or she realises that there is "a system behind it all". Spotting the regularities in the system underlying language enables the language learner to shift perspective from associative predictions (indices) to symbolic predictions in terms of the higher-order regularities that emerge from the system (89). We do not readily find such logical systems (systematic sets of tokens) in nature, with which we can organise our associations – we have to create such systems. When such a system is available it allows a shift in mnemonic strategy from the indexical mode of representation to the symbolic mode. It is worth quoting Deacon (89) in this regard:

> The crucial point is that when such a systematic set of tokens becomes available, it allows a shift in mnemonic strategy that results in a radical

[36] Deacon (1997a:83) argues that while we begin learning symbols through symbol-object correlations, this is only the first step to learning "more crucial" relationships.

> transformation in the mode of representation. What one knows in one way gets recoded in another way. It gets *re-presented*. We know the same associations, but we know them in a different way. You might say we know them from both the bottom up, indexically, and from the top down, symbolically. And because this recoding is based on higher-order relationships, not the individual details, it often vastly simplifies the mnemonic problem and vastly augments the representational possibilities.

Symbolic representation also allows for a vast amount of implicit knowledge. The combinatorial rules that make symbolic associations possible ensure that it is not objects themselves that need to be encoded in memory, but the way in which they can be related. This means that new symbols can be incorporated into the system which is already encoded and related to others, without the need for memorising co-occurring symbols or objects. The more concrete and cumbersome indexical associations can thus be "unlearned", and the more efficient and powerful combinatorial logic of symbol relationships can provide mnemonic support for retrieving and reconstructing these associations when needed. Hence symbolic associations allow us to effectively ignore much of the vast web of word-object, word-word, and object-object indexical associations (Deacon 1997a:302). We are able to quickly determine which indexical associations are relevant and which are not in a particular instance. This allows for vastly accelerated language production and comprehension. In his words: "We become lightning calculators of reference" (302).

Deacon uses the concept of "unlearning" to refer to the necessity of indexical associations being superseded for symbolic reference to emerge (Deacon 1997a:92). A logically complete system of relationships among a set of symbol tokens is necessary to form a basis in terms of which symbolic associations can be made. The difficulty is that symbolic function emerges from the system, and learning begins, as it were, prior to recognising the symbolic function in a token or set of tokens. As Deacon (93) puts it:

> To learn a first symbolic relationship requires holding a lot of associations in mind at once, while at the same time mentally sampling the potential combinatorial patterns hidden in their higher-order relationships. Even with a very small set of symbols, the number of possible combinations is immense, and so sorting out which combinations work and which don’t requires sampling and remembering a large number of possibilities.

The prior associations that will eventually be recoded into a symbolic system might require a great deal of time and effort, but the implicit pattern in the relationship between the indexical patterns that allows for the symbolic recoding itself must be *discovered*. Deacon describes the mental process involved as suppressing one learned set of associations, in favour of another, which is derived from it. First, we need to learn the associations that will

make up the system of symbolic relationships as individual indexical referential relationships. Then a shift in strategy needs to take place, where these learned associations are seen in terms of a derived, higher-order associative strategy and reorganised accordingly. Deacon describes this learning strategy as an example of "insight learning" – the recoding of previously available, but unlinked fragments of information (Deacon 1997a:94). Hence, it seems likely that being inclined to search out new perspectives, or linking available knowledge in new ways, might offer a significant advantage in discovering symbolic relationships. Learning which tokens can and cannot be combined creates a new level of mapping to "semantic features" (95). It becomes possible to add new elements to the existing system in terms of common reference with semantic features already defined in the system, or by identifying new features that can be integrated with existing ones. In this manner, the system of symbols grows and it is even possible to integrate separately constructed systems of symbols.

As Deacon envisages it, there are hierarchies within the realm of symbolic associations and each "level" requires a new process of "insight-recoding" (the learning/unlearning process). He describes more abstract concepts as new levels of symbols coding for other symbolic relationships. The hierarchical character of symbolic associations allows us to establish roughly discrete stages in the process of learning them. At every level of recoding, the number of combinatorial possibilities increases geometrically. Rote learning is often required to discover the underlying symbolic logic at a given level, before moving to the next level (Deacon 1997a:95).[37] Deacon offers this characteristic of symbolic learning as a possible contributing factor to the "crudely stage-like pattern of children's cognitive development" proposed by Jean Piaget. "However," he cautions, "this punctuated pattern of symbolic conceptual development is a reflection of symbolic information processing and not an intrinsic feature of developing brains and minds" (97).

With the shift to symbolic associations, a kind of threshold effect is produced, where a mnemonic shift allows for the transition from associative learning (indexical reference) to categorical guesses among alternatives (symbolic reference). The result is a phenomenal increase in the rate that new lexical items are acquired. Such threshold effects can be seen in children's acquisition of language, with vocabulary growth and a diversification of the types of words being acquired. With the shift to symbolic reference, new grammatical classes or logical groups are discovered and these classes are filled with new lexical items. Hence, discovering the combinatorial rules of a symbol system allows for the discovery of its semantic features. The semantic features, in turn, indicate how the new symbols can be added to the system, without having to relearn any correlations (Deacon 1997a:98-99).

We are presented with a conception of symbolic associations as a hierarchic network of connected nodes that make up a semantic space that is continually changing (similar to the models of neural networks discussed in Chapter 4). One of the most important aspects of this

[37] Deacon (1997a:95) likens this process to a student learning a mathematical operation or a scientific concept by rote, before "getting" the underlying logic of the concept, which constitutes insight.

network of connected nodes is the hierarchies present in it and the constraints that come about as a result of the combinatorial patterns that make up the semantic space. Deacon criticises abstract theories of language that contemplate possible rules for combining unspecified tokens into strings, because such theories often assume that there is no constraint on theoretically possible combinatorial rule systems (Deacon 1997a:100). The symbolic use of tokens is constrained by the possible uses of both individual tokens and the tokens in terms of which they acquire meaning. Deacon holds that a form of regimented combinatorial organisation is a logical necessity for any system of symbolic reference. Without an explicit syntactic framework and an implicit interpretive mapping, it would be impossible to produce unambiguous symbolic information, or to acquire symbols in the first place. Syntactic structure can be seen as the higher-order combinatorial logic or grammar that enables and regulates symbolic reference. The implication of this realisation, in Deacon's own words, is that: "Syntax and semantics are deeply interdependent facets of language" (100).

5. A Universal Grammar After All?

As we have seen in chapter 2, Chomsky's very influential theory of Universal Grammar concludes that the only reasonable explanation for children's ability to acquire language is that it is an innate ability. Deacon emphatically rejects theories of Universal Grammar. However, he does believe that children are born with a certain predisposition to learn human language (Deacon 1997a:102). He holds that the fact that all normal children raised in a particular society will learn the language of that society, proves that human brains are equipped for or predisposed to the language function in some way. But he rejects Chomsky's proposed innate language competence to the extent that it requires knowledge of grammar to be a priori present in the human brain. He argues that this move only manages to relegate its unanswered questions to evolutionary biology and neuroscience, without presenting convincing evidence that such innate knowledge is even biologically possible, calling it an "argument from incredulity" (103). The ability of the human brain to create and learn symbolic systems and to acquire the complex rule systems and immense vocabulary at a very early age seems impossible to proponents of innate grammar, partly because the rules of grammar are simply too complex to be learned by any form of inductive learning and seem to defy attempts at formal description. Deacon argues that theories of innate grammar fail to provide neurological or evolutionary biological bases for their claims, and they ignore the feedback that children receive from the intricate social context in which they learn. He believes the idea of Universal Grammar makes unfounded and wide-ranging assumptions about brains and evolution, which he sets out to identify and counter.

But, Deacon does not go the usual route of critics of Universal Grammar. His argument is not that taking the social context of language-learning into account or that moving beyond an understanding of learning as induction will be enough to counter objections raised

by proponents of Universal Grammar. Such modifications would still not account for children's ability to acquire language so quickly and seemingly quite effortlessly. He asserts that the obvious gap between children and other highly intelligent species, and even sophisticated inductive-learning algorithms, when it comes to acquiring symbolic language cannot be ignored. It is also telling that children's other learning abilities are still relatively restricted when they learn language. From this he concludes that some kind of a priori preparation for language must be present (Derrida 1997a:105). The solution that he offers for this puzzle might be seen as quite surprising: he asserts that Chomsky inverts cause and effect in his analysis of the problem of language learning. The source of prior support for language acquisition does not lie inside the brain, but in language itself.

Deacon begins by taking the evolution of language into account.[38] His thesis is that languages evolved with respect to the human brain, and not according to arbitrary principles or constraints. A child faced with learning a language is aided by the fact that the language would have evolved to be intuitive, or "user-friendly", in a manner of speaking. Language, over time, adapted to people. In this regard, Deacon agrees with Chomsky that the child's knowledge of grammar and syntax is learned in a manner different from words. He even agrees that the rules that govern syntax and grammar are "discovered". But such discovery does not take place through introspection and applying the rules of grammar innate to the brain. In fact, learning by trial-and-error plays a significant role in acquiring syntax, but Deacon observes that children seem able to easily anticipate grammatical rules. In other words, there seems to be a relatively large proportion of *correct* trials in child language learning. There seems to be a strong bias in choosing ways to organise available words, and for Deacon, this points to an apparent ability in children to anticipate regularities in language. He speculates that language might be organised in such a manner that intuitive guesses are more likely to work. Hence, his argument is that language is adapted to "spontaneous" or natural human predispositions, in order to facilitate its acquisition (Derrida 1997a:107-108). "Children's minds need not innately embody language structures, if languages embody the predispositions of children's minds" (109).

Deacon is not suggesting that there exists a "designer" of language that tailors it to human beings' natural capabilities. He explicitly states that language "designs" itself: "Languages just change spontaneously over the course of many generations. Every effort to design a language has flopped" (Derrida 1997a:109). The key lies in the evolution of language, in that children are the vehicles that ensure that language is reproduced. Languages that are adapted to children's "natural assumptions" (biases) about learning, and how language works, have a better chance of being replicated through the generations.

Biological evolution is much slower and much more inflexible than language evolution (cf. Dawkins 1976). Changes in brain structure would take hundreds of thousands of years to become widely represented in the species. Deacon observes that brain structure has been

[38] The reader will recall that Chomsky is extremely sceptical about the ability of evolutionary theory to account for the language capabilities of humans.

notably conserved from the origins of vertebrates, while languages can become unrecognisable within a few thousand years. From this observation he concludes that we have been focussing on the wrong half of evolutionary equation when it comes to flexibility and adaptation.

5.1 AN EVOLUTIONARY UG

Languages are not designed to form logical systems.[39] This allows for the possibility that it might not be possible to assign utility or purpose to the principles of language, and we might not even be able to recognise their design principles. If language structure appears odd or illogical to us, it is because we are judging their design in terms of the wrong criteria or comparing them to inappropriate models. Deacon believes that the most basic design principle of language is not communicative utility, but reproduction (1997a:110). His theory should be distinguished from that of Dawkins, in that the reproduction at issue here is not only that of language, but also that of the users of language. Deacon (110) does view language as an organism (of sorts), however, and as such he believes that languages should be studied in evolutionary terms, rather than to model them as axiomatic rule systems:

> Instead of approximating an imaginary ideal of communicative power and efficiency, or following formulae derived from an alleged set of innate mental principles, language structures may simply reflect the selection pressures that have shaped their reproduction.

In order to be reproducible and reproduced, language needs to adapt in accordance with the structure of children's minds. Simply put, language structures that are more easily and quickly learned by children are more likely to be passed on to further generations and hence to be reproduced. Deacon is careful to emphasise the fact that the analogy of language as an evolving entity should not be taken too far (Derrida 1997a:112). Languages are not independent physical organisms, with intentionality and reproductive systems. But he does liken languages to viruses, in that viruses occupy the border between living and non-living systems – they lack reproductive organs, metabolic processes, and reproductive systems, but they do manage to evolve and adapt efficiently, through insinuating themselves into the cells that replicate them. Deacon describes the relationship between people and language as "symbiotic." This biological metaphor implies a two-way relationship, with language being "good" for people, and people being "good" for language. However, such a conception seems to perpetuate the anthropomorphic tendency of viewing language as an intentional (and conscious?) organism. A misplaced *agency* is implicitly accorded to language, and it might be more useful to think of languages as being "alive" in the sense of a complex system (not

[39] As Deacon (1997a:114) puts it: "Languages are abstractions. They are fuzzy collections of behaviours that happen to be reasonably well described by logical rule systems."

necessarily biological)[40] that is able to adapt to its environment, without the biological and intentional undertones.

In terms of his new model and inverted relationship between language structure and language learning, Deacon presents the following definition of language (1997a:112):

> Languages are inanimate artefacts, patterns of sounds and scribblings on clay and paper, that happen to get insinuated into the activities of human brains which replicate their parts, assemble them into systems and pass them on. The fact that the replicated information that constitutes a language is not organised into an animate being in no way excludes it from being an integrated adaptive entity evolving with respect to human hosts.

It is important to re-emphasise that the adaptation goes both ways – language adapts to human brains (specifically the brains of children), while human (children's) brains adapt to language. Languages are entirely dependent on humans, and it seems safe to say that humans are entirely dependent on language.

Deacon (1997a:115) does not present the "source" of language knowledge to be either in the child's brain, or in the "texts" from which the child learns language, but as highly distributed across the interactions between the child's learning and the evolution of the language community. He sees the process of language *change* as being the key to understanding the child's ability to learn language. This is because the processes that drive language change at a socio-cultural level are the same mechanisms that enable everyday language learning. Another link that he draws in terms of these mechanisms is the source of *language universals*. Deacon acknowledges the existence of grammatical universals, but does not concede that they are somehow hardwired into the brain. In fact, he argues that universal grammatical rules are not located anywhere and that they are not determined. Instead, he suggests that the implicit axioms of grammar have emerged independently in each evolving language, "in response to universal biases in the selections process affecting language transmission" (116). In other words, language universals are convergent features of language evolution.

All natural languages have a common adaptive context, which includes learning biases in children, human mnemonic and perceptual biases, constraints posed by human vocal and auditory abilities, and the requirements of a system of symbolic reference. It is because of these general structural constraints that languages evolve common structural properties, which leads to certain universals among different languages. Convergence is not

[40] Complexity theory, for example, would hold that language is alive in that it is an open system that is in dynamic interaction with its environment (language users) and which adapts over time, while still retaining traces of its history within the system (cf. Cilliers 1998:2-13; Auyang 1998:13; Gell-Mann 1994:17). Seen in these terms, it becomes clear that evolutionary development need not be restricted to phenomena that are biologically "alive" – many systemic phenomena that have the capacity for self-organisation exhibit adaptive behaviour in response to changing environments can be thought of as "living" systems.

limited to language *structure*, however. Word-reference can also be convergent, as Deacon demonstrates with regard to the evolution of colour terms in very different languages and societies (1997a:116).

Deacon sees the evolution of colour terms as an ideal example of the convergence of universals, because it combines the apparent arbitrariness of assigning a name to an entity and, what he terms "a well-understood neurobiological universal" (Deacon 1997a:116). Superficially, it would appear that assigning a colour term to a particular colour should be completely unconstrained – the name can comprise of any combination of human vocal sounds, generally, the human eye is able to see every gradation of colour between certain wavelengths of the spectrum, and any term can be assigned to any point on that spectrum. Interestingly, however, there are universal aspects to the colour terms that are assigned to light frequencies in the world's languages. Deacon does not take this to mean that colour terms are intrinsic to the brain or that the terms are derived from a single parent language. Firstly, there is a hierarchic sequence in which colours are (or aren't) included in a language (117). In languages with the least colour terms there are always at least three terms for colours: black (dark), white (light), and red. The next term included is always green. In languages with more terms, either yellow, or blue, or both are added. After these, the pattern becomes more unpredictable, but generally, additional colours are added in complementary pairs, fitting the prior colours already in the language. In languages with few colours terms, the terms that do exist are applied more broadly to accommodate more of the spectrum of visible light. In addition to this, when faced with a set of colour samples, the best exemplar of particular colour terms is generally agreed upon by people from around the world, regardless of the term they would use to describe that colour. The boundaries between the colours on the spectrum that are named may be fuzzy, but colour terms seem to have what Deacon calls a "category centre" (117). In other words, terms are not arbitrarily mapped onto any point of the colour spectrum visible to the human eye; colours are universally constrained.

The reason for this phenomenon is not too hard to find, when one considers what might be universal to human beings perceiving colour. Deacon defines colour as "a feature that is created by the brain as a means of maximising distinctive experiences of photons striking the cones of the retina in blended streams of different wavelengths" (1997a:118). He goes on to describe the "neural computational logic" of this process. The brain pits the signals from each of the three different types of cones in the retina against one another and derives a difference signal, which best corresponds with the colour discriminations we actually see. We do not see discrete colours, but this process is the source for the biases that underlie the universality of colour naming. It acts as a selective force, causing linguistic references over time to match our universal foci of perceptual experience.

Deacon illustrates his point with the thought experiment, based on a fictional new colour called "grue", which is used to describe the colour of a newly discovered type of

moss.[41] This new word originates in a language that has no other pre-existing colour term other than "red". The new term will begin with a reference that is idiosyncratic and specific to the new object (the moss). But, if the term persists in the language, it will, with time, evolve towards pre-existing biases in human colour conception. This is because, in the absence of a physical sample of the moss, memory of the precise shade of "grue" will be prone to error and imprecision. If there were no biases in human colour perception, the drift of the reference of the term will be random. But, as Deacon puts it, "errors in remembering and communicating the original reference of grue will tend to be biased toward the perceptually most salient and the pragmatically most contrasting opposite colour to red [as the only other colour recognised in our fictional language]: green." (1997a:119). He sees the biases that our brains introduce when we remember and reproduce references as analogous to the process of natural selection. Hence they will favour some variants and not others. The result is that the reference of our new term will evolve within the constraints posed by the human nervous system and the pragmatic constraints of language use. Deacon finds that grue will eventually evolve to refer to green, since there exists no term for green in our fictional language. If such a term did exist, grue would evolve to refer to blue. If terms for both green and blue already existed, it would refer to an intermediary between the two.

This example indicates how universal human neurological biases underlie the evolution of universal term references. Most importantly, for the purposes of this work, is that it indicates how such universals can be accounted for, without the necessity of falling back on some form of genetic pre-programming, or metaphysical entities in the manner of Plato's ideal realm. In Deacon's (1997a:120) words:

> In many ways, primary colour term reference relationships are as invariant and universal in character as is any aspect of the deep logic of human grammars, and yet it is clear that the colour terms themselves are not built into the brain, and that the choice of what colour to designate with a particular word is not dictated by any mental compulsion. What is universally built into the brain is rather a subtle set of perceptual biases that have none of the categorical and symbolic properties of words.

Even weak biases can produce social evolutionary consequences if they are continually present and Deacon goes so far as to state that the effects of such biases can appear to be completely determinate (Deacon 1997a:120). Stronger biases will form much stronger constraints on features such as attention, working memory, sound production, the automation of functions, and on the social selection of syntactic habits. Furthermore, biases from our primate heritage, dominated by visual and manual communication, must also form

[41] The term is borrowed from Nelson Goodman's (1955) discussion on the instability and uncertainty of word reference projection to the future. Whereas Goodman suggests that word reference is only stable due to social entrenchment, Deacon argues that projectability is possible.

strong universal biases in human communication. Deacon believes that it is this heritage that accounts for the fact that so many descriptive and explanatory devices in human language borrow extensively from visual and manipulative analogies, such as spatial metaphors, to refer to semantic dimensions.[42]

It becomes possible to account for language universals, without having to assume that the linguistic features must be hardwired into the human brain, or assuming that any linguistic features are necessary or predetermined. Common perceptual, behavioural, and emotional biases could translate into different languages sharing near universal conceptual patterns of linguistic representation. Hence, the argument has done away with the need for an innate and universal grammar, and replaced it with a social evolutionary process, which continues to evolve with time and further generations of language users. Deacon's point is that there are no "rules" of grammar, whether in the brain or in the language community, which need to be learned in order to grasp language. Language structures are the products of powerful evolutionary processes on many levels, and a set of basic design rules or principles would simply be redundant. Furthermore, innate mental tendencies are a fraction of the selection biases involved in the evolutionary process (Deacon 1997a:121-122). Thus Deacon addresses the paradoxical problems posed by universal grammarians – the objection that grammar is theoretically "unlearnable", the apparently inexplicable ability of children to acquire language and the often illogical rules of grammar and syntax – without falling back on a theory of innate language ability in humans. These "problems" can be understood as result of convergent social evolutionary trends and the parallel adaptations of language structures to the limitations and biases provided by children's brains: "Human children appear pre-adapted to guess the rules of syntax correctly, precisely because languages evolve so as to embody in their syntax the most frequently guessed patterns" (Deacon 1997a:122).

5.2 Learning from Neural Networks

Deacon goes on to discuss the implications of biases in human neural architecture[43] and evolutionary trends for the rapid language acquisition abilities of children.[44] As we have seen it is precisely these abilities that lead theorists like Chomsky to propose the existence of some sort of language acquisition device in children, which is thought to be genetically hardwired into the brain. We have also seen that Deacon proposes that the grammatical and syntactic structures of most (successful) languages have been selected through evolution to compliment the constraints already existing in children's brains. Deacon's suggestion is that the universal biases in the human brain facilitate the language learning process, which implies that children *do* acquire language through learning. Given the extremely complex and vast

[42] Pinker makes a similar argument (cf. 1997:352-362).

[43] Cf. Chapter 4 for a discussion of neural networks.

[44] It seems that there is relative consensus among theorists that some form of pre-adaptation occurred in the hominid lineage that allowed for our language abilities (Christiansen and Kirby 2003:301).

language structure that needs to be acquired, the question arises as to what such a learning process would look like. "Plato's Problem", as formulated by Chomsky, seems relatively insurmountable.

Deacon launches his offensive on Chomskyan scepticism with a discussion of the language abilities of Kanzi, the pygmy chimpanzee, who has demonstrated the most advanced symbolic capabilities exhibited by any non-human animal (1997a:124). Kanzi seems to have learned to understand speech and communicate symbolically (as opposed to indexically) by clinging to his mother while she was being trained to use language. While his mother proved to be a poor learner, Kanzi became very adept at language tasks. Deacon suggests that Kanzi's success was due to the fact that he was immature at the time that he was exposed to language training (1997a:125). He further suggests that the case of Kanzi points to the possibility that there exists a critical period in language acquisition, where a "language-learning mechanism" is activated at a certain developmental stage (126).[45] The fact that Kanzi exhibited a critical period for language acquisition, would furthermore suggest that learning language is facilitated by immaturity in general, rather than by a specifically human characteristic. Just like young children, young chimps have poor memories for details, and they are easily distracted and easily bored. Deacon notes that these apparent impediments to learning seem to put children and young chimps at an advantage when it comes to learning language. Furthermore, and crucially, children and young chimps seem to have a better sense of which stimulants and context parameters to pay attention to than older chimps and children. "Like a child attuned to the phonemes of the local language, Kanzi's whole orientation seems to have been biased by this early experience" (126). An important difference between children and Kanzi, of course, is that chimps do not learn a language in the wild (at least, not in the sense that Deacon uses the concept, i.e. "symbolic communication"). Hence, there is no reason to suppose that chimpanzees would have evolved a language-learning mechanism, or that Kanzi should have a critical period for acquiring language. As Deacon states: "Kanzi's example throws a monkey-wrench into the whole critical period argument as it applies to language acquisition in children" (126). Is it possible that there is something about infancy in general that aids language acquisition?

Generally, theories that hold that because of its complex structure natural grammar should be nearly impossible to learn, understand "learning" in terms of logical induction. In such a scenario, a learner is faced with examples of grammatical sentences, and he or she then derives general grammatical rules from their similarities.[46] "The hypothesized rules are thus only constrained by the relationships implicit in the input strings" (Deacon 1997a:127). Theorists argue that it is almost impossible to inductively derive the rules of a language from the finite set of grammatical sentences that the learner would be exposed to. Deacon argues that this objection is aimed at a logical problem that is not analogous to the pragmatic

[45] See Komarova and Novak (2001) where they maintain that the language acquisition period in humans is the first 13 years.

[46] See Pinker (1997:262-296) for example.

problems that real language users face (128). According to him, inductively deriving correct grammatical rules from a set of example sentences is not the only possible way of ascertaining which sentences will be unambiguous and which will not. He argues that there are many learning processes, and that successful learning depends to a large extent on matching the appropriate process to the structure of the patterns that need to be learned. We have already seen that Deacon holds that symbolic relationships, given their structure, are difficult to learn. He argues that similar difficulties are posed by learning the logic of grammar and syntax, especially because "these facets of language are also surface expressions of the *deep web of symbolic relationships* [my emphasis]" (Deacon 1997a:128). As already mentioned, it seems that the limitations to children's learning abilities – difficulty in conscious memorisation of associations, an inability to concentrate for long periods of time, the brief span of their working memory – may be constraints and biases in learning that are advantageous when it comes to language learning (129).

As we have seen, Deacon defines language, as a system of associations with a deep logic that derives from the indirect systemic logic of symbolic reference. It is highly distributed and non-local and the rules for implementing these relationships tend to form hierarchic patterns. Taking these characteristics into consideration, Deacon contends that the way in which children learn language is akin to the way in which neural networks are trained by cognitive scientists. In other words, when it comes to language, learning is not the "top-down" approach of inductive logic, but something like the pattern recognition abilities such as those exhibited by neural networks.

As we have seen, neural networks are computational models, inspired by an idealised conception of the architecture and interaction of neurons in the brain. They have their origin in mathematical models of neurons, developed by McCulloch and Pitts.[47] These models are variously known as neural networks, distributed processors, or connectionist models (Cilliers 1998:16). When using neural networks as models, theorists borrow design features that are thought to be utilised by information processes in the brain. Functionally, the brain consists of a large network of interconnected neurons, and very crudely, neurons are conceived of as operating by converting inputs to outputs, by means of a form of Boolean logic.[48] Deacon makes use of research done by cognitive scientists into the language-learning

[47] See Chapter 4.

[48] Kauffman (1995:71-79) discusses Boolean networks by means of the analogy of a network of light bulbs that are randomly wired together with an electrical circuit. Each bulb can only have two values, on and off. The bulbs are interconnected so that in the following state, each bulb reacts to the input from the current state of all of the bulbs that it is connected to. Each bulb can be connected to any number of others. As the network proceeds though a series of states, each bulb (node) will react examine input from bulbs connected to it and then switch on or off according to the rules for reacting to other signals assigned to it. The process is repeated again and again. A system that consists of a finite set of bulbs will have a finite set of possible states. Once it is started off, it will run through a sequence of possible states, also called a trajectory. Within a finite system, the system will eventually hit upon a state that it has previously encountered, and it will run through the associated trajectory again. Our system is deterministic in the sense that the behaviour (on/off) of the light bulbs is constrained by the input from the other bulbs that they are connected to and ordered in terms of the trajectory that it settles into.

abilities in neural networks as an indication of the effects that learning biases can have on language learning.

Language can be characterised as a distributed pattern recognition problem, which requires a learning approach which is able to focus learning at "the right" level of analysis. Deacon argues that starting small and simple, with a learning process incapable of tackling the whole problem, might offer this sort of learning constraint" (Deacon 1997a:129). Neural networks can be trained to solve simple language tasks because of their pattern recognition abilities. It is extremely difficult to programme digital computers to accomplish this feat. Deacon (131ff) argues that the discrepancy in abilities between digital computers and neural networks comes down to the fact that computers are programmed using "top-down" approaches with step-by-step instructions for analysis, while networks are trained to "categorise" sets of stimuli and to respond to novel sets that are similar to those on which they were trained. Neural nets are capable of a "stimulus generalisation" of sorts. Deacon (1997a:133) explains:

> A fully trained net that could make correct predictions for novel sentences (analogous to stimulus generalisation) would necessarily embody aspects of the statistical structure of English grammar and syntax in its architecture, even though it would not contain any semantic information. If a net could be trained to do this as well as a person, it could indicate two things: (1) that the statistics of the relationships between grammatical classes of words in the training strings contained sufficient structure from which to recover grammatical regularities; and (2) that the regularities were learnable in some form without explicit (rule-based) error correction.

Research in neural networks would thus suggest that the most likely method in terms of which language is learned is one of "stimulus generalisation". However, work on neural nets further suggests that such "complex mapping problems" which are solved in incremental stages, run the risk of "fooling" neural nets into overlooking global information patterns in favour of the more immediately apparent, local patterns on which they are trained.[49] In order to overcome this tendency inputs into the network need to be recoded "in such a way as to reduce the distributed character of the associations" (Deacon 1997a:133). This is done by introducing biases into the networks against features that are common to suboptimal solutions that overlook more optimal, global solution to input problems. Thus, staged training, starting with a set of simple sentences and followed by sets of more complex sentences produces nets that are eventually able to employ complex syntax. Furthermore, by randomly disrupting

---

Neural systems are conceived of as working along similar lines as Kauffman's system, with neurons being substituted for the light bulbs.

[49] See Deacon (1997a:122-142) for a more detailed discussion of the problem-solving abilities of neural networks and their implications for our conception of language-learning.

recurrent inputs during early training, more global regularities of the input become more redundantly represented throughout the net and thus form biases that favour large-scale statistics of word associations (Deacon 1997a:34). Hence, neural networks demonstrate that it is (135):

> …possible to design a device that could learn to predict grammatically correct sentence structures in a purely inductive fashion, given nothing more than a corpus of positive examples of allowable texts – exactly what the UG theorists had said was impossible. The key to accomplishing what many had "proved" was impossible was to structure the learning process differently at different stages of learning. What was available for learning at early stages was "filtered" (either by the training sequence or the competence of the net), so that only some aspects of the input were available at any time. This set up learning biases that just happened to correspond with crucial structural features of the problem space. Although the information embodied in this incremental learning strategy was extrinsic to the language data presented to the net, it was far less explicit than a Universal Grammar, or indeed any specific grammatical information. There is no difficulty imagining how such generic constraints could happen to be available to support the learning of grammatical speech…it [shows] how important the learning structure is to what can and cannot be learned. More important, it suggests that the structure of language and the way it must be learned are linked. What may be essentially inaccessible under quite general learning conditions may become more accessible under more limited conditions.

Deacon suggests that work on neural networks indicates why children (and Kanzi) are adept language learners and why a critical language learning period exists. Children are unable to remember the details of specific word associations they are slow to map words to objects that co-occur in the same context, they only remember very global structure-function relationships of sentences, and find it difficult to hold more than a few words in short-term memory at a time (Deacon 1997a:135). The suggestion is that these constraints on their learning abilities during their early years are actually advantageous for children when learning language. Just like neural networks that have been presented with biases, children with their limited attention spans and memories for associations and details would disregard immediate (local) information by default, and would thus attend to more general (global) patterns in the sentences that they are being presented with. In Deacons words, "the relevant large-scale logic of language ‘pops out’ of a background of other details too variable for them to follow, and paradoxically gives them a biased head start" (135). Furthermore, symbolic associations, which are quintessential examples of highly distributed relationships, would be acquired in a

similar manner. It requires postponing commitment to immediately obvious associations, until less obvious distributed relationships are acquired (Deacon 1997a:136):

> Only by shifting attention away from the details of word-object relationships is one likely to notice the existence of superordinate patterns of combinatorial relationships between symbols, and only if these are sufficiently salient is one likely to recognise the buried logic of indirect correlations and shift from a direct indexical mnemonic strategy to an indirect symbolic one.

Deacon's position is that grammar and syntax both inherit the constraints implicit in the logic of symbol-symbol relationships (Deacon 1997a:36). Furthermore, syntactic regularities are essential to discovering the combinatorial logic that underlies symbols. Therefore, acquiring the shift to symbolic mnemonic strategies and acquiring grammar are intertwined learning problems, which are more easily surmounted by learners who "get" the bigger picture, while losing track of the details. Deacon concludes that it is unsurprising that the optimal time for learning grammatical and syntactic regularities in language coincides with the period when symbolic reference is first discovered. Immature brains are an advantage in the first steps toward overcoming the complex learning problem of shifting from the indexical to the symbolic referential strategy. As brains mature, this advantage falls away, but brains become more adept at acquiring the complex details of vast symbolic systems.

Deacon suggests that language structures have evolved to take advantage of the learning biases in younger children, in that those languages that are more easily acquired at a younger age would have a greater chance of being replicated. On the other hand, he argues that the advantages conferred by language would have caused the human brain, in turn, to evolve in such a way so as to further facilitate the acquisition and use of language in humans. His argument is not only that immaturity provides biases that enable language acquisition, but that human brains would furthermore have evolved biases that facilitate language learning to such an extent that virtually all human beings are able to acquire at least a modicum of language.

Thus, instead of an innate language organ or universal grammar in the form of a "hopeful monster", Deacon puts forward universal biases that exist in the human brain by virtue of natural selection. He thus accounts for Peirce's concept of *abduction*, without having to postulate grammatical knowledge that is innate in that it is genetically included.

6. Aspects of Language and Evolution – Contra Peirce and Chomsky

Unfortunately, restricted space prohibits us from examining Deacons' analysis of the structure of the brain in terms of the influence of language in any detail. He devotes an extensive part of *Symbolic Species* (1997a:145-320) to aspects of the human brain that seem

to have evolved to support language.[50] He shows that the structure of the human brain has been significantly influenced over the course of its evolution by the requirements posed by language. Deacon argues that Kanzi would probably not progress beyond the level of symbolic sophistication of a three-year-old human child (Deacon 1997a:255), because the fundamental constraints on language evolution are learning and memory difficulties, and not limited motor abilities. As a pygmy chimp, Kanzi would not possess the brain-organisation necessary to surmount the symbolic-learning threshold. "Symbol acquisition abilities provide the pacemaker for language evolution from which other adaptations for language must derive their usefulness" (255).

Deacon argues that the use of symbolic reference caused changes in the brain over the course of evolution. His argument is one of Baldwinian evolution, where human behaviour served to amplify and bias the process of natural selection by modifying the context in which natural selection takes place (Deacon 1997a:322). Thus, human beings changed the environment in which selection takes place through adopting the use of symbolic representation, which created a new environmental niche which presented a new set of selection pressures for their offspring. If such altered conditions remain consistent, its selection pressures could eventually lead to favoured genetic characteristics becoming prevalent throughout a species. The argument is not that learned adaptations are replaced in a one-to-one basis with instinctual counterparts[51], but that the new developmental constraints could cause already existing genetic biases to be amplified or lessened (324).[52] Hence, the argument is not that evolution produced a "language organ" in the genes, but that it recruited various regions and abilities of the hominid brain so that they can support symbolic reference. Eventually, these initial symbolic abilities would have been applied in more aspects of human life and would have become more complex, hence increasing selection pressures for brain structures that can support language functions. Hence, language and the brain exerted

---

[50] Also see Deacon (1997b:337-357) and Deacon (2004:288-290).

[51] This seems to be Pinker's position (see 1994:242-243, 362-369; 1997).

[52] To quote Deacon (1997a:326) in this regard:

> Unlike the analogue to Lamarckian inheritance of traits acquired by habit, where adaptive responses are presumed to be mapped one-to-one from outside to inside the genome, any traits that ease the costs imposed by behavioural adaptation to new conditions will be selectively favoured by Baldwinian selection and thereby subject to genetic assimilation… Inevitably, a given adaptive response can be biased in its probability of appearance during development in a very large number of diverse ways, and all may be simultaneously modified by Baldwinian selection… Of all the forms of adaptation, the flexibility to learn new behavioural responses during one's lifetime can produce the most rapid and radical evolutionary consequences. Indeed, the ability to learn and thus inherit acquired behaviours may be one of the most powerful sources of evolutionary change. It provides an organism with access to a repertoire of potential adaptations, and so amplifies and extends the range of the behavioural predispositions that can be "sampled" by natural selection. A learned behavioural response can become genetically assimilated to become a behavioural predisposition by virtue of the costs that it imposes on the organism. There are costs in terms of learning time, costs for failing to learn or learning incorrectly, and costs for simply being inefficient. Individuals who, for any number of reasons, learn more quickly and reliably and implement this behaviour more efficiently will benefit in reproductive terms.

reciprocal evolutionary pressures on one another. Languages whose structures ran counter to the supporting structures of the human brain will simply have ceased to exist. However, it is important to note that because the rate of adaptation is so much more rapid for languages than it is for human brains, languages have adapted much more than brains have. The result is that only the most invariant and general features of language would exist long enough to influence brain evolution (Deacon 1997a:329). Thus, Deacon concludes that there is little possibility of mental adaptations to specific syntactic structures. Only those features of language that are common to essentially every language, and which pose consistent and invariant demands on neural processes, would contribute to brain structure. Hence: "the key requirement for genetic assimilation is the existence of some invariant sensorimotor features or invariant mnemonic features of the adaptation" (330).

Deacon's further argument is that the discontinuity between stimulus associations and symbolic reference associations makes it *impossible to assimilate symbols genetically* (Deacon 1997a:332). We have seen that symbolic associations are essentially arbitrary with respect to various language communities and hence do not present features that are invariant for periods of time long enough to influence brain structure. Conversely, some animal calls seem to be fixed indexical referential signals that do stay invariant long enough to be genetically assimilated (such calls would include laughter and sobbing in the case of humans). Similarly, grammatical functions do not exhibit neural computational invariance (332-333). Deacon argues that the same grammatical operations can be represented by very different brain regions, and that the majority of the deep structures of grammar that are considered to be candidates for UG are the least likely features of language that would have evolved neural supports. He concludes that if there are innate rules of grammar, they could not have gotten there by genetic assimilation through Baldwinian processes, but only through some form of "miraculous accident" (333). Scientifically speaking, however, recourse to such a "hopeful monster" theory is unacceptable, especially given that there may be a scientifically legitimate alternative explanation available. Deacon's point of departure is to identify an aspect of language that is ubiquitous, namely, the unprecedented cognitive computational demands posed by symbolic reference (334-335). It is also likely that a predisposition evolved to attend more closely to some stimulus features or biases toward certain realms of learning (337-338). He is emphatic, however, that "[n]o innate rules, no innate general principles, no innate symbolic categories can be built in by evolution… The noncorrelative nature of symbolic reference has cut language off from forces that shape biological evolution, and instead has shifted the burden of adaptation to a new level of information transmission" (339).

He (Deacon 1997a:266) finds that widely distributed neural systems contribute in a coordinated manner to create and interpret symbolic relationships. Among these systems, the prefrontal cortex seems to be involved in the construction of novel symbolic relationships, including those involved in language learning. Deacon believes that the prefrontal cortex is crucial for making the shift from indexical associations to symbolic associations (270). Deacon

is emphatic that language functions should not be seen as "mapping onto" brain functions – there may be a number of different ways to achieve the same language function:

> I suggest that we need to think of language functions in [terms of the requirements of the functional goal], especially those that are imagined to be "deep" to the surface structure of speech. We need to treat them as composite behavioural products or logically defined outcomes, as opposed to neural operations. There need not be any specific association between a brain region or connection and a class of linguistic operations, and there may even be many alternative neurological means of achieving the same symbolic end. This does not assume that the human brain lacks local specialisations for language, or that language abilities can be described in terms of general learning mechanisms alone. And it does not depend on any denial of area specialisation within the cerebral cortex. It implies only that the neural distribution of language functions need not parallel a linguistic analysis of those same functions.

Areas in the brain that have to do with language were recruited by the principle of natural selection from structures that originally evolved for very different functions. This, coupled with the fact that language is itself subject to evolutionary demands quite different from those on the brain, indicates that the brain has been designed according to a logic very different than that of language as one of its cognitive performances or behaviours (Deacon 1997a:287-288). This observation by Deacon presents another stumbling block for the Chomskyan project of deducing certain "facts" about the "language organ" from studying the structure of grammar. Hence, not only does Deacon indicate how language universals and language acquisition can be possible without the existence of an innate universal grammar, he demonstrates the unfeasibility of the methodological assumptions on which Chomsky's project is based. And he manages to do this in terms of a methodological approach that Chomsky is extremely sceptical of – viewing language as the product of evolution by means of natural selection. This insight also provides important constraints for theories that are quick to see a parallel between natural language and "mentalese" or natural language and symbol manipulation in digital computers, especially given that the structure of "mentalese" is rarely speculated upon.[53] It is worth quoting Deacon (293) in this regard:

> Once we abandon the reification of language areas as modular language algorithm computers plugged into an otherwise non-linguistic brain, it becomes evident that language functions may be widely distributed and processed simultaneously in many places at once. They may also be distributed according to a computational logic that is not necessarily obvious from the eternal speech

[53] An exception is Pinker – see Pinker (1997), for example.

signal. Our experience of the linearity of speech belies an internal parallelism and hierarchy of the underlying cognitive processes. Indeed, one possible interpretation of the hierarchic logic of grammatical and syntactic operations is that they evolved to accommodate this mismatch between cognitive processes and production constraints. This is important information to keep in mind for linguistic analysis.

7. DEACON'S THEORY

In terms of his analysis of reference into three modalities (iconic-indexical-symbolic), Deacon gives his own account of how information may be represented in the brain. His theory is highly speculative and contingent upon a great deal of further research. Its relevance to the current discussion lies in its illustrative value – here we have a model of a possible way that language manifests in the brain which is based on neurological research, and which is willing to forgo many of the favoured philosophical assumptions that constrain more abstract theorising on the subject. Deacon manages to fulfil Peirce's wish to root philosophy within scientific developments while using Peirce's own semiotic in the process. Ironically, Deacon seems more willing than Peirce to relinquish some philosophical assumptions (such as Peirce's desire to reserve a place for God in his semiotic) and as such formulates a theory that does not veer off into the realm of the metaphysical as Peirce's does.

In terms of the manifestation of symbolic reference in the brain, Deacon (1997a:300) argues that brain structures necessary for the analysis of symbols are distributed across many areas of the brain. However, the systemic character of symbolic reference would suggest that, even though the representation of symbolic associations is distributed, similar classes of words ought to share neural commonalities. Symbolic referential relationships would be produced by the convergence of different neural codes from various brain systems. Deacon argues that word comprehension and retrieval processes, as symbols, are the results of combinations of simpler associative processes (indices). Hence, "as each supportive representational relationship is brought into play in the process of producing or comprehending a word, each corresponding neural substrate is activated in different phases of the process" (Deacon 1997a:300).

Seeing that the symbolic representational relationship is hierarchic and is constructed from lower levels of representation (from icon to index to symbol), Deacon speculates that their neural representation would have a hierarchic structure as well. As the lowest level of representation, iconic relationships would map onto processes with single sensory modalities (Deacon 1997a:300). Similarities between phonemes or visual experiences in different contexts would be examples of such iconic relationships. Deacon speculates that such simple iconic processes would probably be highly localised, and perhaps represented in the cerebral cortex. Whereas he believes that words and familiar objects might require compound iconic analysis involving more than one sensory dimension or modality, he argues that in most

practical contexts only a few features would need to be assessed in order to recognise an icon, "because the competing alternatives are limited" (301). The link between icon features is indexical (see chapter 1):

> The mental representation of a complex object is based on a correlation of numerous iconic associations in different dimensions or modalities which predict one another's presence. The more complicated the object or relationship, the more numerous the iconic and indexical assessments that are required to recognise it. In this regard, there is only a difference in numbers and diversity of icons involved between learning to recognise objects, on the one hand, and learning to recognise relationships between objects or between events, on the other. Recognition and prediction processes can be streamlined as we learn to focus sensory analysis on only the assessment of the most relevant features.

Indexical associations are associations between words and objects and Deacon believes these associations to involve cross-modality relationships, such as sound, vision, etc. Thus, word analysis would be highly distributed across the brain and cannot be limited to a single modality. As we have seen, symbolic reference develops when i) a set of indexical associations is established between signs and objects in experience, ii) a systematic set of indexical associations between signs is established in terms of "logical alternation and substitution correlates", iii) when a correspondence is noticed between the sign-sign system and implicit relationships between the objects to which the signs refer (Deacon 1997a:301). When this happens a referential shortcut is created where the indexical association can be bypassed and the relationships implicit in the signs can be used to refer to relationships between objects and events. Concrete indexical associations are unlearned and the indexical (and iconic) relationships are recoded into symbolic relationships, which creates a mnemonic shortcut that allows information transmission and reception to be accelerated and compressed (302). The ignored indexical relationships still implicitly ground word reference, and can be invoked if necessary. Neurologically speaking, Deacon speculates that there may be "convergence" zones within the cerebral cortex where sensorimotor traces of experiences might be interlinked, hence, where images and experiences may be linked with particular word sounds. These associations may provide the indexical associations of words (302). However, he argues that their symbolic association – meaning – involves "something more" (303). This something more would include the associative relationships between words and the logic of how they map onto indexical relationships. Deacon concludes (309):

> The symbolic functions, the grammatical and representational relationships, are not processed in any one place in the brain, but arise as a collective result

of processes distributed widely within the brain, as well as within the wider social community itself. Virtual reference is only virtually realised.

Deacon's makes a convincing case for a hypothesis that there is no dedicated language module or faculty in the brain, with crucial implications for theories on both language and mind/brain. As we have already seen, the existence of an innate universal grammar is not only very unlikely, but also unnecessary, given that language universals can arise out of the co-evolution of both language and the brain. Furthermore, the apparently distributed character of language throughout various brain regions minimises the possibility that the structure of language functions in the brain can be directly deduced from the structure of language itself. It also minimises the possibility that language processing in the brain can be likened to the information processing operations of serial digital computers. Language structure itself cannot be expected to follow any "externally" determined logical principles, in that the logic according to which it is structured is a combination of the requirements of language learners and the gradually modified structures of an ape brain, that may originally have evolved to support very different functions.[54] Deacon's case is convincing because of his point of departure – the theory of evolution by means of natural selection. Such an approach has proved to be an extremely successful explanatory principle at various levels of biological explanation. There does not appear to be any salient reason, other than various metaphysical commitments, that human beings – as biological organisms – should be approached in terms of any other theoretical assumptions. It is also clear that language and the mind are fundamentally entwined and have to be studied in conjunction with one another.

An extremely obscure aspect of the co-evolution of language and mind is the origin of language. What prompted hominids to cross the symbolic threshold? Deacon argues that many of the scenarios posed as sources of the origin of language (group living, sharing information, organising hunts, etc.) are not mutually exclusive alternatives of why language originated, but are rather a list of domains to which symbolic communication has successfully been introduced (Deacon 1997a:350). Furthermore, the value of each of these uses would have contributed selection pressures for the elaboration of symbolic abilities. Inevitably, explaining symbolic origins is a highly speculative exercise; the co-evolutionary adaptations between the brain and language are extremely complex and with obscure cause-and-effect relationships. Furthermore, there is very little surviving archaeological evidence that can plausibly shed light on the process (see Deacon 1997a:365-375). It is entirely possible that the scenario in terms of which the symbolic threshold was crossed will never be reconstructed. Deacon presents his own vision of the process, which, while it may very well be inaccurate, is edifying.

Deacon (1997a:378) believes that the first symbolic systems were not full-blown languages. Much social communication probably took place through call-and-display

[54] In Deacon's (1997a:326) words: "Evolution seldom follows straight lines."

behaviours such as those of some modern primates. As we have seen, learning to use symbols requires learning a system of symbols, which is useless until the whole system of associations is acquired. Thus Deacon argues that the shift to symbolic communication must have conferred a significant advantage on those who acquired it in order to offset the cost of the acquisition (379). He also argues that such expensive adaptations usually occur in the context of intense sexual selection, as can be seen with the cumbersome tails of mail peacocks, for example.

Deacon notes that *Homo sapiens* have a highly unusual mating pattern and speculates that this singular set-up led to the evolution of a singular trait – symbolic communication (Deacon 1997a:384-385). In the case of *Homo sapiens* both males and females contribute to the rearing of offspring, adult humans are bound by long-term and exclusive sexual access and exclusion rights to particular individuals, and such bonded couples continue to live in large cooperative social groups that include many unrelated males and females (385). Such a setup is highly unusual in a world where in the case of other pair-bonded mates, exclusive sexual access is enforced through isolation.

Another feature that also stands out in our prehistoric ancestors, and which is unusual for primates, is the advent of the use of meat within initially foraging societies. The appearance of stone tools from approximately 2.5 million years ago presumably corresponds with a shift from foraging behaviour to scavenging (and later hunting) for meat. Deacon (Deacon 1997a:386) argues that this novel food source placed immense pressure on the social group organisation of our ancestors. Whereas women in foraging societies provide an equivalent amount of calories to the men, the situation changes when meat becomes the primary food source as the group. Meat is more difficult to acquire than nuts and fruit, etc. and a pregnant or nursing woman would be unable to provide amounts of meat equal to those of the unencumbered men. Hence, men have access to a resource vital to incapacitated women, and children. The resultant dependency on men for an essential resource, and the presence of other fertile men and women within the social group complicate group dynamics significantly. Whereas women need a guarantee that the father of her offspring will not abandon her for another women and continue to provide the nutrients necessary for keeping her and her offspring alive, men need a guarantee that the offspring that he is providing for are indeed his.

Evolutionary speaking, potential and significant conflicts of interests arise with regard to procreation, and both men and women need to ensure that their investment of limited time, energy, and resources will indeed lead to viable and legitimate offspring, which ensures that their genes are procreated. Furthermore, the cooperation of other members of other relatives and friends becomes necessary for acquiring meat and ensuring that the vulnerable members of the group receive it (Deacon 1997a:387). Large, cooperative, mixed-sex social groups with significant male contribution to the care of offspring produces requirements that are highly unusual within the natural world. Deacon argues that this pattern of social-sexual organisation tends to undermine itself in the course of evolution because it produces a highly volatile social

structure that is susceptible to disintegration (388). The only case where group living and male provisioning seem to occur successfully is where limited and unambiguous sexual access is maintained, such as happens in the case of social carnivores (392).

Deacon notes that the shift from polygyny to pair-bonding with *Homo erectus* and *Homo habilis* coincides with the first development of stone tool technologies and the first increase in relative brain size (Deacon 1997a:392-396). He argues that distinctive social arrangements in terms of specialised social communication needed to accompany these changes in order to insure a viable and stable reproductive strategy. The strategy that arose was that of solving the problem symbolically (397).

Whereas indexical communication can only refer to things and events by virtue of a concrete link with it, it cannot refer to abstract and intangible objects. Inevitably, referring to a reproductive strategy would require reference to abstract concepts that are meant to ensure reciprocity in specific social relationships. Included among the things that need to be referred to are: identification of specific individuals and remembering their past behaviours, the ability to detect cheaters who do not reciprocate, devising and implementing punishment for cheaters, identifying which individuals are available as mates and which are not, terms for sharing resources must be set, current and future obligations must be set out, etc (Deacon 1997a:398-399). All of this information can only be represented symbolically. Hence, Deacon suggests that the original impetus for crossing the symbolic threshold was, in effect, the marriage agreement (400).[55]

Deacon argues that the example of Kanzi and other primates that have acquired some symbolic capability shows that there need not have been anything intrinsic to the human brain that gave rise to the use of symbols (1997a:402). These examples also indicate that an extensive external support network (akin to the chimp trainers) would have been necessary to support the initial shift to symbolic representation of the essentially ape-brains of our ancestors. He speculates that the advent of a symbolic culture probably progressed in fits and starts with many trials and errors before a context was created in which symbolisation

---

[55] To quote Deacon (1997a:400-401) directly:

> Marriage contracts establish both vertical lineal symbolic relationships and horizontal affinal symbolic relationships. Marriage, in all its incredible variety is the regulation of reproductive relationships by symbolic means, and is essentially universal in human societies. It is pre-eminently a symbolic relationship, and owing to the lack of symbolic abilities, it is totally absent in the rest of the animal kingdom. What I am suggesting here is that a related form of regulation of reproductive relationships by symbolic means was essential for early hominids to take advantage of a hunting-provisioning subsistence strategy… But without symbols that refer publicly and unambiguously to certain abstract social relationships and their future extension, including reciprocal obligations and prohibitions, hominids could not have taken advantage of the critical resource available to habitual hunters. The need to mark these reciprocally altruistic (and reciprocally selfish) relationships arose as an adaptation to the extreme evolutionary instability of the combination of group hunting/scavenging and male provisioning of mates and offspring. This was the question for which symbolisation was the only viable answer. Symbolic culture was a response to a reproductive problem that only symbols could solve: the imperative of representing a social contract.

could be supported. Furthermore, he suggests that the necessary contextual support would have been made available by ritual (402). Ritual would have shifted the attention away from concrete and immediate indexical associations to the abstract relations between a set of created signs. Deacon argues that ritual in fact still plays this role within the contemporary context of symbol-learning, although in a much less obvious way than in his hypothetical scenario. Hence, in repeating the same set of actions with the same set of objects over and over again it becomes possible to make the transmission from explicit sign-object associations to implicit sign-sign association (Deacon 1997a:402-403). Rituals are still employed to define abstract social relationships and establish habitual thoughts and actions. Admittedly, simple and even crude symbolic systems can serve the purposes of establishing a social contract such as a marriage agreement, and it is probable that the advantages posed by symbolic representation were appropriated by various other aspects of human life, hence providing selection pressures for an ever-increasing upward trend in the complexity of our symbolic abilities.

To sum up the argument (Deacon 1997a:406-407):

> …demonstrating that [social] relationships exist and providing some way of marking them for future reference so that they can be invoked and enforced demand the explicit presentation of supportive indices, not just from reproductive partners but from all significant kin and group members… The symbol construction that occurs in these ceremonies is not just a matter of demonstrating certain symbolic relationships, but actually involves the use of individuals and actions as symbol tokens. Social roles are redefined and individuals are explicitly assigned to them… As with all symbolic relationships, social roles are defined in the context of a logically complete system of transformations; and because of this all members of a social group (as well as any potential others from the outside) are assigned an implicit symbolic relationship when any one member changes status… Tokens that served an indexical purpose within the ritual symbol construction become symbolic because of it. Rings, ceremonial clothing, and ritual scarring are indices of having participated in a symbolic transmission, and thereby can become symbols of the same relationship. Giving the universality of human marriage and the embedded ritual and ceremony, it seems reasonable to imagine that hints of the original logical structure of the first symbolic rituals still echo in the marriages that men and women make today.

Hence, the earliest forms of symbolic communication need not have had anything to do with language as we know it today. The details of modern language probably derived from other, and very divergent evolutionary causes (1997a:409). The crucial step for establishing the linguistic ability, however, was the ability to think symbolically, which over time led to

changes in brain physiognomy in order to support this novel and infinitely beneficial evolutionary anomaly. It is important to highlight an implication of this co-evolutionary thesis of mind/brain and language, namely that the evolution of language most likely took place over a distributed "design space" which cannot be seen to be internal or external to the brain – it took place at the interface where Baldwinian evolutionary processes ensure that cultural evolution can affect biological evolution.

8. CONCLUSION: THE SYMBOLIC MIND

At the end of his marathon exposition Deacon devotes some space to a number of philosophical implications that his theory might have. The implications are vast and far-reaching and merit at least another volume dedicated to them exclusively. Important to note for our purposes is that Deacon's co-evolutionary theory sheds some light on why questions of mind and language seem to be inextricably intertwined, both in our common-sense intuition and in more formal theorising on the matter. If Deacon (and other evolutionary theorists) is indeed correct in his co-evolutionary hypothesis, the human mind as we know it today is the result of our ancestors hitting upon a novel evolutionary niche and "crossing the symbolic threshold". The "discovery" of symbolic reference gave rise to a novel evolutionary context to which our initially ape-like brains adapted; so much so, that we seem to have a fundamentally symbolic cognitive disposition. Indeed, Deacon argues that most of the aspects of the human "mind" that we consider to be different from "mere" animal sentience or "mere" brain processes, are those processes in the human brain that make use of symbolic reference. Our access to symbols allows us to "represent" the abstract, the hypothetical, the implicit, the realm of other minds, and the realm of our own mind and the self (Deacon 1997a:423-432). These abilities are not inborn, but are acquired in the process of acquiring culture and language, and thus are inherently social human traits. Symbolic representation enables us to use the iconic and indexical representations in our own memories to reconstruct the experiences of others through the mediation of symbols – "In a very real sense, this gives us the ability to share a virtual common mind" (427).

In the light of Deacon's hypothesis we can conclude that the (very influential) theorists that had been discussed in the previous chapters all had significant and vital insights into significant aspects of the problems to do with explaining mind and language. However, the phenomena in question are extremely complex and involve a vast web of related phenomena, many of which fall into different scientific and academic disciplines as these have traditionally been conceived of. The vastness of the subject matter, in conjunction with various conventions and assumptions within the specific disciplines in which these theorists worked, led to each theorist running up against certain conceptual dead-ends that could not be overcome in terms of the specific paradigm in which they worked.

Hence, Peirce (Chapter 1) argued for the amalgamation of philosophy and science (including evolutionary theory). He recognised that our cognitive structure constrains the way in which we experience the world and thus our knowledge of the world, and was sceptical of

the Cartesian conception of the mental realm as being formal and disembodied, unrelated to the material world. This caused him to argue that our knowledge is always knowledge in relation to the material world and he argued that it is in the course of learning to manipulate the signs in terms of which we acquire knowledge that our first assumptions in terms of which we interpret the world are formed. He proceeded with an attempt to formulate a theory of signs (semiotic) that accords with the facts rather than with inherited conventions with regard to the subject. However, Peirce could not overcome his conception of free will and the accompanying requirement that it not be "mechanical" in the manner that he conceived matter to be. Free will and Peirce's desire to retain some of his protestant convictions in his evolutionary approach led to the seemingly extreme metaphysical position that all matter is, in fact, embodied mind that operates according to semiotic principles.

Chomsky (Chapter 2) recognised that the behaviourist stimulus-response conception of acquiring language was hopelessly inadequate to account for the extensive knowledge that apparently needs to be acquired in order to make use of language, which has a seemingly endless capacity for generating novel and appropriate sentences (Plato's problem). This insight allowed him to look to the structure in which language presumably manifests in the brain, which he dubbed the "language faculty". He speculated that the only way in which to overcome Plato's problem is to acknowledge that our knowledge of grammar is innate, and manifests in some sort of language faculty in the brain. This led him to conclude that the structure of mind, as it pertains to the language faculty, can be deduced from the structure of grammar. The insight that aspects of language are influenced by the way in which our minds/brains are structured is crucial to understanding language and brain and Chomsky (although inspired by Peirce) has deservedly become one the most influential theorists in fields pertaining to language and mind. However, his overly theoretical approach to the language faculty and scepticism towards "scientific" (especially evolutionary) approaches to these questions have proved equally influential, without necessarily being justified. Furthermore, his insistence that the "language faculty" is a separate module within the brain is equally untenable given the empirical evidence that has been generated in recent years.

In Chapter 2 it became apparent that Chomsky's own revisions of his early theories indicate that he is aware that the bond between external language structure and language as it is embodied in the mind is much more tenuous than he believed it to be, however, he seems unable to address this difficulty in terms of his own "first assumptions". We saw that he remains unable to account for meaning in terms of language structure and for what he termed "the creative aspect of language use" (being able to generate a virtually infinite amount of sentences relevant to a virtually infinite array of situations).

While Derrida (Chapter 3) recognised the explanatory potential of the structuralist approach to language, he also recognised certain deficiencies within this approach. Hence, whereas he endorsed the argument that meaning in language does not arise from anything inherent to objects, signs, or minds, but rather from the differences within the language structure, Derrida appropriated Peirce's semiotic to amend the structuralist approach. He

added an extra dimension to the structuralist system, arguing that temporality and spacing also contribute to meaning within the system. Derrida realised that the "language system" also encompasses the linguistic memories of language users, whether it be their cognitive memories, or the collective manifestation of "memory" in archives, for example. Furthermore, the context in which meaning is created also influences meaning to a significant extent. Derrida's theory is essentially a theory of *reference* – how signs evoke meanings for sign/language users. He indicated that the language system, memory, and individual interpretive contexts are inextricably linked and signs gain their referential capacity from this structure as a whole. Derrida also extended the concept of language to include all signs that convey information, hence endorsing Peirce's semiotic position that our phenomenological access to the world is entirely mediated by signs. However, as we have seen, Derrida's hypothesis is entirely theoretical, and very much situated within the continental phenomenological tradition. He gives no indication of how it would manifest in the mind/brain, or even how we would go about establishing whether his theory is accurate. Hence, it should be understood as a critique of a particular metaphysical tradition and its preconceptions and preoccupations, rather than as a positive hypothesis.

It was only with the developments in logic in the 1930's and resultant realisation that the mind may be modelled according to mathematical principles, that the idea began to take hold that the mind could be studied in a manner similar to other "external" phenomena. The significance of this conceptual breakthrough can hardly be overemphasised. Gradually the mind would cease to be the exclusive domain of philosophers (an epithet that also applied to psychologists at that stage) and their thought experiments. As was discussed in Chapter 4, the resultant sciences of mind set about constructing artificial minds that mimic the abilities of human minds. However, due to what appears to be a misinterpretation of what Turing envisaged with the universal Turing machine, and the latent mind-body dualism that still lurked in many theoretical constructs, theorists tended to view mental operations as akin to logical operations – universal, disembodied, axiomatic, and radically decontextualised.

Models constructed in accordance with these principles failed to emulate human minds, however, and the (computational) science of mind in its various manifestations lost some of its momentum. The realisation emerged that the environment that provides input to the "mind" and a notion of the purposes in terms of which the "mind" operates were necessary to frame the problems that a mind would address. Towards the end of the twentieth century the conviction arose that these problems could be overcome with the incorporation of the theory of evolution by means of natural selection. The theory of evolution could provide a credible means for determining the "purpose" of mind (to ensure that their "possessors" survive and reproduce) and the input that they are likely to receive (environmental and bodily stimuli that the organism's sense are capable of perceiving and which are relevant to the survival and reproductive goals of the individual).

As became clear in the course of the final chapter, Deacon proceeded to incorporate all of these fundamental insights into a co-evolutionary theory of mind and language, which

combines both a computational and an evolutionary approach to mind, and that manages to overcome many of the limitations of the theories from which they arose. Thus, Deacon takes an essentially evolutionary approach to both language and mind. He works from the assumption that both phenomena developed in our ancestors by means of the principle of natural selection because they facilitated their survival and reproduction. As such he has a problem horizon in terms of which to analyse the workings of the mind, in that the mind/brain processes "compute" information from the body and the environment in terms of the fundamental purpose of keeping the body alive and enabling it to reproduce its genes.

Deacon accepts Chomsky's argument that language structure and mind structure are related, and explores what this relation may be. Deacon's own contention is that those aspects of the human mind that are uniquely human are the result of the selection pressures that were put on the minds of our ancestors by the adoption of a (initially crude) system of symbolic communication. Deacon envisages the symbolic aspect of mind along the lines of Peirce's semiotic and a Derridean distributed structure of traces – with information (signs) being minimally encoded in the mind in terms of icons (which can be likened to traces). Icons can be indexically related, while indices are related symbolically. The argument is that all sentient beings interpret the world iconically, which is the basic requirement for sentience. Furthermore, all sentient beings can relate instances of iconic information through associative learning and instinct. Relating indices symbolically, however, is based on convention and can only take place if the relevant symbolic conventions have been internalised.

Deacon (like Peirce) argues that these conventions are internalised through learning language. He rejects the Chomskyan notion of innate UG, but agrees that Plato's problem remains pertinent. Deacon argues that what is innate to human minds/brains is the ability to acquire language within a critical period of the child's development. He argues that language adapted to certain biases within the human mental structure in order to ensure that it is acquired. Furthermore, he believes that the shift from innate indexical interpretation to symbolic interpretation is facilitated by the organisation of the human brain during the optimal period for language acquisition. The immature human brain ensures that the underlying (symbolic) structure of language is concentrated on, rather than the specific (indexical) details of specific word references. He argues that the underlying symbolic structures of languages converges on universals in the same way that colour terms converge in universals across all human societies. The universal aspects of languages reflect cognitive universals in the brain that are not restricted to language functions, as is the case with UG.

Hence, Deacon constructs an evolutionary theory of a symbolic mind, which explains how and why mind and language structures are related; how language is acquired, without having to insist on an unsubstantiated innate grammar or modular language faculty; how signs refer; why symbolic language only evolved in Homo sapiens; and why Homo sapiens has mental abilities that seem to diverge radically from those in the rest of the natural world. The argument here is not by any means that Deacon provides the last word on mind-language evolution. What is being argued is that Deacon has managed to provide a

comprehensive and convincing theory that spans all of the aspects of mind-language interaction that have traditionally been highlighted as pertinent theoretical questions. Furthermore, he has done so as far as possible by means of the scientific evidence available to him. In this regard his theory remains open to falsification in that he does not commit himself to overtly metaphysical positions. Where the scientific evidence is sketchy, such as with the archaeological evidence for the initial shift to symbolic communication, he acknowledges the extremely speculative character of his theory. Hence, Deacon's thesis is an indication of how inherited limitations, such as disciplinary boundaries and metaphysical convictions, can be overcome in order to address questions as complex as those of mind and language. His methodology points to the only way in which these questions can be answered with any measure of success.

CONCLUSION

Philosophy is to be studied, not for the sake of any definite answers to its questions, since no definite answers can, as a rule, be known to be true, but rather for the sake of the questions themselves; because these questions enlarge our conception of what is possible, enrich our intellectual imagination and diminish the dogmatic assurance which closes the mind against speculation…
[Russell 1962 (1912):161].

Two quotations by Bertrand Russell opened this work and it seems fitting to conclude with a quotation also by him. In *The Problems of Philosophy* [1962 (1912):7] Russell argues that philosophy is the attempt to answer ultimate questions. In order to fulfil this role, it needs to be critical of the assumptions and dogmas common to "ordinary life and even the sciences" (7). He goes on to point out (141ff) that most of the matters that occupy the space in the writings of philosophers have to do with "*a priori* metaphysical reasoning" aimed at proving such things as the dogmas of religion, the rationality of the universe, the illusoriness of matter, and so forth. However, he argues that such theses are in vain and that knowledge of the universe is not to be obtained by metaphysics and proposed proofs that make conclusions based on the laws of logic. Russell severely limits the kinds of things that we are able to prove (145):

> Thus we are left to the piecemeal investigation of the world, and are unable to know the characters of those parts of the universe that are remote from our experience.

He is quite content with this conclusion, claiming that it is "in harmony with the inductive and scientific temper of our age" where it has been found that "very little can be proved *a priori* from considerations of what must be" [1962 (1912):145]. It appears that many apparent *a priori* impossibilities or necessities are in fact the results of "obstinate mental prejudices" [Russell 1962 (1912):147]. Writing early in the twentieth century, Russell gives an illuminating description of "the temper of our age" which merits quotation [1962 (1912):147] :

> The attempt to prescribe to the universe by means of *a priori* principles has broken down; logic, instead of being, as formerly, the bar to possibilities, has become the great liberator of the imagination, presenting innumerable alternatives which are closed to

> unreflective common sense, and leaving to experience the task of deciding, where decision is possible, between the many worlds which logic offers for our choice. Thus knowledge as to what exists becomes limited to what we can learn from experience – not to what we can actually experience, for, as we have seen, there is much knowledge by description concerning things of which we have no direct experience.

Russell argues that our knowledge consists of both *a priori* and empirical knowledge. Hence, philosophical knowledge does not differ essentially from scientific knowledge – philosophers do not have a unique source of wisdom or insight that remains unattainable for practitioners of other disciplines. He cannot but come to the conclusion that what sets philosophy apart from science is *criticism* [1962 (1912):149]. Philosophy as a discipline is characterised by its critical attitude towards principles adopted in science and in ordinary life, examining them for inconsistencies and invalid assumptions. Philosophy only adopts such principles when, to paraphrase Russell, it finds no reason for rejecting them.

Russell is not advocating complete scepticism however. He conceives of philosophy as a positive endeavour, which advances logical arguments in the light of facts [1962 (1912):150-152]. Thus, philosophy should consider "each piece of apparent knowledge on its merits" and retain whatever knowledge seems consistent and justified. Of course, cast in these terms, nothing indemnifies philosophy from error, seeing that knowledge is judged in terms of principles that may in themselves be flawed; philosophy is inevitably constituted within a given paradigm. Russell thus argues that the most that philosophy can lay claim to is that it reduces the risk of error in what becomes accepted knowledge. This is no small achievement, however, given that the accumulation and validation of knowledge is such a painstaking and fraught affair, very much prey to prevalent prejudices, assumptions, wishful thinking, custom, and even individual or institutional agendas. Philosophy keeps the possibility of speculation alive, which may result in possibilities being raised that would not otherwise have been conceived.

It is in this spirit that this thesis approached the work of the various theorists that have been discussed. Collectively, their work represents and addresses some of the major assumptions that have shaped thought on mind and language and their interrelation over the course of the twentieth century. And the value in studying their work lies in the philosophical enterprise of critically scrutinising these assumptions, especially in terms of their origins, but also in terms of their internal consistency, and justifiability in the light of recent scientific discoveries. With this in mind, this work is in part a historical overview of seminal positions in theories of cognition with an eye on the genealogy of some of the certainties in contemporary theories of mind, cognition, and language. But it is also a positive enterprise – demonstrating the efficacy and predictive scope that an evolutionary (and inter-disciplinary) approach embodies.

In the course of the work it became clear that all of the discussed theorists, while having made major and advantageous conceptual shifts with regard to basing theories of language and mind on scientific developments, were all restricted in their vision by their own metaphysical allegiances. It has been argued that the philosophical lineage of many of these metaphysical presuppositions could be

traced back to a mind-body dualism that has been implicit within the western philosophical tradition since Plato. This dualism resulted in a pervasive and entrenched "mind-first" paradigm, where the mental is regarded as prior to and privileged over (to use Derridean terminology) the material. It was also argued that such a dualism is the inevitable result of our common-sense (folk psychological) conception of the world, which further complicates the task of arguing for moving away from a categorical mind-body distinction.

We saw that both Chomsky and Putnam traced contemporary positions on mind, primarily the mind-body distinction, to Descartes' reaction to the new mechanistic scientific world view, as put forward by Galileo and Newton. Descartes, strongly influenced by his religious loyalties, could not account for free will and an insubstantial soul in terms of this new world view. Hence the move of conceiving of the mind as non-material and thus not subject to the apparently incontrovertible laws of physics. But his compromise was never a happy one, and Descartes never managed to convincingly solve the problem posed by the very obvious interaction between mind and body.

Despite Descartes' blatantly maladroit solution to mind-body interaction (claiming that the interaction takes place through the mediation of the pineal gland), mind-body dualism, with a mind subject to its own esoteric and unknown laws, became established within mainstream philosophy. As Chapter 1 demonstrated, even Peirce, who advocated the incorporation of science and even evolutionary theory into philosophy, could not bring himself to argue for a material mind. His scientific understanding was shaped by the classic picture of a mechanically causal universe, where particles of matter interact through contact, and where there is no place for "esoteric" forces that act over distances. However, we have also seen that Peirce realised that language (in the form of semiotic signs) was fundamentally intertwined with the mind. With regard to language he realised that the meaning of signs could profitably be accounted for in relational and structural terms. Thus, he viewed signs in terms of a hierarchical system that gain meaning from being in a triadic relationship between the sign, a mind, and context.

In this manner Peirce found himself in a quandary – his theory of signs implies that the mind and signs interact and constitute one another (he argued that we cannot think without signs). However, his conception of the material world prohibited him from conceiving of the mind as a material entity. The main reason for Peirce's unease with a material mind is that he conceived of a material world as being wholly *determined* by the linear interaction between the particles that make up the universe. An extended mind, linked to the matter of the universe, would be wholly determined by its interaction with the particles of the universe. Peirce could not think of a way to reconcile a material mind with free will with this world view. Hence Peirce's recourse to a universal mind. It should be remembered that Peirce worked in a world-view that conceived of the universe of being permeated with protoplasm. One of the qualities of protoplasm – or "life-slimes" – is that it feels. Furthermore, Peirce saw ideas as possessing energy with which to affect other signs, and also as having an intrinsic quality of feeling. Since, feeling is a quality of protoplasm, this means that feelings can be propagated through the universe in manner similar to the classical paradigm of interacting material particles. Peirce thus concluded that ideas (signs) are not limited to the mind, and because of their extension through their quality of feeling can interact mechanically with protoplasm. Finally, Peirce

reconciled the mechanical understanding of the universe, the concept of protoplasm, and his insight that thoughts consist of (triadic) signs and his need to account for free will and the soul by conceiving of the universe as consisting of ideas or a universal mind. In order to construct his metaphysical system, Peirce attributes the traditionally mental property of feeling to matter, and the traditional material property of extension to mind. His theory is not materialistic given that mind is not governed by mechanical law; in fact, the "original law" becomes the law of mind – the law of association – and the laws of matter become special instances of this law. Peirce argues that his theory should be judged through comparing its consequences with our common-sensical experience of our minds, which he of course believes is a favourable comparison.

In terms of the scientific paradigm of his day, Peirce's "objective idealism" could be considered to be a possible and legitimate scientific hypothesis. He does not introduce esoteric or metaphysical entities into his project. On the contrary, his theory examines the possible logical implications of the scientific paradigm of his day for the mind-body problem. His universal mind may seem hopelessly metaphysical to contemporary sensibilities, which generally shun concepts such as protoplasm or matter that can feel, but would have seemed quite acceptable in the late nineteenth century. And while the scientific paradigm would change over the next hundred odd years, Peirce's central concerns with the mind-body problem would remain.

Through his semiotic, Peirce anticipated the central role that language would play in theories of mind in the twentieth century, as well as the focus on the structure of language in theories on language and meaning. One of the most influential theorists in this regard would be Noam Chomsky. As we saw in Chapter 2, Chomsky believed that the study of language would give us insight into the structure of the mind. Thus, Chomsky too believed that language and mind were fundamentally interrelated – to such an extent that he expected the logical structure of the mind to mirror the logical structure of language. By developing a "science of language", he believed he would also be developing a science of mind. His project was to study the mind and "metal-phenomena" in naturalistic terms, and he rejected most work done on mind on the grounds that they made unjustified "non-naturalistic" assumptions.

We saw that Chomsky argues that there is no coherent mind-body question and that questions of representation or intentionality are radically misconceived. He believes that contemporary theories that cast certain philosophical questions in mind-body terms have failed to take account of the collapse of Cartesian dualism with Newton's postulation (and science's acceptance) of gravity. This led to the collapse of the mechanistic understanding of the universe. To paraphrase his argument, in a milieu where science invokes "fields of force, curved space, infinite one-dimensional strings in ten-dimensional space" etc., it becomes all but impossible to equate the "physical" or the "material" to that which is subject to "mechanical causes". In other words, the mind-body question presupposes an outdated and discredited understanding of what physical entities are – an understanding that does not make sense outside of a mechanical philosophy. Chomsky thus argues that the mind-body question can only be raised in its Cartesian form on the grounds of unjustified dualistic assumptions.

In an attempt to develop a naturalistic understanding of mind and language, and inspired by Peirce's concept of abduction (see Chapter 1), Chomsky identified "Plato's problem". This problem

was based on the realisation that, given the complexity and almost unlimited combinatorial possibilities of language, it is impossible for children to learn language from scratch and through induction. Chomsky concluded that human beings are *designed* to learn language, and thus proposed the existence of a Universal Grammar – a uniform human cognitive state from which the grammar of particular languages can be derived by language learners, which eliminates the need for learning grammar from input sentences from the environment.

Chomsky thinks of UG as a kind of "language acquisition device" within the brain, situated within the (idealised) language faculty. The language faculty, as a "mental organ", is assumed to be autonomous and to actually exist in the brain as a modular subsystem of the brain. From analysing the structure of the language faculty, Chomsky proceeds to explain many of the properties of sound and meaning and the relations between expressions across a variety of languages. He argues that many of these properties derive from our "inner nature" as determined by the initial state of our language faculty, which accounts for the universal aspects of the different human languages. Although he can only characterise the properties of language and of grammar in abstract terms, Chomsky believes that with the progress of science it might become possible to know something about the physical representation of grammar and of the structure of the language faculty, although he is quite sceptical about the ability of science to address complex phenomena. He argues that his theories are not without empirical content, however, because it is foreseeable that discoveries in neurophysiology or in the study of behaviour and learning might cause us to revise or even abandon a given theory of language or a particular understanding of grammar.

Chomsky also argues that it is the faculty of language that contributes to the construction of a "common-sense understanding", which yields the common-sense world that all people have in common. We saw that Pinker and Churchland both address the role that common-sense understanding (or folk psychology) plays in our understanding of the world. Churchland (Chapter 4) holds that it is the general conception of the "mental" in western philosophy can be traced to our common-sense understanding, or folk psychological descriptions of such phenomena. Chomsky argues that common-sense understanding, like language, is ubiquitous within the human species and seems to be acquired with as little effort (as opposed to scientific understanding, for example). A common-sense understanding situated in the language faculty would explain how our ideas seem able to "anticipate" experience. Chomsky envisions "internal concepts" provided by the language faculty, which primarily have to do with those human interests that relate to our biological constitution and requirements. However, despite this overtly biological attribution to the language faculty, we saw that Chomsky insists that it is "science" and not evolutionary theory that should provide the naturalistic framework in terms of which the language faculty should be studied.

In his early work, Chomsky restricted himself to those aspects of language that could "be sensibly pursued", namely the study of the nature, acquisition, and use of linguistic competence. He believed that the state of knowledge available to him would allow for progress to be made on the way in which classic problems of language can be formulated and thus allow for novel approaches. Over time, his conception of UG and how particular grammars come to be specified underwent various variations. Finally, in step with contemporaneous theoretical trends, Chomsky came to think of the

language faculty as a "computational system" and as such he attempted to approach it in terms of general computational principles. With his minimalist programme, Chomsky attempted to formulate a theory of language as an optimal response to certain computational problems. This led him to do away with many of the structures he postulated in earlier generative grammar, claiming that "Nature can't be that ugly".

Chomsky's attempt to formulate a theory of the language faculty in terms of an optimal computational approach reflects the concerns and difficulties that faced attempts in cognitive science to model the mind as a computational system (chapter 4). In the absence of a biological framework, determining what an "optimal" solution to computational problems faced by the brain would look like becomes an almost impossible task. Just what are the "computational problems" that the brain is supposed to address? Furthermore, the assumption that the mind would exhibit an "optimal" design to meet its particular computational demands seems to an arbitrary and unreasonable restriction to impose on a biological phenomenon. As became clear in our discussion on the evolution of biological systems (chapter 4), nature works within the constraints of a natural world of previously realised designs. Thus, we cannot expect biological phenomena to mirror the logic of our designed artefacts – an insight that needs to be incorporated into computational theories of mind if they are to model an accurate problem horizon in terms of which a system operates. (We will come back to this point).

Chomsky began his career when much of cognitive science was in its infancy. Thus, despite his scientific inclinations and his proposed naturalistic approach to the problems of mind, one of the few avenues open to him (as with early theorists in philosophy of mind) was to study the mind in terms of language. While this approach was an improvement on the behaviouristic approach and on deducing "mental characteristics" from introspective reports, Chomsky's method cannot be said to be wholly scientific. The strong correlation that Chomsky draws between the structure of mind and the structure of language can be attributed to the semiotic and structuralist traditions that must surely have influenced him. However, there seems to be no reason why the "language faculty" should be modular (there is much scientific evidence that language is in fact highly distributed throughout the brain). In the absence of a modular language faculty that interacts with the brain in a highly constrained manner, it becomes difficult to draw any conclusions about the mind from studying the structure and grammar of language. And even if the language faculty were modular, there is no reason to assume that the analysis of the structure of language would render any insight into such a structure of the modular faculty. (In Chapter 4 we discussed similar objections to viewing the mind in terms of "propositional attitudes").

Chomsky's naturalistic approach to mind could conceivably have led him to incorporate insights from both computational theories of mind and biological analyses of the origin and purpose of the brain, seeing that he ultimately believes that the mind is part of the brain. However, we have seen that Chomsky is strongly opposed to an evolutionary approach to mind and language. A clue to this resistance on his part is perhaps to be found in his later work, where he argues that language should be an optimal solution to computational problems faced by the mind, even while he admits that empirical observation seems to suggest that language does not adhere to this requirement. Chomsky cannot conceive of nature as being "ugly". Some computational approaches to mind, like a certain

conception of "pure" physics, promise a "disembodied" analysis of phenomena. In this conception, phenomena can be reduced to elegant, "optimal", and presumably simple physical laws.

On the other hand, evolutionary "design" exhibits a kind of "compromise position". Evolutionary design is constrained by competing demands. Pinker (chapter 4) pointed out that language design is almost certainly a compromise between the divergent communicative needs of speakers and hearers, such as minimal demands on the cognitive and other language-related systems in learning and producing language, while conveying the maximum amount of information possible in order to facilitate understanding. Furthermore, evolution does not "design" biological phenomena teleologically. Already realised designs in ancestors, accidental but beneficial mutations, and changes in the environment make for cobbled-together, yet extremely effective, biological designs. (If a design were not extremely effective it would quickly become extinct soon enough and thus disappear from the gene pool). Thus, Chomsky's requirement of optimal design seems to have more to do with his conceptions of what the logic of both language and the mind should look like; and perhaps with a wish to be able to continue with his speculative analyses of both phenomena. The alterative might imply that his subject matter is better studied by the life-sciences than through linguistic analysis.

Chomsky's project was partly a reaction to the shortcomings he saw in structuralism and behaviourism in their treatment of language and in analytic philosophy in its conception of both language and mind. One of Chomsky's major criticisms of structuralism was that it failed to account for the creative aspect of language use, and for its semantic content. However, despite various attempts throughout his body of work, Chomsky cannot account for either the creative aspect of language use or for meaning and his worked focussed mainly on the structure of language. As with Chomsky, Derrida appropriated Peirce's semiotic to amend the structuralist approach. However, Derrida's analysis would focus mainly on meaning. Significantly, he would develop a theory of meaning in terms of the structure of the system – something that Chomsky was never able to do.

Like Chomsky, Derrida was very aware of the metaphysical tradition in terms of which the sign had been understood in the past, and he traced traditional conceptions if the sign to a particularly Greco-Christian world-view. Here thought or *logos* is conceived of as a realm of pure perception or divine intelligibility, uncontaminated by worldly contingencies and the fallibility of the senses. The sign is conceived of as consisting of a "worldly" and possibly fallible signifier, while the signified is the true and unambiguous meaning to which the signifier refers. The intelligible truth underlying sensible appearances (the meaning underlying the signifier) is supposedly accessible through rationality, that ideal state of the mind (which can be linked to the divine, as Descartes had done) which is not fooled by fallible worldly appearances. Logos is a necessary condition for true knowledge. Thus, Derrida linked the signifier/signified distinction to the postulation of *logos* and to the parallel idea of a creator God.

What Derrida thus aimed to show was that what historically seems self-evident within a given metaphysical tradition, need not be self-evident at all. Situated within the logocentric tradition, Derrida does not propose to overcome or surpass this metaphysical stance, but proposes to disrupt it "from within". Metaphysics, here, is understood to mean the conditions of possibility of a particular epistemology – in many ways his position is akin to that of Peirce's abduction and Chomsky's UG. His

position is that it is not possible to "step out of" a given metaphysical position, but that "nothing is conceivable without a certain metaphysical stance". The most that one can hope to do is to identify the historical and systematic origins of certain metaphysical assumptions, and to realise that these assumptions are not inevitable or necessary. In terms of Derrida's view Peirce's ultimate recourse to theology in order to ensure the possibility of definite knowledge and universal truths is not an anomaly but an inherent part of his, and Derrida's, metaphysical tradition. Derrida does not take the same recourse, however. Instead of trying to ground ultimate, objective, and certain truth in a phenomenon external to human cognition, Derrida settles for the "provisional" truth of meaning that is constituted within a system of differences. However, ultimate truth, independent of human situatedness in the world, will always remain a metaphysical fantasy.

As was seen in Chapter 3, Derrida, through the concept of *différance*, added an extra dimension to the structuralist language system, arguing that *temporality* and *spacing* also contribute to meaning within the system. Derrida realised that the "language system" also encompasses the linguistic memories of language users, whether it be their cognitive memories, or the collective manifestation of "memory" in archives. Furthermore, he, like Peirce, argued that the context in which meaning is created also influences meaning to a significant extent. Hence, he argued that in order to account for the meaning of a signifier, one had to take a relational approach, understanding the sign in terms of its relation to other signs within the language system (as did Saussure), the context in which it was being used (a position also held by Peirce), and in terms of the instances in which the person understanding the sign would have come across it before, as well as the person's anticipation of how it can be used. Thus, he significantly expanded both Saussure's and Peirce's relational understanding of language, while incorporating the embodiment and the history (or more accurately histories) of the system of signs into the equation. As we saw, this crucial insight is mirrored in neural network theory, where feedback is introduced into the computational system, allowing for the network to compute some of its output an input and thus build up a rudimentary memory.

Derrida argued that a way to overcome the closure of a traditional cause-and-effect framework with regard to language, and to formulate a concept of cause and effect *appropriate* to a system of *différance*, is to adopt the idea of the *trace* (see Chapter 3). A trace is neither a cause nor an effect, but the condition of possibility for signification. Hence, the movement of signification is possible because of *différance*, in that every element (*trace*) present to a signification is related to elements other than itself. Elements, or traces, always incorporate elements or traces of those signs to which they were related in the past, and traces to which they can be related in the future. Any sign is understood in terms of signs to which it was related in the past, and in terms of other signs it may conceivably be related to in the future. Derrida envisages that signs can thus be seen in terms of a relational system of traces – elements primary to signs which relate signs not only to other concurrent signs in the system of language, but also to past and future instances of those signs. Put in these terms, it becomes clear that Derrida's analysis of how meaning in constituted mirrors the structure of neural networks (chapter 4). It can thus be argued that Derrida developed a model which theoretically postulates how meaning could come to about in a distributed computational system (like the brain) in a manner other than through representing language symbols or propositions, as was a prominent

position in philosophy of mind. However, Derrida's project was fundamentally phenomenological, and he gave no indication of how his theory could be empirically verified (or falsified). Or of how language manifests itself in the mind/brain. As such, there does not seem to be much reason to accept his postulation over that of Peirce, or Chomsky, or various other philosophers of language and mind, or even our own common-sensical conceptions of both language and mind.

It is wholly impossible to judge the veracity of theories without a paradigm in terms of which one can define concepts such as "truthfulness" or "logical possibility". Within the current milieu the primary frameworks for establishing truth conditions and theoretical acceptability of hypotheses concerning natural objects are provided by the natural sciences. As discussed in chapter 4 the "material eliminativists" rejected hypotheses concerning the natural world that are formulated in terms of our common-sensical, or folk psychological understanding of the world (cf. the discussion on Churchland for example). Such theorists argue that hypothesis and disciplines that have fallen into disrepute, such as phlogiston and alchemy (and protoplasm), were perfectly coherent concepts and disciplines in terms of our common-sense conception of the world. Furthermore, as Chomsky points out, many of our scientific breakthroughs are "breakthroughs" precisely because they run counter to our common sense conception of the world, and hence were extremely difficult to "discover" or formulate. Scientific methodology is developed to overcome our intuitive assumptions in order for us to study phenomena on their own terms, as it were. There is no reason why the mind (and its relation to language) should be exempt from this trend. Our folk psychology might suggest that the mind is "disembodied" and "non-material", and that we think in terms of natural language. However, as with other scientific breakthroughs, there is no reason to adhere to our common-sense understanding of phenomena if science presents us with an alternative, viable, and justifiable scientific understanding of them. We saw that the "mind-first" paradigm has proven to be insidious, even hampering to the efforts of proponents of an amalgamation of science and philosophy, e.g. Peirce and Chomsky, to overcome untenable philosophical presuppositions. But until these assumptions are subjected to scientific scrutiny, there is no reason for us to accept these positions any more than there is reason to accept that the heavenly bodies revolve around the earth.

This position is further strengthened by developments in logic early in the twentieth century that indicated a way in which the mind can be thought of that would make it amenable to scientific (empirical) study. Chapter 4 traced this development, where the possibility was raised that the mind can be viewed as an information-processing entity. We saw that the development of the computer further strengthened this metaphor, in that it provided a working model of a functional information processing machine. The possibility arose that the processes in the mind are material in the same way in which information processes are material. In this view the mind need not be viewed as a mysterious result of the "pulley system" that Descartes proposed as a model for how the body functions. The conception of "material" entities could be expanded to include the effects of seemingly immaterial and disembodied entities (information) on the nervous systems. Furthermore, viewing the mind as "computing" or "processing" information that it gleans from the senses, raises the possibility that both the external "material" world and the internal "mental" world are subject to interpretation. The idea arose that the mental world can be rooted in the physical world through the concepts of information,

computation, and feedback, and that the mind can be modelled in those terms. In this view mental phenomena such as beliefs, intentions, and memories are composed of information; information which is embodied in patterns of activity within the brain.

However, the newly developed sciences of mind did not completely overcome the dualist conception of mind as something separate and relatively disassociated from the body. In many ways the early approaches to mind in these disciplines, which approached the mind as a digital or symbol-processing computer, was still very much caught up in the Enlightenment view of the mind as being the seat of rationality, conceived of along the lines of deductive logic. In this view, the information being processed in the mind is context-independent, and could be considered to be "objective", in that logical methodology is universally applicable. Certainly, the notion of mind as a biological entity did not feature in this research programme.

There was early enthusiasm for the conviction that the mind could be accurately modelled and thus that it would be possible to create artificial intelligence. However, over time the digital paradigm came to be regarded as not being equal to the task of creating artificial intelligence. Connectionism, which originated from an idealised model of the neuronal architecture of the brain, seemed to have more success in this regard. As was seen in Chapter 4, this was largely due to the parallel and distributed computational processes embodied by neural networks, instead of the digital, symbolic processing of digital computers. Processes such as distributed representation of information, memory, and feedback were emulated, allowing for computational models that could learn, generalise, and adjust its structure in response to its environment. All in all, connectionism seemed to be a more promising approach to modelling the cognitive processes of the nervous system. However, there were some shortcomings in this model. Firstly, they did not come close to emulating the complex and diverse abilities of natural nervous systems. Furthermore, they required intensive design and training in order to accomplish relatively simple tasks. How was it possible that such systems arose "naturally" in the animal kingdom? And how does "design" occur in natural neural networks?

There are significant parallels between the early computational approaches to mind and twentieth century philosophy of mind. Much of the subject matter that would characterise twentieth century philosophy of mind derives from attempts to clarify our knowledge of the "objective world" in terms of our immediate and presumably inherently subjective experience. As we have seen, the inauguratory move of twentieth century analytical philosophy (and ultimately philosophy of mind) was the integration of this prevailing concern with a programme of linguistic and logical analysis. As with structuralism, analytical philosophy believed that the key to establishing meaning (objective meaning in this instance) was the formal analysis of the language in terms of which we articulate (subjective) experience. A crucial difference between early computational approaches to mind, and early philosophy of mind was that the mind was still essentially a mysterious and inaccessible entity to philosophers of mind. The only way that the mental was accessible to these theorists was through introspection, or via the first-person reports of the introspection of others. Thus there was no conception of how the mind could be modelled or analysed through scientific methodology, which meant that there was a strong resistance to "reducing" the mind to the physical.

A very important factor in the resistance to the "reduction" of the mental to the physical in philosophy of mind is the early commitment within the discipline to studying the mental in terms of linguistic analysis. Despite the subsequent move towards materialism within the discipline, the assumption remained that a first-person linguistic account of "metal events" is a legitimate, and indeed the only legitimate, way of describing the mind. This assumption was so entrenched, that the early (and some later) physicalist approaches expected the language reports of mental experiences to be somehow directly "translatable" to the mental substrate. In this manner, the intuitive (and presumably correct intuition) that language is somehow fundamentally intertwined with the mind has caused an excessive and unwarranted role being attributed to linguistic analysis in the study of mind.

As the computational metaphor took hold in theories connected to mind, and the idea that some aspects of the mind could be considered to something akin to information processing became more widespread, the argument remained that there are some aspects of mind that are still irreducible "phenomenological". There is something that it is "like" to be a mind, to experience things first-hand, which cannot be understood in material terms. Hence, the proper subject matter for philosophy of mind became those phenomenological aspects of mind, such as intentionality, beliefs, desires, and qualia.

We have seen that there is currently a trend towards an "embodied embedded" approach in the cognitive sciences, which is accompanied by the "rediscovery" of Darwin and his theory of natural selection. It seems that the computational approach had come up against an insurmountable hurdle: thought constitutes more than computation. What remained critically absent in traditional computational approaches to mind were concepts of the environment in which the mind operates (receives its input) and a notion of the purpose of mind (what its input is used for), as well as its historical development. It now seems highly improbable that the human mind can successfully be conceived of as an abstract, disembodied, logic-machine.

In Chapter 4 we saw that Pinker argues that the structure of the mind is best studied by approaching it with the assumption that it was designed by natural selection to solve the kinds of problems that the species has been faced with for most of its existence. The rationale for this approach is of course Darwin’s theory that natural selection seizes upon random, advantageous mutations in species because such mutations enhance the organism’s fitness in the struggle for existence and hence ensures that it increases its progeny.

Evolutionary theory offers clear criteria for when a trait can be attributed to natural selection, namely that it exhibits complex design for a specific function, as well as the absence of an alternative process capable of explaining the complexity. Seeing that the theory of natural selection is the only successful account of the origin and development of complex biological structures, the position in this thesis is that language and mind should be studied in terms of natural selection. It is very difficult to imagine how a nonadaptationist process such as genetic drift or a general law of growth could result in an arrangement of matter that would result in such complex structures, which were then fixated within the population. The advantage of Darwin’s theory is it that allows for seemingly teleological adaptation, in the absence of some form of foresight or a designer. In this approach, it is generally accepted that the mind needs to be understood in the context in which it is used and for which it was developed

(note that the context can change over time). It is assumed that the structure and abilities of the mind is fundamentally linked to the role that it can play in terms of the "goals" of natural selection, namely survival and reproduction.

With regard to what the mind and language are "for" Pinker shows how evolution can answer this question, in that it enables us to "reverse-engineer" the structure of the mind. He argues that language and mind allow humans to create new knowledge by "mentally playing out" possible combinatorial interactions between object in the world, and then predicting an outcome. In order to accomplish this, humans make use of intuitive theories of objects, forces, pasts, places, manners, states, substances, hidden biochemical essences, and for other animals and people, beliefs and desires – a "folk science" made possible by both language and mind. Human languages combine words into grammatical sentences and use the underlying propositions to reason about invisible and relevant entities such as absent people, animals, noteworthy places, etc. Mental maps represent the important places, events, and agents and their functioning. In this understanding, language is also a means for exchanging knowledge and thus increasing the benefit of knowledge that might have been acquired by expensive trial-and-error for which someone else had paid the cost. Pinker furthermore argues that the mind is "what the brain does", i.e. the processing of information, thus thinking is seen as "a kind of computation" (Chapter 4). In this understanding the evolution of mind is inherently linked to the evolution of the brain.

Thus, Pinker combines the ideas of the computational theory of mind and the theory of the natural selection to explain the origins of both mind and language. The result is a comprehensive theory, incorporating a wide range of disciplines, including anthropology, palaeontology, psychology, philosophy, biology, cognitive science, neuroscience, and so forth. The interdependence of the brain/mind, language, the evolutionary environment, human physiology, the social environment, language structures, learning abilities, the reproductive forces driving natural selection, and a range of other factors is taken into consideration in such accounts.

It is generally accepted that a computational approach to the mind in this context cannot ignore the substrate in which such computation takes place, namely the brain. Hence, a computational "machine" in this sense does not equal a "Turing machine" or an "idealised neural network". Furthermore, it is generally accepted that structure of the nervous system and the body both enable and constrain the mind's computational abilities. Various roles are attributed to language, in this regard, both in terms of its role as a source of vital information which the mind "computes", and as an entity that can literally shape the structure of a brain (to a greater or a lesser extent, depending on the theory). A system dedicated toward solving particular problems (obstacles to survival and reproduction, for example) cannot accumulate information from its environment at random, but must be equipped with a set of "core truths" about its environment and a set of rules from which the implications of these truths can be deduced. Pinker refers to these rules as the "rules of common sense". Hence, by means of the "rules of common sense" the mind can deduce *relevant* implications from what it knows. We saw that Pinker argues that generations of philosophers overlooked the implication of having to "frame" knowledge in terms of relevance to the organism because of the apparent effortlessness of the common-sense in terms of which they thought. It is only with the advent

of artificial intelligence and its attempt to duplicate our common-sense in computers that the problem of framing knowledge became a serious topic of study. Our common-sense understanding of the world now seems to be a function of the structure of our minds rather than incontrovertible and universal laws (this insight was, of course, anticipated by Peirce's concept of *abduction*).

It is the structure of the mind and its resultant "common-sense" orientation to the world that explain the existence of beliefs and desires within the organism. Anthropomorphically speaking, intelligence entails having desires, pursuing them, and using beliefs about the world that are effective in that they are at least approximately or probabilistically true. It is a function of our understanding of organisms that they act in order to accomplish some goal (food, shelter, movement, and the like) that causes us to view them as having beliefs and desires. Pinker refers to this orientation in organisms as "intuitive psychology" (note the correlations between this view and Churchland's concept of "folk psychology") and it is a powerful tool for predicting, controlling, and explaining behaviour.

Here we have an indication of how a computational-evolutionary approach to language and mind can address some of the central issues raised in philosophy of mind, such as beliefs, desires, as well as an issue addressed by Peirce, Chomsky, and even Churchland namely the notions of abduction (Peirce), Universal Grammar (Chomsky), or folk psychology (Churchland). The question of why people seem to be inclined towards certain aspects of the world or certain views of the world can be explained in terms of evolutionary "goals". The human nervous system has evolved in such a way so as to attend to information and views that have the greatest possibility of being true and thus increases the organism's chance of making correct interpretations and predictions about the world and to act in a manner that will ensure its survival. If a particular organism were to habitually make incorrect interpretations and predictions it would not survive long enough to procreate and thus its inefficient "Folk Psychology" would thus disappear from the species. This is, of course, a hopelessly simplified formulation of the theory, but it is sufficient to indicate how an evolutionary theory would account for human beings' abductive abilities and common-sense understanding of the world. Thus we have identified a "problem horizon" – something that was lacking in computational theories of mind, where a system's problem horizon is wholly determined by a human designer or programmer.

Language has a central role to play here. From an evolutionary perspective, the persistence and universality of language would suggest that it is beneficial for human survival and procreation, and hence has been "selected for" during the evolutionary process by providing a competitive edge to those who possessed it. While the computational-evolutionary approach minimises the role that language has to play in studying the mind, it sees language as an important factor in the evolution of the mind. Language also has a crucial role to play in survival in that it allows for crucial information to be shared, which enhances human beings' ability to virtually construct models of the world, which facilitate analysis, prediction, and interpretation of how the world (and its inhabitants) "acts".

Deacon (Chapter 5) gives an evolutionary account that plausibly explains why the intuitive view that language and mind are incontrovertibly related is so strong in both folk psychology, and philosophies and sciences of mind. He explicitly links the evolution of the human mind as we know it with the evolution of human language. In this manner he manages to formulate an even more comprehensive theory of mind and language evolution, which has the capacity to address many if not

most of the philosophical quandaries discussed in the course of this work. As we saw, Deacon questions much of the accepted wisdom in many of the disciplines that are concerned with language and/or mind and explores the implications that an evolutionary and radically interdisciplinary approach to language and the mind will have on such received wisdoms. Furthermore, he makes use of all of the primary theorists and insights discussed in the course of this work in constructing his theory of the co-evolution of language and the brain. Hence, Deacon attempts to rectify faulty assumptions and present alternative interpretations of the evidence that we already have at our disposal regarding languages, brains, and evolution.

Deacon sees the evolutionary origins of language as an entry point into the question of the logic linking our cognitive functions to brain organisation (as we saw a major point of contention within functionalism). His point of departure is asking the question: "Where do human minds come from?", and his answer is: From our ability to create and manipulate *symbols*, which stems from the novel form of mind that developed in the interface between the evolution of the human brain and human language.

As we saw, Deacon argues that human language is not just a mode of communication, but also the outward expression of an unusual mode of thought – *symbolic representation*. He believes that other animals do not have access to symbolisation and the virtual world that comes with it. Furthermore, the human ability to think symbolically is not innate, but *develops* with the internalisation (learning) of the symbolic process that underlies language. Hence, only minds with the ability to communicate symbolically can think symbolically. In order to underscore his position Deacon makes use of Peirce's analysis of the sign and his division of signs into a hierarchical structure consisting of icons, indices, and symbols. It is crucial to remember that Deacon does not equate *sign* with *symbol* – symbols are one type of sign. In keeping with Peirce's semiotic, Deacon is careful to emphasise that objects are not intrinsically icons, indices, or symbols, but are *interpreted* to be so. The interpretation depends on the response that they produce. He proposes that all sentient beings can make use of icons and indices to interpret and refer to the world. (Deacon calls this instance of biological semioticity, *brain-based awareness*.) Only human beings, however, have the ability to think and interpret the world in symbolic terms.

Deacon argues that it is the symbolic nature of words that gives them their meaning. In terms of Peirce's theory, Deacon believes that this symbolic nature requires interpretants (see Chapter 1) over and above those required by other forms of reference (i.e. iconic and indexical). These interpretants are the result of a great deal of additional learning on the part of humans. As with Derrida, Deacon argues that the symbolic basis of words is further mediated by their underlying systemic connections. In being interpreted, words are interpreted in terms of the many iconic and indexical associations (traces) that they have with other words.

Deacon, like Derrida, criticises the approach of much of western philosophy to language as being caught up in a "metaphysics of presence", to use Derrida's formulation, where the "disclosure" of the "truth" is the presentation of the thing itself to consciousness or the mind in the form of a signifier (see Chapter 3). The traditional conception of reference seems to have myopically been focussed on an object-reference relationship, where two phenomena are involved that have to be

somehow present to one another (either physically or in terms of shared characteristics) in order for reference to be established. Derrida's inversion of the spoken vs. written language dichotomy serves to highlight the same point that Deacon is making – correlation is not a prerequisite of (symbolic) reference. In fact, the symbol must be able to function in the radical absence of its referent. Furthermore, Deacon reaches the same conclusion that Derrida does in another respect – symbolic reference is distributed across the systemic relationships of the symbols within the language (and neural) system. This means that not only can words function as straightforward indexical signs, but most of them can be used to point to complex sets of indices that are connected by means of complex syntactical relationships. Through the use of a system of symbols (words) with a logic that does not rely on spatial-temporal (indexical) correlation for reference, a virtual "reality" is created based on a symbolic logic. Everyone who has access to the system of symbols and has the ability to think in symbolic terms can engage in semiotic activity, which is freed from iconic and indexical referential constraints. This allows for the conceptual, the abstract, and even the fantastical to have meaning.

What is it that allows human beings to be able to learn and implement a system of symbolic associations, while animals seem unable to do so? Deacon argues that a form of regimented and possible combinatorial organisation is a logical necessity for any system of symbolic reference. Without an explicit syntactic framework and an implicit interpretive mapping, it would be impossible to produce unambiguous symbolic information, or to acquire symbols in the first place. Syntactic structure can be seen as the higher-order combinatorial logic or grammar that enables and regulates symbolic reference. Thus, he presents us with an argument for the fundamental interdependence of syntax and semantics. But where does such an implicit syntactic framework come from? And why is this framework seemingly uniform across all human populations? On the face of it, Deacon's argument seems to be a version of what Chomsky has dubbed, Plato's Problem, and seems to support the existence of a Universal Grammar.

Deacon emphatically rejects Chomsky's postulation of the existence of a Universal Grammar, however, seeing this theory as an illegitimate recourse to a "freak mutation"-theory. In terms of Deacon's project, a legitimate argument in terms of evolutionary theory needs to be found for UG. He does believe, however, that children are born with a certain predisposition to learn human language. He holds that the fact that all normal children raised in a particular society will learn the language of that society proves that human brains are equipped for, or predisposed to, the language function in some way. But he rejects Chomsky's proposed innate language competence to the extent that it requires knowledge of grammar to be present *a priori* in the human brain, and that it ignores the feedback that children receive from the intricate social context in which they learn. He argues that this move only manages to relegate its unanswered questions to evolutionary biology and neuroscience, without presenting convincing evidence that such innate knowledge is even biologically possible.

Deacon observes that brain structure has been notably conserved from the origins of vertebrates, while languages can become unrecognisable within a few thousand years. From this observation he concludes that we have been focussing on the wrong half of evolutionary equation when it comes to flexibility and adaptation. His hypothesis is that *languages* evolved with respect to the human brain, and not according to arbitrary or logical principles or constraints. A child faced with

learning a language is aided by the fact that the language would have evolved to be intuitive, or "user-friendly", to facilitate its learning by children. Deacon argues that there seems to be a strong bias in children when choosing ways to organise learned words, and for Deacon, this points to an apparent ability in children to anticipate regularities in language. He speculates that language might be organised in such a manner that children's intuitive guesses are more likely to be correct. Furthermore, he does not present the "source" of language knowledge to be either in the child's brain, or the environment from which the child learns language, but as highly distributed across the interactions between the child's learning and the evolution of the language community.

However, Deacon also shows that the structure of the human brain has been significantly influenced over the course of its evolution by the requirements posed by language, although at a much slower pace than the structure of language has adapted. He argues that only the most invariant and general features of language would exist long enough to influence brain evolution. Thus, Deacon concludes that there is little possibility of mental adaptations to specific syntactic structures. Only those features of language that are common to essentially every language, and which pose consistent and invariant demands on neural processes, would contribute to brain structure. Hence, the key requirement for genetic assimilation is the existence of some invariant sensorimotor features or invariant mnemonic features of the adaptation. As to aspects of the human brain that seem to have evolved to support language, the argument is not that evolution produced a "language organ" in the genes, but that it recruited various regions and abilities of the hominid brain so that they can support symbolic reference. Eventually, these initial symbolic abilities would have been applied in more aspects of human life and would have become more complex, hence increasing selection pressures for brain structures that can support language functions. The co-evolution of language and brain-structure implies that languages are not designed to form formally logical systems. This allows for the possibility that it might not be possible to assign utility or purpose to the principles of language, and we might not even be able to recognise their design principles.

Deacon's further argument is that the discontinuity between stimulus associations and symbolic reference associations makes it *impossible to assimilate symbols genetically.* He concludes that if there are innate rules of grammar, they could not have gotten there by genetic assimilation. Deacon concludes that only the aspect of language that is most ubiquitous would have influenced brain structure. And he identifies this ubiquitous aspect as the unprecedented cognitive computational demands posed by symbolic reference. He is adamant that no innate rules, innate general principles, or innate symbolic categories can be built in by evolution. The noncorrelative nature of symbolic reference has cut language off from forces that shape biological evolution, and instead has shifted the burden of adaptation to a new level of information transmission (language). This insight also provides important constraints for theories that are quick to see a parallel between natural language and "mentalese" or natural language and symbol manipulation in digital computers (such as that of Fodor, for example).

Over and above Plato's Problem, Deacon also addresses the existence of *language universals* in this manner. Deacon acknowledges the existence of grammatical universals, but does not concede that they are somehow hardwired into the brain. In fact, he argues that universal

grammatical rules are not located anywhere and that they are not determined. Instead, he suggests that the implicit axioms of grammar have emerged independently in each evolving language, in response to universal biases in the selection process affecting language transmission. In other words, language universals are convergent features of language and brain evolution. Thus he does away with the need for innate knowledge of grammar and an innate "language faculty" while presenting a biologically and evolutionary plausible alternative to these assumptions.

Deacon's argument has done away with the need for an innate and universal grammar, and replaced it with a social evolutionary process, which continues to evolve with time and further generations of language users. His point is that there are no "rules" of grammar, whether in the brain or in the language community, which need to be learned in order to grasp language. Language structures are the products of powerful evolutionary processes on many levels, and a set of basic design rules or principles would simply be redundant. He then proceeds to explain how language (and universals) is acquired, despite the apparent insurmountability of Plato's Problem. Deacon's suggestion is that the universal biases in the human brain facilitate the language learning process, which implies that children *do* acquire language through learning. Given the extremely complex and vast language structure that needs to be acquired, the question arises as to what such a learning process would look like.

Deacon holds that symbolic relationships, given their structure, are difficult to learn. Furthermore, similar difficulties are posed by learning the logic of grammar and syntax, especially because these aspects of language are also the surface expressions of the *deep web of symbolic relationships*. Deacon argues that the "language learning period" in which children acquire language with amazing facility corresponds to certain deficiencies in children's learning capabilities. Children have a short attention span, they have difficulty in memorising associations, and a relatively brief period of working memory. However, he speculates that that the limitation to children's learning abilities during this period may provide the constraints and biases in learning that are necessary to facilitate rapid language learning. As argued, Deacon presents us with a conception of symbolic associations as a hierarchic network of connected nodes that make up a semantic space that is continually changing (similar to the models of neural networks discussed in Chapter 4). One of the most important aspects of this network of connected nodes is the hierarchies present in it and the constraints that come about as a result of the combinatorial patterns that make up the semantic space.

Deacon defines language, as a system of associations with a deep logic that derives from the indirect systemic logic of symbolic reference. It is highly distributed and non-local and the rules for implementing these relationships tend to form hierarchic patterns. Taking these characteristics into consideration, Deacon contends that the way in which children learn language is akin to the way in which neural networks are trained by cognitive scientists (see Chapter 4). Thus, language learning is not the "top-down" approach of inductive logic, but something like the pattern recognition abilities such as those exhibited by neural networks. The suggestion is that, just like neural networks that have been presented with biases to facilitate their learning (see Chapter 5), children with their limited attention spans and limited memories for associations and details would disregard immediate (local) information by default, and would thus attend to more general (global) patterns in the sentences that they are

being presented with. In this manner, children would be intent on the large-scale logic of language, the background of other details being too variable for them to follow. They would thus be focussed on syntactical and grammatical patterns rather than particular word associations. Paradoxically, their learning deficiencies give them a head start in learning language. Furthermore, symbolic associations, which are quintessential examples of highly distributed relationships, would be acquired in a similar manner. It requires postponing commitment to immediately obvious associations, until less obvious distributed relationships are acquired.

The implication here is that acquiring the shift to symbolic mnemonic strategies and acquiring grammar are intertwined learning problems, which are more easily surmounted by learners who see the bigger picture, while losing track of the details. Deacon concludes that it is unsurprising that the optimal time for learning grammatical and syntactic regularities in language coincides with the period when symbolic reference is first discovered. Immature brains are at an advantage in the first steps toward overcoming the complex learning problem of shifting from the indexical to the symbolic referential strategy.

In terms of the manifestation of symbolic reference in the brain, Deacon argues that brain structures necessary for the analysis of symbols are distributed across many areas of the brain. However, the systemic character of symbolic reference would suggest that, even though the representation of symbolic associations is distributed, similar classes of words ought to share neural commonalities. Symbolic referential relationships would be produced by the convergence of different neural codes from various brain systems (similar to the distribution of "weights" through neural networks – see Chapter 4). His position is that that word (symbol) comprehension and retrieval processes are the results of combinations of simpler associative processes (indices). This implies that as supportive representational relationships are activated by the process of producing or comprehending a word, each corresponding neural substrate is activated in different phases of the process.

Seeing that the symbolic representational relationship is hierarchic and is constructed from lower levels of representation (from icon to index to symbol), Deacon speculates that their neural representation would have a hierarchic structure as well. As the lowest level of representation, iconic relationships would map onto processes with single sensory modalities (e.g. visual or olfactory input). Similarities between phonemes or visual experiences in different contexts would be examples of such iconic relationships. Deacon speculates that simple iconic processes would probably be highly localised, and perhaps represented in the cerebral cortex, whereas words and familiar objects might require compound iconic analysis involving more than one sensory dimension or modality. He believes that in most practical contexts only a few features would need to be assessed in order to recognise an icon; the link between iconic features is indexical (see Chapter 1). These iconic associations may provide the indexical associations of words. However, he argues that their symbolic association – meaning – involves "something more". This something more would include the associative relationships between words and the logic of how they map onto indexical relationships. Indexical relationships are recoded into symbolic relationships, which create a mnemonic shortcut that allows for information transmission and reception to be accelerated and compressed. The ignored indexical

relationships still implicitly ground word reference, and can be invoked if necessary. Neurologically speaking, Deacon speculates that there may be "convergence" zones within the cerebral cortex where sensorimotor traces of experiences might be interlinked, hence, where images and experiences may be linked with particular word sounds.

Deacon's theory of how the world comes to be "represented" in the brain sheds some light on why questions of mind and language seem to be inextricably linked, both in our common-sense intuition and in more formal theorising on the matter. Furthermore, if he is indeed correct that the human mind as we know it today is the result of our ancestors hitting upon a novel evolutionary niche and learning how to make use of symbolic reference, his theory presents us with a novel way of addressing the apparently mysterious character of mental phenomena. His theory supports the intuition that mental phenomena are different from material phenomena, without implying that we need to dismiss mental phenomena as insurmountably mysterious. In fact, he raises the possibility that "irreducibly" mental phenomena can be understood in naturalistic (scientific) terms. Central to this possibility is the conception of mental phenomena as computational, in that they are the result of the information processing of an organism that has to navigate the world in terms of the information available to it, with the goal of surviving and reproducing. Furthermore, it requires us to conceive of human language structure and the mental processes that support it as inherently symbolic – not symbolic in the Turing-Fodor sense (see Chapter 4) but in the Peirce-Deacon sense (Chapters 1 and 5).

Thus, the argument is that the ability to think symbolically is central to the ability to form intricate, novel, and abstract representations of the world. It is the contention here that these representations constitute what we generally refer to as the "mental". Because only human beings are capable of acquiring the ability to think symbolically, only human beings experience mental phenomena. We alone are able to experience a world of beliefs, hopes, fears, and desires that are not fundamentally linked to our immediate experiences. Symbolic reference affords us access to the world of the imagination. With this in mind, it becomes possible to address the philosophical questions of dualism and other minds, for example.

In a cognitive structure based on the learned ability to think symbolically, both subjective and objective experience become instances of symbolic representation. Both our conceptions of ourselves and of other people are symbolic constructs, based on the terms of which we process the information that we gain both from the apparently "external" and "internal" worlds. As both Pinker and Churchland argued, we construct a folk psychology regarding other people in order to interpret and anticipate their behaviour and facilitate our navigation of the world. Folk psychology essentially assumes that other people have minds similar to ours which give rise to habits, beliefs and desires similar to ours. This belief is generally born out by the fact that our anticipations so often turn out to be correct. If other people did not have minds similar to our own, there would be little reason to believe that we can ascribe motivations and actions to them with any accuracy. The possibility that the agents that we are confronted with are zombies and mechanically act in a manner that mimics those of people with minds disregards the fact that we attribute the characteristic of being a "zombie" precisely on the ground that such entities do *not* act in a way that we would expect entities with minds to act (the same

consideration applies to the leeway that we would grant children or animals, for example). They are fundamentally mind-less. Ultimately, there is no way to answer the thought experiment that holds that we may be duped into attributing mental characteristics to machines or similar entities that in *every way* act as we would expect an entity with a mind to act, and that report first-person experiences. For all intents and purposes such "machines" would have minds, just as we are "machines with minds."

Folk psychology is applied relatively unreflective. However, when the need arises, we subject our own motivations and those of others to intense scrutiny, through constructing a model in terms of which those motivations can be analysed. Deacon (1997a:426) argues that the ability to take the others' perspectives is not inborn, but is ingrained by means of learning culture, and particularly language. While there are arguments to be made for at least some aspects of our folk psychology to be innate (Pinker's position), Deacon's contention seems extremely plausible. Deacon also suggests that much of folk psychology and animal "theories of mind" might be based on indexical experience, where repeated correlation between given behaviours results in an indexical interpretation of behaviour. However, the ability to formulate an abstract and "not-yet-experienced" conception of the motivations, desires, etc. of other people, requires the ability to formulate symbolic anticipations. This ability arises from the ability to spot, extricate, and represent regularities from indexical experience (also indirect indexical experience, through stories for example) and to make novel attributions and anticipations on these grounds. Hence, the ability to interpret experience from a novel and different perspective is based on the ability to think (and represent models) symbolically.

A similar process takes place in accounting for one's own beliefs, motivations, and desires, etc. On the one hand they can be described as "unique" and "mental" because they are based on *a particular* organism's information processing processes, in a particular context and geared towards particular goals that are wholly focussed on that organism and its context. However, to conclude that such mental entities are accessible through introspection is erroneous. A similar symbolic model needs to be constructed to interpret one's own mind, as for the minds of others (as Freud already made clear with his "discovery" of the unconscious). Furthermore, one might have a convenient folk psychological theory in terms of which to interpret one's own behaviour which may be useful most of the time, but which is not necessarily accurate.

The notion of free will can also be addressed through the ability to construct symbolic models of possible future worlds. Human beings' actions and reactions are not completely determined by their (indexical) experience of the world. We can construct and review possible scenarios and future consequences without having to experience them first. In this manner it becomes possible for humans to override the behaviourists' envisioned stimulus-response interaction with the world. Such behaviourism can also be part of our interaction with the world; however, our ability to represent alternative courses of actions (thus, to use *reason*) undermines any simplistic deterministic conception of human nature.

In a very minimalistic sense it might be possible to conceive of symbolic reason also as determining action. However, if one is able to generate different possible courses of action, and then choose one in terms of the goals that one hopes to achieve (which are ultimately guided by our overriding desire to survive and reproduce), one's action cannot be determined by the bogey-man of

mechanical natural laws. For those proponents of free will that require that we are in no way "caused" to take one given course of action instead of another might argue that such a conception of "free will" is not free enough. However, if absolutely nothing were to cause one to take one course of action rather than another, one might very well ask "Why act at all?" One has to conceive of some cause of action and decisions, in the sense that action requires a reason to act. There has to be some goal that needs to be achieved, otherwise there would be no way in which to determine whether a given action is "reasonable", and not just entirely random. Hence, an evolutionary-symbolic-computational approach to free will provides i) a problem horizon (a reason to act); ii) a way to represent alternative courses of action (thus a non-deterministic domain in which to act), and iii) a way to determine whether action is random, or can be attributed to a mind (thus it allows for reason, in terms of applicability to goal that needs to be reached). I would argue that this approach thus provides for free will. Furthermore, it provides for free will within a universe wholly subject to natural laws and with no recourse to obscure "mentality".

With regard to that old favourite of philosophy of mind, qualia, I would argue that it can generally be explained in the above terms as well. The fact that there is something "that it is like" associated with mental phenomena can partly be explained in terms of the self-reflective nature of our symbolic models of the world. Mental representations are symbolic representation encoded into a model of the world that incorporates the organism/person's experiences into that model. In this process, the iconic and indexical representations that are based on "first-hand" experience (thus information from the organism's own senses rather than from symbolic sources such as language) are incorporated into the organism's theoretical and symbolic construct of the world. I would argue that those aspects of mental phenomena that are linked to first-hand iconic and indexical experiences, rather than on the theoretical constructs that we create from these icons and indices, constitute those aspect of the mental realm that we usually refer to as "qualia". This should explain the different "feel" that some mental phenomena have. Such mental phenomena are different from other representations of the world in that they have to do with the actual sensual experiences of a given organism. Qualia are thus not mysterious or irreducibly "mental". They are reducible to personal experiences that have been encoded in neural information and incorporated into a representational construct of the world that includes the organism that is generating the construct (the self) into that picture.

Part of this thesis has been dedicated to the attempt to indicate that it is possible to account for the mind and mental phenomena in naturalistic terms. The argument is that this is possible, without reducing the mind to the classical conception of a fundamentally determined material entity. A conception of a material mind is possible that still allows for such ostensibly "immaterial" phenomena as free will, beliefs, desires, and qualia. Although this analysis is not presented as conclusive or complete, it is presented as wholly possible and plausible, within a legitimate materialistic scientific framework.

In summary, we saw that the shift towards studying the concept of information through the computational sciences was crucial for redefining the common-sense conception of what constitutes mind and matter. A fundamentally mechanical conception of the material world is no longer tenable in a world where seemingly disembodied and "immaterial" computations have significant real-world

effects. The possibility arose that the seemingly immaterial mind would become known through scientific experiment in the laboratory, rather than through the *a priori* analyses of the philosophers. However, as we have seen, such hopes were overly ambitious and premature. The computational approach to the mind was still firmly under the influence of a philosophical mind-body dualism. Furthermore, information-processing in the mind was equated with a contemporaneous conception of language processing in the mind, which were on the whole radically speculative and theoretical. Generally, other practical dimensions of meaning have been regarded as slightly disreputable, secondary elements, too far removed from the pure theory of representation. These factors led to the material aspects of computation being disregarded – to the detriment of the models being produced within the discipline. In terms of prevailing conventions, it seemed impossible to overcome the quandaries that "science of mind" found itself in – a situation which appeared to justify the continuation of philosophical approaches to mind and language that took very little notice of developments in the sciences of mind. It seemed obvious that the mental comprised elements that were not amenable to scientific study, and hence required a discipline dedicated to analysing the "metaphysical" aspects of mind.

However, the metaphysical aspects of mind are nothing more than a formalisation of our common-sense assumptions about what the mental "is like", as Nagel puts it. We have discussed Chomsky's, Pinker's, and Churchland's scepticism with regard to the value of general common-sense understandings of phenomena in that most accepted scientific hypotheses (e.g. quantum theory) are radically different from our common-sense understanding of the world. It has become apparent that our perceptions and knowledge are fundamentally structured by the organisation, orientation, and purpose of our cognitive structure, as argued in one way or another by all of the theorists that we have studied. Thus there is no reason to believe that the world accords with our seemingly "immediate" perceptions (or introspections) of it. Furthermore, the methodology in terms of which mind (and language as it relates to mind) has been studied, amounts to little more than a contemporary scholasticism, based on a tradition of interpretation and reinterpretation of certain canonical ideas, as has been shown with our discussion on Chisholm's apparent misrepresentation of Brentano's conception of intentionality, as well as the fundamental misunderstanding of what Turing envisaged with the universal Turing machine. Hence, the argument here is that the study of mind needs to be tempered by empirical – non-common sensical – knowledge as far as possible, as well as be subjected to critical historical oversight, to root out outdated conventions within the inherited body of theory. There seem to be many obstacles to such a move, however, ranging from the resistance of some of those theorists dedicated to the study of mind and language, to the radically interdisciplinary requirements of such an approach.

With the move towards evolutionary analyses of language and mind, the possibility has arisen that the mind might well be amenable scientific study. Not in terms of the "pure" physics that theorists like Chomsky would seem to favour, but in terms of the biological or life sciences. As Deacon's work indicates, viewing the mind as a biological, information-processing entity, the primary purpose of which is to ensure the survival and reproduction of its possessor, offers a way out of many apparent theoretical quandaries. This is even more so when language is approached as a product of evolution

that also contributes to the structure of the mind. It seems more and more probable that the problems of philosophy of mind as traditionally conceived – explaining the "phenomenal feel" of states of mind, the apparent lack of spatial location of states of mind, the intentionality of mental states, the counter positioning of introspection and perception, and the apparent inaccessibility of the mental are, in fact, pseudo-problems, rooted in conventional "common-sensical" assumptions about mind rather than in actual attributes of the mental. If the mind, as "something that the brain does", is the product of computational processes where the human organism gains information about the world in various modalities through its senses, and structures and interprets that information in terms of the inherent structure of the brain and symbolic thought provided by language learning, the evolutionary goals of the organism, and past experiences, there is no reason to assume that our "introspective knowledge" is any less subject to interpretation and "material interference" than knowledge that we gain through "perception". There is also no reason to believe that mental knowledge or states are more easily accessible or more trustworthy than "material knowledge". Hence, there is no reason why the mental should not be studied through a radically interdisciplinary science of mind.

The argument here is not that an evolutionary approach to mind and language, and particularly a co-evolutionary theory in the spirit of that by Deacon, is fundamentally correct and incontrovertibly proven. The argument is that, in terms of Russell's philosophical scepticism and the "scientific temper of our age", there can be no justification for dismissing such evolutionary approaches out of hand, especially if such a dismissal is based on unjustified philosophical allegiances to an *a priori* and metaphysical tradition. There is no reason why the mental needs to be restricted to *a priori* musings only when there is the possibility that it can also be approached through empirical study. The role of the philosopher in this regard should not be to act as the keeper of esoteric and exclusive wisdom regarding the mystical realm of the mental, but to critically evaluate the scientific paradigm in which the mind is studied, critiquing theoretical assumptions and allegiances and ensuring that preconceived prejudices do not lead to untenable, but comfortable "certainties". Furthermore, the philosopher, more than any other theorist, would have to keep in mind the caveat raised by Russell that we are unable to know the characters of those parts of the universe that are remote from our experience, and entertain the possibility that there are some aspects of the universe that we may never know. The role of philosophy would not be to "fill in the blanks" through metaphysical speculation, but to ensure that theoretical explanations retain some measure of scientific credibility.

BIBLIOGRAPHY

Arbib, Michael A. (2003). The Evolving Mirror System. *In*: Morten H. Christiansen and Simon Kirby eds. (2003). *Language Evolution* Oxford: Oxford University Press.

Aronson, Jerrold L. (1969). Connections: A Defence of Peirce's Category of Thirdness. *Transactions of the Charles S. Peirce Society*, Summer (5), 158-172.

Austin, John L. (1980). *How to do Things with Words*. Oxford and New York: Oxford University Press.

Auyang, S. Y. (1998). Foundations of Complex-System Theories in Economics, Evolutionary Biology, and Statistical Physics. Cambridge: Cambridge University Press.

Bak, Per (1997). How Nature Works. The Science of Self-Organised Criticality. Oxford: Oxford University Press.

Bernstein, Richard J. (1965). Action, Conduct, and Self-Doubt. *In*: Richard J. Bernstein, ed. *Perspectives on Pierce: Critical Essays on Charles Sanders Peirce*. New Haven and London: Yale University Press.

Bickerton, Derek (2003). Symbol and Structure: A Comprehensive Framework for Language Evolution. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution* Oxford: Oxford University Press.

Blackburn, Simon (1995) Language, history of the philosophy of. *In*: Ted Honderich, ed. *The Oxford Companion to Philosophy*. Oxford and New York: Oxford University Press.

Block, Ned (2002[1978]). Troubles with Functionalism. *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings.* Oxford: Oxford University Press.

Botha, Rudolph (2004). Windows with a view on language evolution. *European Review* 12 (2), 235-243.

Bracken, Harry M. (1994). Chomsky's variation on a theme by Descartes. *In*: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge.

Brentano, Franz (2002). The Distinction between Mental and Physical Phenomena. *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings*. Oxford: Oxford University Press.

Burks, Arthur W. (1980). Man: Sign or Algorithm? A Rhetorical Analysis of Peirce's Semiotics. *Transactions of the Charles S. Peirce Society*. Fall (16), 279-292.

Burrow, J. W. (1968). Introduction. In: Charles Darwin. The Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life. London: Penguin Books.

Carnap, Rudolf (1967 [1928])). *The Logical Structure of the World*. Rolf A. George, trans. Berkley: University of California Press.

Carnap, Rudolf (2002 [1933]). Psychology in Physical Language. *In*: David J. Chalmers, ed. Philosophy of Mind. *Classical and Contemporary Readings*. Oxford: Oxford University Press.

Chalmers, David J. (1995). The Puzzle of Conscious Experience. *Scientific American*. CCLXXIII, 80-86.

Chalmers, David J. (2002a). Foundations. *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings*. Oxford: Oxford University Press.

Chalmers, David (2002b). Consciousness and Its Place in Nature. *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings*. Oxford: Oxford University Press.

Chisholm, Roderick M. (2002[1957]). "Intentional Inexistence". *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings*. Oxford: Oxford University Press.

Chomsky, Noam (1957). *Syntactic Structures.* The Hague: Mouton and Co.

Chomsky, Noam (1966). *Cartesian Linguistics. A Chapter in the History of Rationalist Thought.* New York and London: Harper & Row.

Chomsky, Noam (1972). *Language and Mind.* New York: Harcourt Brace Jovanovich, Inc.

Chomsky, Noam (1975). *Reflections on Language.* New York: Pantheon House.

Chomsky, Noam (1977). *Essays on Form and Interpretation.* New York: North-Holland.

Chomsky, Noam (1980). *Rules and Representations*. Oxford: Basil Blackwell.

Chomsky, Noam (1986). Knowledge of Language: it's Nature, Origin, and Use. London: Praeger.

Chomsky, Noam (1994). Chomsky, Noam. *In*: Samuel Guttenplan, ed. *A Companion to the Philosophy of Mind*. Oxford: Blackwell Publishers.

Chomsky, Noam (1995a). Language and Nature. *Mind,* 104 (413), 1-61.

Chomsky, Noam (1995b) *The Minimalist Program.* Cambridge, Massachusetts: MIT Press

Chomsky, Noam (2000a). *The Architecture of Language*. Mukherji, N, Patnik, B. N. and Agnihotri, R. K. eds. New Delhi: Oxford University Press.

Chomsky, Noam (2000b). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, Noam (2004). The Generative Enterprise Revisited. Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi. Berlin and New York: Mouton de Guyter.

Christiansen, Morten H. and Simon Kirby (2003a). Language Evolution: Consensus and Controversies. *TRENDS in Cognitive Sciences.* 7 (7), 300-307.

Christiansen, Morten H. and Simon Kirby (2003b). The Hardest Problem in Science? *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution.* Oxford: Oxford University Press.

Churchland, Paul M. (2002[1981]). *Eliminative Materialism and Propositional Attitudes. In*: David J. Chalmers, ed. Philosophy of Mind. Classical and Contemporary Readings. Oxford: Oxford University Press.

Cilliers, Paul (1989). Brain, Language and Consciousness. A Post-Structural Neuropsychology. Thesis delivered for the degree of Master of the Arts at the University of Stellenbosch.

Cilliers, Paul (1998). Complexity and Postmodernism. Understanding Complex Systems. London and New York: Routledge.

Clark, Andy (2001). Reasons, Robots and the Extended Mind. *Mind and Language*. 16 (2), 121-145.

Cohen, Morris R. (1998). Introduction. *In*: Charles S. Peirce. *Chance, Love, and Logic.* Morris R. Cohen, ed. Lincoln and London: University of Nebraska Press.

Colapietro, Vincent (1987). Is Peirce's Theory of Signs Truly General? *Transactions of the Charles S. Peirce Society.* 87 (23), 205-234.

Copeland, Jack (2000a). What is a Turing Machine? *AlanTuring.net* Available from: http://www.alanturing.net/turing_archive/pages/Reference.html [Accessed 14 August 2006].

Copeland, Jack (2000b) The Church-Turing Thesis. *AlanTuring.net* Available from: http://www.alanturing.net/turing_archive/pages/Reference.html [Accessed 14 August 2006].

Corballis, Michael C. (2003). From Hand to Mouth: The Gestural Origins of Language. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution.* Oxford: Oxford University Press.

Coveney, Peter and Roger Highfield (1996). *Frontiers of Complexity. The Search for Order in a Chaotic World.* Farber and Farber: London.

Cowley, Stephen J (2002). Why brains matter: an integrational perspective on The Symbolic Species. *Language Sciences* 24, 73-95.

Critchley, Simon (1996). Derrida: Private Ironist or Public Liberal? *In*: Chantal Mouffe, ed. *Deconstruction and Pragmatism*. London and New York: Routledge.

Culler, Jonathan (1994). On Deconstruction. Theory and Criticism after Structuralism. London: Routledge.

Davidson, Iain (2003). Archaeological Evidence. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution* Oxford: Oxford University Press.

Darwin, Charles (1985). The Origin of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life. London: Penguin Books.

Dawkins, Richard (1976). *The Selfish Gene.* Oxford: Oxford University Press.

Dawkins, Richard (1999). *The Extended Phenotype. The Long Reach of the Gene*. Oxford: Oxford University Press.

Deacon, Terence (1997a). The Symbolic Species. The co-evolution of language and the human brain. Penguin: London.

Deacon, Terrence (1997b). What makes the Human Brain Different? *ProQuest Social Science Journals. Annual Review of Anthropology*. 26, 337-357.

Deacon, Terrence (2000). Evolutionary perspectives on language and brain plasticity. *Journal of Communication Disorders.* 33 (4), 273-291.

Deacon, Terence (2003). UG and Semiotic Constraints. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution.* Oxford: Oxford University Press.

Deacon, Terrence (2004). Monkey homologues of language areas: computing the ambiguities. *TRENDS in Cognitive Sciences*. 8 (7), 288-290.

Dennett, Daniel C. (1991). *Consciousness Explained.* London: Penguin.

Dennett, Daniel C. (1994) Intentional Systems. *In*: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Vermont: Bradford Books.

Dennett, Daniel C. (1995). Darwin's Dangerous Idea. Evolution and the Meanings of Life. New York: Simon and Schuster.

Dennett, Daniel C. (2004). *Freedom Evolves*. London: Penguin Books.

Derrida, Jacques (1976). *Of Grammatology*. Baltimore and London: Johns Hopkins University Press.

Derrida, Jacques (1978). *Writing and Difference* Alan Bass, transl. Chicago: The University of Chicago Press.

Derrida, Jacques (1981a) *Dissemination*. Chicago: The University of Chicago Press.

Derrida, Jacques (1981b) *Positions* Chicago: University of Chicago Press.

Derrida, Jacques (1986). *Margins of Philosophy.* Sussex: The Harvester Press.

Derrida, Jacques (1996). Remarks on Deconstruction and Pragmatism. In: Chantal Mouffe, ed. *Deconstruction and Pragmatism.* London and New York: Routledge.

Derrida, Jacques (1997a). Signature Event Context. *In*: Jacques Derrida *Limited Inc.* Evanston, IL: Northwestern University Press.

Derrida, Jacques (1997b). Limited Inc a b c… *In*: Jacques Derrida. *Limited Inc.* Evanston, IL: Northwestern University Press.

Derrida, Jacques (1997c). Afterword. *In*: Jacques Derrida. *Limited Inc.* Evanston, IL: Northwestern University Press.

Delaney, C. F. (1992). Peirce on the Hypothesis of God. *Transactions of the Charles S. Peirce Society.* 28 (4), 725-739.

Deledalle, Gérard (1992) Peirce's "Sign": Its Concept and Its Use. *Transactions of the Charles S. Peirce Society*. 28 (2), 289-301.

Dreyfus, Herbert, L. (1981). From Micro-Worlds to Knowledge Representation. *In*: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

Dunbar, Robin I. M. (2003). Origin and Evolution of Language. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution* Oxford: Oxford University Press.

Dupuy, Jean-Pierre (2000). *The Mechanization of Mind. On the Origins of Cognitive Science*. Princeton and Oxford: Princeton University Press.

Eco, Umberto (1990). *The Limits of Interpretation*. Bloomington and Indianapolis: Indiana University Press.

Fitzgerald, John J (1976). Ambiguity in Peirce's Theory of Signs. *Transactions of the Charles S. Peirce Society*. 76 (12), 127-134.

Flanagan, Owen (1991). *The Science of Mind.* Cambridge, Mass.: MIT Press.

Flanagan, Owen (1995). The History of the Philosophy of Mind. *In*: Ted Honderich, ed. *The Oxford Companion to Philosophy*. Oxford and New York: Oxford University Press.

Fodor, Jerry A. (1975). *The Language of Thought.* New York: Cromwell.

Fodor, Jerry A. (1990). Why There Still Has to Be a Language of Thought. In: William G. Lycan, ed. *Mind and Cognition: A Reader.* Cambridge, Massachusetts: Basil Blackwell.

Fodor, Jerry (1998). The Trouble with Psychological Darwinism. *In*: London Review of Books 20(2).

Fodor, Jerry (2000). The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology. Cambridge, MA: MIT Press.

Fodor, Jerry (2005). Reply to Steven Pinker 'So How *Does* the Mind Work?' *Mind and Language.* 20 (1), 25-32.

Freud, Sigmund (1901). *The Interpretation of Dreams.* Standard Edition, Volumes 4-5. London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, Sigmund (1925). *A Note upon the "Mystic Writing-Pad."* Standard Edition, Volume 19. London: Hogarth Press and the Institute of Psycho-Analysis.

Freud, Sigmund (1950 [1895]). *Project for a Scientific Psychology*. Standard Edition, Volume 1. London: Hogarth Press and the Institute of Psycho-Analysis.

Fukui, Naoki and Mihoko Zushi (2004). Introduction. In: Noam Chomsky. The Generative Enterprise Revisited. Discussions with Rina Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi. Mouton de Gruyter: Berlin.

Gallie W. B. (1966). *Peirce and Pragmatism.* New York: Dover Publications, Inc.

Gell-Mann, Murray (1994). The Quark and the Jaguar. Adventures in the Simple and the Complex. London: Little Brown Company.

Glendinning, Simon (2004). Language. *In*: Jack Reynolds and Roffe, Jonathan eds. *Understanding Derrida*. New York and London: Continuum.

Gödel, Kurt (1931). On Formally Undecidable Propositions of Principia Mathmatica and Related Systems. Meltzer, B. trans. 1962. New York: Basic Books Inc.

Goodenough, Ursula and Terrence W. Deacon (2003). From Biology to Consciousness to Morality. *Zygon.* 38 (4), 801-819.

Goodman, Nelson (1955). *Fact, Fiction, and Forecast*. London: University of London.

Goudge, Thomas, A. (1950). *The Thought of C. S. Peirce*. New York: Dover Publications, Inc.

Gould, Stephen, J. and Lewontin, Richard, C. (1979). The Spandrels of San Marco and the Panglossian Programme: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London.* 205 (1161), 281-288.

Gould, Stephen, J. and Vrba, Elizabeth (1981). Exaptation: A Missing Term in the Science of Form. *Paleobiology*. 8, 4-5.

Gould, Stephen J. (1980). Is a New and General Theory of Evolution Emerging? *Paleobiology*. 6, 110-130.

Gould, Stephen J. (1982). Darwinism and the Expansion of Evolutionary Theory. *Science*. 216, 380-387.

Gould, Stephen (1989). *Wonderful Life. The Burgess Shale and the Nature of History*. New York: Norton.

Gould, Stephen J. (1991). *Bully for Brontosaurus*. New York: Norten.

Gould, Stephen J. (1992a). *The Confusion of Evolution*. New York Review of Books, November 19, 47-54.

Gould, Stephen J. (1992b). Life in a Punctuation. *Natural His*tory. 101, 10-21.

Gould, Stephen, J. (2000) "More Things in Heaven and Earth" In: *Alas, Poor Darwin.* London: Random House.

Greenlee, Douglas (1976). Peirce's Concept of the Sign: Further Reflections. *Transactions of the Charles S. Peirce Society*. 76 (12), 135-147.

Gregory, Richard L. (1998). Turing. *In*: Richard L. Gregory, ed. *The Oxford Companion to the Mind*. Oxford and New York: Oxford University Press.

Hanson, Norwood, R. (1965). Notes Toward a Logic of Discovery. *In*: Richard J. Bernstein, ed. *Perspectives on Pierce: Critical Essays on Charles Sanders Peirce*. New Haven and London: Yale University Press

Hart, Kevin (2004) Religion. In: Jack Reynolds and Roffe, Jonathan eds. *Understanding Derrida*. New York and London: Continuum.

Haugeland, John (1981a). Semantic Engines: An Introduction to Mind Design. In: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

Haugeland, John (1981b). The Nature and Plausibility of Cognitivism. *In*: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Vermont: Bradford Books.

Hauser, Mark D. and Tecumseh W. Fitch (2003). Uniquely Human Components. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution* Oxford: Oxford University Press.

Hauser, Mark D., Noam Chomsky and Tecumseh W. Fitch (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*. 298, 1569-1579.

Hausman, Carl (1998). Infinitesimals as Origins of Evolution: Comments Prompted by Timothy Herron and Hilary Putnam on Peirce's Synechism and Infinitesimals. *Transactions of the Charles S. Peirce Society*, 34 (3), 627-640.

Hayles, Katherine (1999). *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature*, and Information. Chicago and London: Chicago University Press.

Herbenick, Raymond, M. (1970). Peirce on Systems Theory. *Transactions of the Charles S. Peirce Society,* 6, 84-98.

Hofstadter, Douglas, R. (1984). *Gödel, Escher, Bach: An Eternal Golden Braid*. Harmondsworth: Penguin Books.

Holland, John (1998). *Emergence: From Chaos to Order.* Massachusetts: Helix Books.

Hopcroft, John, E. and Ullman, Jeffrey, D. (1969) *Formal Languages and Their Relation to Automata*. Massachusetts: Addison-Wesley Publishing Company.

Hurford, James R. (2003). The Language Mosaic. *In*: Morten H. Christiansen and Simon Kirby eds. *Language Evolution* Oxford: Oxford University Press.

James, William (1890). *The Principles of Psychology.* London: McMillan.

Jackson, Frank (1982). Epiphenomenal Qualia. *Philosophical Quarterly*. XXXII (127), 127-136.

Johnson, Barbara (1981). Translator's Introduction. *In*: Jacques Derrida, *Dissemination.* Chicago: University of Chicago Press.

Johnson, Christopher (1993). *System and Writing in the Philosophy of Jacques Derrida*. Cambridge: Cambridge University Press.

Kampf, Louis (1994). Review of Chomsky, Cartesian Linguistics. In: *Noam Chomsky. Critical Assessments*. Carlos P. Otero, ed. London: Routledge.

Katz, Jerrold and Fodor, Jerry (1994). What is wrong with the philosophy of language? *In*: *Noam Chomsky. Critical Assessments*. Carlos P. Otero, ed. London: Routledge.

Katz, Jerrold J. (1994). Mentalism in Linguistics. *In*: *Noam Chomsky. Critical Assessments*. Carlos P. Otero, ed. London: Routledge.

Kauffman, Stuart A. (1993). The Origins of Order: Self-organization and Selection in Evolution. New York: Oxford University Press.

Kenny, Anthony (1994). Language and the Mind. *In*: *Noam Chomsky. Critical Assessments*. Carlos P. Otero, ed. London: Routledge.

Keslo, Scott J. A. (1999). Dynamic Patterns: The Self-organisation of Brain and Behaviour. Cambridge: MIT Press.

Komarova, Natalie L and Nowak, Martin A. (2001). Natural selection of the critical period for language acquisition. *Proc. R. Soc. Lond*. 268, 1189-1196.

Kuhn, Thomas S. (1970). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. (1959). What the Frog's Eye Tells the Frog's Brain. *In*: William C. Corning and Martin Balaban eds. (1968). *The Mind: Biological Approaches to its Functions*

Lieberman, Philip (2003). Motor Control, Speech, and Evolution. *In*: Christiansen, Morten H. and Kirby, Simon eds. *Language Evolution.* Oxford: Oxford University Press.

Liszka, James Jakób (1990). Peirce's Interpretant. *Transactions of the Charles S. Peirce Society.* 26 (1), 7-62.

Livingston, Paul (2002). Experience and Structure. Philosophical History and the Problem of Consciousness. *Journal of Consciousness Studies.* 9 (3), 15-33.

Love, Nigel (2004). Cognition and the Language Myth. *Language Sciences.* 26, 252-544.

Lumsden, David (2002). Crossing the symbolic threshold: a critical review of Terrence Deacon's The Symbolic Species. *Philosophical Psychology*. 15 (2), 156-171.

Luria, A. R. (1994). Scientific perspectives and philosophical dead ends in modern linguistics. *In*: *Noam Chomsky. Critical Assessments*. Carlos P. Otero, ed. London: Routledge.

Matthews, P. H. (1994). Language as a mental faculty: Chomsky's progress. *In*: *Noam Chomsky. Critical Assessments*. Carlos P. Otero, ed. London: Routledge.

McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*. 7, 115 - 133.

McDermott, Drew (1981). Artificial Intelligence Meets Natural Stupidity. John Haugeland, ed. *In*: *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

McGilvray, James (1999). *Chomsky. Language, Mind, and Politics.* Cambridge: Polity Press.

Melrose, Robin (1995). The seduction of abduction: Peirce's theory of signs and indeterminacy in language. *Journal of Pragmatics.* 23, 493-507.

Melvil, Yuri, K. (1966). The Conflict of Science and Religion in Charles Peirce's Philosophy. *Transactions of the Charles S. Peirce Society*, 2, 33-50.

Meyers, Robert G. (1992). Peirce's New Way of Signs. *Transactions of the Charles S. Peirce Society.* 28 (3), 505-521.

Minsky, Marvin (1981). A Framework for Representing Knowledge. *In*: John Haugeland, ed *Mind Design. Philosophy, Psychology, Artificial Intelligence*. . Montgomery, Vermont: Bradford Books.

Moravcsik, J. M. E. (1994). Linguistic theory and the philosophy of language. *In*: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge.

Morten, Adam (1995). Turing Machine. *In*: Ted Honderich, ed. *The Oxford Companion to Philosophy*. Oxford and New York: Oxford University Press.

Mukherji, Nirmalangshu et al. (2000) Editors' Preface. *In*: N. Mukherji, N, B. N. Patnik, and R. K. Agnihotri, eds. *In The Architecture of Language*. New Delhi: Oxford University Press.

Nagel, Thomas (1982). What is it Like to be a Bat? *In*: D. R. Hofstadter and D. C. Dennett. *The Mind's I. Fantasies and Reflections on Self and Soul*. London: Penguin

Nagal, Thomas (1993). The Mind Wins! *In*: *The New York Review of Books*. March 4 1993.

Nathan, P. W. (1998). Nervous System. *In:* R. L. Gregory, ed. *The Oxford Companion to the Mind.* Oxford: Oxford University Press.

Newell, Allen and Simon, Herbert, A. (1981). Computer Science as Empirical Inquiry: Symbols and Search. *In*: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

Newmeyer, Frederick, J. (1994). Has there been a 'Chomskyan Revolution' in Linguistics? *In*: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge

Newmeyer, Frederick J. (2003). What Can the Field of Linguistics Tell Us About the Origins of Language? *In*: Morten H. Christiansen and Simon Kirby, eds. *Language Evolution.* Oxford: Oxford University Press.

Newmeyer, Frederick, J. (2004). Typological evidence and Universal Grammar. *Studies in Language.* 28 (3), 527-548.

Norris, Christopher (2004) Metaphysics. *In*: Jack Reynolds and Jonathan Roffe, eds. *Understanding Derrida*. New York and London: Continuum.

O'Conner, Timothy and Robb, David (2003a). Preface. *In*: Timothy O'Conner and David Robb, ed. Philosophy of mind: contemporary readings. London and New York: Routledge.

O'Conner, Timothy and Robb, David (2003b). Introduction. *In*: Timothy O'Conner and David Robb, ed. *Philosophy of mind: contemporary readings*. London and New York: Routledge.

Otero, Carlos P. (1994a). Chomsky and the rationalist tradition: support for innateness, metaphysics vindicated and a rare kind of intellectual history. *In*: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge.

Otero, Carlos P. (1994b). Prolegomena to a future evolution in metaphysics: the emergence of *homo loquens* and physical law. *In*: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge.

Parker, Kelly (1994). Peirce's Semeiotic and Ontology. *Transactions of the Charles Peirce Society*. 30 (1), 51-75.

Patterson, Collin (1998). Evolution: Neo-Darwinian Theory. *In*: Richard L. Gregory, ed. *The Oxford Companion to the Mind*. Oxford and New York: Oxford University Press.

Patton, Paul (2004). Politics. *In*: Jack Reynolds and Jonathan Roffe, eds. *Understanding Derrida*. New York and London: Continuum.

Peirce, Charles S (1965). *The Collected Papers of Charles Sanders Peirce Vol. I-VI.* Charles Hartshorne and Paul Weiss, eds. Cambridge, Mass: The Belknap Press of Harvard University Press.

Peirce, Charles S (1998). *Chance, Love, and Logic.* Morris R. Cohen, ed. Lincoln and London: University of Nebraska Press.

Penrose, Roger (1989). *The Emperor's New Mind*. Oxford: Oxford University Press.

Peterson, John (1991). Can Peirce be a Pragmatist and an Idealist? *Transactions of the Charles S. Peirce Society*. 221-235.

Picton, Phil (1994). *Introduction to Neural Networks*. London: The McMillan Press Ltd.

Pinker, Steven and Bloom, Paul (1990). Natural Language and Natural Selection. *Behavioural and Brain Sciences.* 13, 707-784.

Pinker, Steven (1995). *The Language Instinct*. New York: William Morrow and Company, Inc.

Pinker, Steven (1997). *How the Mind Works.* New York and London: W. W. Norton and Company.

Pinker, Steven (2002). *The Blank Slate*. The *Modern Denial of Human Nature*. London: Penguin Books.

Pinker, Steven (2003). Language as an Adaptation to the Cognitive Niche. *In*: Morten H. Christiansen and Simon Kirby, eds. *Language Evolution.* Oxford: Oxford University Press.

Pinker, Steven and Jackendoff, Ray (2005). The faculty of language: what's special about it? *Cognition.* 95, 201-236.

Place, U. T. (2002 [1956]). Is Consciousness a Brain Process? *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings*. Oxford: Oxford University Press.

Plato (1981).*The Republic*. Harmondsworth: Penguin Classics

Pollack, Jordan, B. (1989). Connectionism: past, present, future. *Artificial Intelligence Review.* 3 (1), 3-20.

Popper, Karl R. (1966). Of Clouds and Clocks: An Approach to the Problem of Rationality and the Freedom of Man. St Louis, MO: Washington University Press.

Prigogine, I. (1980). From Being to Becoming. Time and Complexity in the Physical Sciences San Francisco: Freeman and Company.

Prigogine, I. and Strengers, I. (1984). *Order Out of Chaos: Man's New Dialogue with Nature.* London: Heinemann.

Putnam, Hilary (1980). *Mind, Language, and Reality. Philosophical Papers, Volume 2.* Cambridge: Cambridge University Press.

Putnam, Hilary (1981). Reductionism and the Nature of Psychology. *In*: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

Pylyshyn, Zenon (1981). Complexity and the Study of Artificial and Human Intelligence. *In*: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

Pylyshyn, Zenon (1994). Rules and representations: Chomsky and representational realism. *In*: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge.

Quine, Willard Van Orman (1960). *Word and Object*. Cambridge Mass.: MIT Press.

Radford, Luis (2004). Rescuing Perception: Diagrams in Peirce's theory of cognitive activity. Available from: http://laurentian.ca/educ/lradford/rescuing.pdf (Accessed 10 August 2005).

Ransdell, Joseph (1976). Another Interpretation of Peirce's Semiotic. *Transactions of the Charles S. Peirce Society.* 12 (2), 97-121.

Raposa, Michael L. (1991). Peirce and Modern religious Thought. *Transactions of the Charles S. Peirce Society*. 27 (3), 341-369.

Ricoeur, Paul (1979). *Main Trends in Philosophy.* London: Holmes and Meier.

Rorty, Richard (1996a). Remarks on Deconstruction and Pragmatism. *In*: Chantal Mouffe, ed. *Deconstruction and Pragmatism*. London and New York: Routledge.

Rorty, Richard (1996b).Response to Simon Critchley. *In*: Chantal Mouffe, ed. *Deconstruction and Pragmatism.* London and New York: Routledge.

Ruse, Michael (1995). Evolution. *In*: Ted Honderich, ed. *The Oxford Companion to Philosophy*. Oxford and New York: Oxford University Press.

Russell, Bertrand (1962). *The Problems of Philosophy*. London: Oxford University Press.

Ryle, Gilbert (2002 [1949]). Descartes' Myth. *In*: David J. Chalmers, ed. *Philosophy of Mind. Classical and Contemporary Readings.* Oxford: Oxford University Press.

Saussure, Ferdinand de (1983). *Course in General Linguistics*. Roy Harris, transl. London: Duckworth.

Searle, John, R. (1981). Minds, Brains, and Programs. In: John Haugeland, ed. *Mind Design. Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books.

Searle, John R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Searle, John R. (1992). *The Rediscovery of Mind*. Cambridge Mass.: MIT Press.

Searle, John, R. (1994). Chomsky's revolution in linguistics. In: Carlos P. Otero, ed. *Noam Chomsky. Critical Assessments*. London: Routledge.

Searle, John R. (1990). Is the brain a Digital Computer? *APA Proceedings*. *64* (3), 21-37.

Sharpe, Mathew (2004). Psychoanalysis. In: Jack Reynolds and Jonathan Roffe, ed. *Understanding Derrida*. New York and London: Continuum.

Short, Thomas L (1981). Semiosis and Intentionality. *Transactions of the Charles S. Peirce Society*. (17), 197-223.

Short, Thomas L (1996). Interpreting Peirce's Interpretant: A Response to Lalor, Liszka, and Meyers. *Transactions of the Charles S. Peirce Society*. 32 (4), 488-541.

Smart, J. J. C. (2002 [1959]). Sensations and Brain Processes. *In*: David J. Chalmers, ed. Philosophy of Mind. *Contemporary and Classical Readings*. Oxford: Oxford University Press.

Smith, John E. (1965). Community and Reality. *In*: Richard J. Bernstein, ed. *Perspectives on Pierce: Critical Essays on Charles Sanders Peirce*. New Haven and London: Yale University Press

Smith, Kenny (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*. 14 (1), 65-84.

Smith, Neil (2000) Foreword. *In*: Neil Smith, ed. *New Horizons in the Study of Language and Mind.* Cambridge: Cambridge University Press.

Spivak, Gayatri Chakravorty (1976). Translator's Preface. *In*: Jacques Derrida. *Of Grammatology.* Baltimore and London: Johns Hopkins University Press.

Thayer, H. S. (1996). Peirce and Truth: Some Reflections. *Transactions of the Charles S. Peirce Society*. 32 (1), 1-10.

Thompson, D'Arcy (1942). *On Growth and Form*. Cambridge: Cambridge University Press.

Tiercelin, Claudine (1998). Peirce's Objective Idealism: A Defence. *Transactions of the Charles S. Peirce Society*. 31 (1), 1-28.

Tiles, Mary E. (1998). Gödel. *In*: Richard L. Gregory, ed. *The Oxford Companion to the Mind*. Oxford and New York: Oxford University Press.

Tomasello, Michael (2003). Different Origins of Symbols and Grammar. In: Morten H. Christiansen and Simon Kirby, eds. *Language Evolution.* Oxford: Oxford University Press.

Tomasello, Michael (2004). What kind of evidence could refute the UG hypothesis? *Studies in Language.* 28 (3), 642-645.

Tooby, John & DeVore, Irven (1987). The reconstruction of hominid behavioral evolution through strategic modelling. *In*: W.G. Kinzey, ed. *The evolution of human behavior: primate models.* New York: SUNY Press.

Toulmin, Stephen. (1990). *Cosmopolis: The Hidden Agenda of Modernity.* The Free Press, New York.

Turing, Alan, M. (1936). On Computable Numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc*. 42 (2), 230-265.

Turing, Alan M. (1948). Intelligent Machinery. *In*: B. Meltzer. and D. Michie, eds. (1969) *Machine Intelligence*. Edinburgh: Edinburgh University Press.

Turing, Alan (1950a). Computing Machinery and Intelligence. *MIND : A Quarterly Review of Psychology and Philosophy*. LIX (236), 433-460.

Turing, Alan M. (1950b). Can a Machine Think? *In*: James R. Newman, ed. (1956) *The World of Mathematics. Volume 4*. Simon & Schuster.

Turing, Alan M. (1952). The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London* 641 (237), 37-72.

Von Bertalanffy, Ludwig (1973). General Systems Theory. Foundations, Developments, Applications. Harmondsworth: Penguin.

Weiss, Paul (1965). Biography of Charles Sanders Peirce. *In*: Richard J. Bernstein, ed. *Perspectives on Peirce: Critical Essays on Charles Sanders Peirce.* New Haven: Yale.

Wells, Rulon (1965) Peirce as an American. *In*: Richard J. Bernstein, ed. *Perspectives on Pierce: Critical Essays on Charles Sanders Peirce*. New Haven and London: Yale University Press.

Wittgenstein, Ludwig (1958). *Philosophical Investigations.* Oxford: Blackwell Publishers.

Wunderlich, Dieter (2004). Author's response: Is there any evidence that refutes the UG hypothesis? *Studies in Language.* 28 (3), 646-647.

Young, P. W. (1998). Memory. *In*: Richard L. Gregory, ed. *The Oxford Companion to the Mind.* Oxford: Oxford University Press.