Of One Mind:
Proposal For A Non-Cartesian Cognitive Architecture

Linda Cochrane

A Thesis
In the Department
of
The Individualized Program

Presented in Partial Fulfilment of the Requirements
For the Degree of
Doctor of Philosophy (Individualized Program) at
Concordia University
Montreal, Quebec, Canada

September 2014

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Linda Cochrane

Entitled: Of One Mind: Proposal for a Non-Cartesian Cognitive Architecture

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Special Individualized Progam)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. S. Shaw

_____ External Examiner
Dr. R. Singh

_____ External to Program
Dr. A. Bale

_____ Examiner
Dr. R. De Almeida

_____ Examiner
Dr. M. Hale

_____ Thesis Supervisor
Dr. C. Reiss

Approved by: _____
Dr. K. Schmitt, Graduate Program Director

September 19, 2014

_____
Dr. P. Wood-Adams, Dean
School of Graduate Studies

ABSTRACT

Of One Mind: Proposal for a Non-Cartesian Cognitive Architecture

Linda Cochrane, PhD
Concordia University, 2014

Intellectually, we may reject Cartesian Dualism, but dualism often dominates our everyday thinking: we talk of "mental" illness as though it were non-physical; we tend to blame people for the symptoms of brain malfunctions in a way that differs from how we treat other illnesses. An examination of current theories of mind will reveal that some form of dualism is not always limited to the non-scientific realm. While very few, if any, cognitive scientists support *mind-body* dualism, those who support the view of the mind as a symbol-manipulator are often constrained to postulate more than one cognitive system in response to the failure of the symbol-system model to account for all aspects of human cognition.

In this dissertation, I argue for an empiricist, rather than a realist, theory of perception, for an internalist semantics, and for a model of cognitive architecture which combines a connectionist approach with highly-specialized, symbolic, computational component which includes functions that provide input to a a causally-inert conscious mind. I reject the symbol-system hypothesis and propose a cognitive architecture which, I contend, is biologically-plausible and more consistent with the results of recent neuroscientific studies. This hybrid model can accommodate the processes commonly discussed by dual-process theorists and can also accommodate the processes which have proved to be so problematic for models based on the symbol-system hypothesis.

ACKNOWLEDGEMENTS

Completion and approval of this thesis was made possible through the enthusiastic and enlightened guidance of my academic supervisor, the inimitable Dr Charles Reiss, who was able to make the experience enjoyable. Acknowledgement is also due to Dr Murray Clarke who was involved during earlier developmental stages of this project.

Others to whom I owe thanks are Dr Ketra Schmitt, Graduate Program Director and Ms Darlene Dubeil, Coordinator, of the Individualized Programs (INDI) department, Dr Dana Isac for her very useful critical analysis especially in the area of linguistic philosophy, Dr Vesselin Pekov for advice during the preparation for the Defence, and Dr Dorothea Bye for her continued support and encouragement during the whole process.

Table of Contents

List of Figures.....viii

List of Tables.....ix

1.Introduction.....1

1.1 Overview.....1
1.2 Background.....1
1.3 Dissertation Thesis.....2
1.4 Dissertation Outline.....4

2.An Internalist Semantics.....6

2.1 Introduction.....6
2.2 Part I – Why Empiricism.....7
2.2.1 Realism and Empiricism.....7
2.2.2 Perception and the Senses.....9
2.2.3 Scientific Theories and Empiricism.....11
2.3 Part II: Why an Internalist Semantics.....12
2.3.1 The Causal Theory of Reference.....14
2.3.2 Internalist and Externalist Semantics.....15
2.3.3 Externalist Theories of Meaning.....15
2.3.4 A linguistic theory of meaning.....15
2.3.5 A psychological theory of meaning.....17
2.3.6 Determination of Extension.....19
2.3.7 Determination of Reference.....20
2.3.8 Meaning of Natural-kind terms.....22
2.3.9 Externalist Semantics and Linguistic Puzzles.....24
2.3.10 An Internalist Theory of Meaning.....28
2.3.11 Reference and Interpretation.....28
2.3.12 Metaphors and Figurative Language.....31
2.3.13 Concepts, Categories, and Meaning.....32
2.3.14 The Term-Reference Link.....33
2.4 PART III: Conclusions.....34

3.Cognitive Architectures.....38

3.1 Introduction.....38
3.2 Part I: Representationalism.....38
3.2.1 Mental Representations.....40
3.2.2 Empirical Evidence for Mental Representations.....43
3.2.3 What are Mental Representations?.....45
3.2.4 Representationalism.....46
3.3 Part II: Rival Cognitive Architecture Models.....49
3.3.1 The Computational Theory of Mind.....49
3.3.2 Connectionism.....57
3.4 Part III: Strengths and Weaknesses.....61

3.4.1 Modularity....61
3.4.2 Propositional Attitudes....68
3.4.3 Language of Thought....74
3.4.4 Natural-language and Cognitive Architecture....76
3.4.5 Levels of Cognitive Architecture....78
3.4.6 Learning....80
3.4.7 Comparison of Cognitive Architectures....81
3.4.8 Systematicity, Productivity, and Compositionality....83
3.4.9 The Intentional Stance....85
3.4.10 Co-extension and Co-reference....86
3.5 Part IV: Conclusions....86

4.Dual-Process Theories of Mind....90

4.1 Introduction....90
4.2 Part I: Dual-process Theories – Standard View....90
4.3 Part II: Problems with the Standard View....92
4.3.1 Conscious and Unconscious Processes....92
4.3.2 Language and Reasoning....97
4.3.3 Implicit and Explicit Learning....98
4.3.4 Evolutionarily Old versus Evolutionarily Recent....100
4.3.5 Modularity and Dual-Process Theories....101
4.3.6 Cognitive Correlates of Consciousness....102
Part III: Dual Cognitive Architectural Models....102
Part IV: Conclusions....104

5.Proposal for a Hybrid Cognitive Architecture....106

5.1 Introduction....106
5.2 Part I: Proposal for a Cognitive Architecture Model....107
5.2.1 Introduction....107
5.2.2 Overview....108
5.2.3 The Image Encoding-Decoding Systems....109
5.2.4 Concepts and Categorization....111
Concepts and Connectionism....112
Connectionism and Innate Concepts....115
Connectionism and Concept Individuation....116
5.3 Part II: Cognitive Attributes 118
5.3.1 Conscious and Unconscious Processes....118
5.3.2 Rational and Intuitive Reasoning....118
5.3.3 Preference and Inference....120
5.3.4 Natural Language and Reasoning....121
5.3.5 Consciousness and Volition....123
5.3.6 Consciousness and Natural Language....124
5.3.7 Implicit and Explicit Learning....128
5.3.8 Evolutionarily Old versus Evolutionarily Recent....129
5.3.9 Modularity and Dual-Process Theories....130
5.3.10 Cognitive Correlates of Consciousness....131
5.4 Part III: Possible Criticisms....131
5.4.1 Connectionism and Computation over Symbols....132
5.4.2 The Mind as Computer....133

5.4.3 The Computational Brain.....135
5.4.4 Neuroscience and Cognition.....138
5.4.5 Non-conscious and Conscious Mind Distinction.....139
5.5 Part IV: Conclusions 140

6.Bibliography.....143

List of Figures

Illustration 1: Sensory Receptors....9
Illustration 2: Müller-Lyer Illusion....10
Illustration 3: Jastrow's Duck-Rabbit....28
Illustration 4: Necker Cube....28
Illustration 5: René Magritte's "The Treachery of Images", 1928-29....29
Illustration 6: "The Cat" Figure 1 (Dennett 1990)....29
Illustration 7: Visual Recognition and Action (© McGill University)....30
Illustration 8: Example of "seeing as"....39
Illustration 9: A conceptual schematic of the active intermodal mapping hypothesis....43
Illustration 10: Example of Interpretation by the Visual System....47
Illustration 11: Rectangle Illusion....64
Illustration 12: Experiencing Conscious Free Will....95
Illustration 13: Proposed Hybrid Architecture - Simplified View....109
Illustration 14: Churchland 2007 - Fig 8-2....114
Illustration 15: "Hawk/goose dummy" (Tinbergen 1951)....116
Illustration 16: Psychological View of Memory and Cognitive Processing....124
Illustration 17: Natural Language Phonetic Input Processing and the Conscious Mind....125
Illustration 18: Learning and the Conscious Mind....128

List of Tables

Table 1: Levels of Word-Shaping by Children.....44
Table 2: Cognitive Architecture - Comparison.....81
Table 3: Dual-Process Theory – Differences between System 1 and System 2.....91
Table 4: Dual-Process Theories and Cognitive Architectures — Comparison.....103

1. INTRODUCTION

1.1 OVERVIEW

Over the latter half of the last century, developments in physics, chemistry, biology, and computer science have led to a predominantly (scientific) realist view of the world and a computational theory of mind (CTM) which has dominated particularly in philosophies of mind and language, in psychology, and in linguistics. In this dissertation, I argue for an empiricist (neo-Kantian), rather than a realist, theory of perception, for an internalist semantics, and for a model of cognitive architecture which combines a connectionist approach with a computational theory. In this model, *computation*, as defined by CTM[1], is restricted to the cognitive processes involved in encoding thoughts in a form required for input to the conscious mind and for re-use in non-computational thought processes.

1.2 BACKGROUND

Cognitive science is founded on the idea that we can learn much about the brain from the computer and a major tenet of cognitive science is that the mind is a symbolic system. Supporters of the position that we have symbol-processing minds, contend that the mind has a structure analogous to that of a digital computer. In this view, thoughts[2] are higher level brain states with a syntactic structure; thought processes are computational operations defined over such structures; and computational operations are sensitive only to the formal, syntactic properties of the structures. Not all cognitive scientists agree on the identification of brain and computer, but most cognitive scientists support some form of the Symbol-System Hypothesis (SSH) with some taking SSH literally; some regarding SSH as nothing more than a helpful metaphor or research-provoking analogy; and still others holding that the brain is *some* type of computer but not a symbol-manipulator.

In 1960, Hilary Putnam proposed the computer analogy as an answer to mind/body dualism: the body was like computer hardware (the physical computing machine); and the mind was like the computer software (the program controlling the operation of the computer). Turing Machines are simple abstract computational devices the purpose of which is to aid in investigating the extent and limitations of what can be computed. Mental states

[1] Zenon Pylyshyn (1993), a proponent of SSH, claims that: mental processing is Turing-type *computation*; knowledge is encoded by properties of the brain in the same general way that the semantic contents of the computer's representations are encoded, namely, by physically instantiated symbol structures.

[2] *Thinking* may be defined informally as: having the faculty of thought; capable of a regular train of ideas; a mental activity, not predominantly perceptual, by which one apprehends some aspect of an object or situation based on past learning and experience.

can be compared to the functional or "logical" states of a computer, but humans, unlike computers, are often inconsistent in their patterns of preference and choices and their rational preference function can change with time. In other words, humans are "Probabilistic Automata" (Putnam 1967).

A thinking computer would have to have action-directing inner-processes that match humans in adaptability (Copeland 1993). Unfortunately, designing such a computer has proved highly problematic and, further, the hypothesis that the mind is a symbol-manipulator *tout court* has problems of its own. Even though mental states can be compared to the functional or "logical" states of a computer, even Putnam observed that "[I]t is somewhat unlikely that either the mind or the brain is a Turing Machine" and it is "more likely that the interconnections among the various brain states and mental states of a human being are probabilistic rather than deterministic and that time-delays play an important role" (Putnam 1967). Jerry Fodor, one of the strongest supporters of SSH, argues that Turing's account of computation is, in at least two respects, *local* given that it does not look past the form of sentences to their meanings, and that it assumes that the role of thoughts in a mental process is determined entirely by their internal (syntactic) structure. It can be shown that there are several rational processes which are not local in either of these respects. Fodor states that, even though "Turing's theory of computation is far the best thing that we could offer":

> Wherever either semantic or global features of mental processes begin to make their presence felt, You reach the limits of what Turing's kind of computational rationality is able to explain. As things stand, what's beyond these limits is not a problem but a mystery. (Fodor 1998b)

1.3 Dissertation Thesis

I argue in this dissertation that Fodor's "mystery" is a result of insisting on the idea that the brain is, at least in most respects, some kind of Turing-style computer; that is, of insisting that computation over symbols is the best model of human cognition. I take the position that Connectionism provides the best model of cognition and that the intuition that cognitive processes constitute computation over symbols arises from the way these processes appear to the conscious mind. In the model I am presenting, the conscious mind is like a computer screen in being an "interface" to the hidden, non-conscious, cognitive processes, and whatever thoughts are encoded in a form that can appear in the conscious mind are like computer interface icons. The "icons" that appear in the conscious mind, like the icons on a computer screen, are causally-inert – they appear to consciousness only after the relevant

cognitive processes have taken place[3]. In the model I am proposing, all commonly discussed features of consciousness (such as visual awareness, self-consciousness, and qualia) are encoded as "icons" in the conscious mind through much the same mechanisms. The search for neural correlates of consciousness is an empirical matter, but one possible area of research might be as to whether the "appearance" of icons in the conscious mind is possibly the result of the activation of a form of "alarm" system which is normally the result of a physiological reaction (e.g., an increase in epinephrine) in response to stress whether environmental or psychological. What would differentiate the human conscious mind from that of other sentient beings is that some icons in the human conscious mind are linguistically-encoded.

The connectionist approach, which holds that the mind is not like a computer but is a network with multiple connections between simple active units, is, I contend a better candidate as a model of human cognition. It can accommodate local, as well as the global mental processing which which has proved so problematic for symbol-manipulation theories. One basis for the symbol-manipulation model is, I believe, a non-substantiated reverse-engineering from the surface structure of natural language. The apparently sequential nature of thought processes is deceptive and is a result of the encoding of thought processes in a form which has proved suitable for communicating with oneself ("inner voice" or "inner language") as well as with conspecifics. I reject the claim that logical reasoning relies on a language of thought that uses a syntactical structure reflected in that of natural language. The appearance that it is so reliant results, I contend, from assuming that the form of the reasoning as it appears to consciousness is actually the form of the reasoning processes themselves. This reverse-engineering is no more justified than using it to claim that the structure of the output from a digital computer is the structure of its internal processes; that the icons on a computer screen resemble the computational processing that they represent.

Thoughts that can be "broadcast" to the conscious mind are sequential and are amenable to being explained in symbol-manipulation terms, but many, perhaps most of our cognitive processes do not appear to be translatable into a symbolic or a language-like medium. Input to the cognitive processes, items in the conscious mind, and output as public communication are all sequential combinations of symbols and syntax which are very suited to being explained using the symbol-manipulation theory, but processes using the massively

[3] This analogy of the conscious mind as a computer screen, with "icons" representing but not resembling thoughts that appear to it, is owed, in part, to Donald Hoffman's proposal that we should think of "our sensory systems as constituting a species specific user interface. A user interface, like the windows interface on your laptop, is useful because it does not resemble what it represents" (Hoffman 2010).

parallel neural architecture of the brain are better explained and described by the use of a connectionist model. The proposed hybrid cognitive architecture, which takes into account recent developments in psychology, neuroscience and linguistics, is, in addition, more capable of answering the traditional questions raised by philosophers of mind and philosophers of language.

1.4 DISSERTATION OUTLINE

The structure of the arguments supporting the thesis of this dissertation is as follows:

- **Chapter 2** sets the background for supporting semantic internalism which is the position underlying the design of the cognitive architecture presented in the final chapter. Different theories of perception are examined and rejected because positions taken in the realist versus empiricist debate are relevant to which theory of semantics is supported. A realist position is reflected in the externalist semantics that has dominated philosophy of language since Putnam's influential 1975 paper "The Meaning of 'Meaning'". While it is possible to be both a supporter of semantic internalism *and* a realist, this chapter presents arguments against any of the realist positions and argues for adopting a neo-Kantian position on we perceive the world and develops some of the background assumptions used in developing the hybrid cognitive architecture proposed in the final chapter. Chapter 2 includes some detailed examination of those philosophical theories advanced since the "Linguistic Turn"[4] which have had such a strong impact on developments in cognitive science (e.g., the Computational Theory of Mind, the Language of Thought, and so on).

- **Chapter 3** discusses various important theories of cognitive architecture, including the representational/computational and connectionist models and identifies their main strengths and weaknesses. Many of the main arguments, especially from philosophy, in support of the Computational Theory of Mind and the symbol-system hypothesis are examined. Arguments are then presented in support of the connectionist model as being more consistent with psychological and neuroscientific evidence and as more biologically plausible.

- **Chapter 4** presents an analysis of several dual-process and dual-system theories of mind before offering, in Chapter 5, an alternate view which, I contend, is more consistent with current neuroscientific evidence. These dual-process and dual-system

[4] "Linguistic Turn" is the term given by Richard Rorty (1967) in reference to the focus of philosophy, and other disciplines in the humanities, on the relationship between philosophy and language.

theories are examined in detail in order to identify those attributes which are crucial for any theory of mind and, hence, for the cognitive architecture which is the subject of this dissertation.

- **Chapter 5** presents arguments for a hybrid cognitive architecture which combines a connectionist model with a computational model. Computation over symbols, in this model, is restricted to the components used for encoding and decoding items which may be "broadcast" to the conscious mind or re-used as input to other cognitive processes. Arguments are offered that the proposed cognitive architecture has the attributes which are associated with standard views of the differences between the two systems of dual-process and dual-system theories as identified in Chapter 4. Possible objections to the model I am proposing are presented and discussed. The chapter concludes with some suggestions for research projects.

---

2. An Internalist Semantics

2.1 Introduction

After reading Hilary Putnam's influential 1975 paper "the Meaning of 'Meaning'" in which he famously stated that "meaning just ain't in the head", we cannot fail to notice that he adopted certain metaphysical positions. In making his claim, he assumed that there is a mind-independent world, that that world is structured ("nature is carved at the joints"), that the way we perceive entities in the mind-independent world directly reflects that structure, that natural language can be studied as an external artifact, and that there is a non-psychological component to the meaning of linguistic terms. In this chapter, I reject his externalist semantics and the underlying assumptions.

The adoption or design of a cognitive architecture is constrained by whether the underlying theory of semantics is externalist or internalist. While it is, in principle, entirely possible to endorse both a cognitive architecture based on the symbolic computational model and an internalist semantics, most supporters of the symbolic computational model, especially its most outspoken advocates among philosophers of mind and philosophers of language (Jerry Fodor, Hilary Putnam, Zenon Pylyshyn, Steven Pinker, and many others), support some form of semantic externalism. The rival cognitive architecture, *connectionism*, is, in contrast, inconsistent with semantic externalism; its design renders it necessarily internalist. Given that I argue for a cognitive architecture that is predominantly connectionist, I also argue for an internalist semantics and, hence, need to counter arguments presented in support of externalism.

In this chapter, I examine and criticize different theories of perception and advocate an empiricist, neo-Kantian position. I argue that we can never know the world as it is, that we cannot know the objects of experience as things-in-themselves. Positions taken in the realist versus empiricist debate are also relevant to which theory of semantics is supported. A realist position is reflected in the externalist semantics that has dominated philosophy of language since Putnam's influential 1975 paper "The Meaning of 'Meaning'". The neo-Kantian position I am advocating leads to support for an internalist semantics.

Chapter 2 is divided into three parts:

- Part I consists of arguments in support of an empiricist, neo-Kantian position concerning the relationship between the mind and the external, mind-independent

world. I argue that sensory input and perceptual processing of the input cannot provide direct access to a mind-independent world and, further, that advances in physics, quantum mechanics in particular, undermine the realist's view that reality is two-way-independent of appearance.

- Part II presents arguments for an internalist semantics. I take a cognitive stance and argue that treating natural language as if it were a mind-independent object leads to puzzles and paradoxes which do not arise for internal language theorists. I argue that, contrary to Putnam's famous statement that "meaning just ain't in the head" (Putnam 1975), meaning cannot be anywhere else. I then offer an internalist perspective which owes much to the I-Language approach to linguistics (see Chomsky 1980, 1986, 2000/2005, 2006; Isac & Reiss 2008; and others) and to Jackendoff's "Unconscious Meaning Hypothesis" (Jackendoff 2012).
- Part III presents conclusions supporting an internalist approach which forms one of the assumptions used in the selection of the cognitive architecture proposed in this dissertation.

2.2 Part I – Why Empiricism

While most thinkers accept the existence of a mind-independent world, they do not all agree on how we perceive that world; they posit different theories of perception. Theories of perception include those of *common-sense* (or "naive" or "direct") *realism* for which the senses provide a direct awareness of the external world; *scientific realism* for which scientific theories provide, or may eventually provide, a true account of the world; *anti-realism* which either denies the objective reality of certain types of entities ("unobservables"), denies truth conditions to verification-transcendent statements about these entities, or, in its *constructive empiricist* version, holds that all that is required of scientific theories is that they be *empirically adequate*; and *empiricism* for which claims about the world rest solely on experience and for which knowledge about the external world is always subject to revision being no more than probabilistic and tentative.

2.2.1 Realism and Empiricism

For scientific realists, progress in science leads to theories that are closer to the truth about objective reality, and the belief that it is possible to transcend experience, thereby to discover the causes of events even in the absence of direct observation. The empiricist

position, on the other hand, rests on the assumption that, as Immanuel Kant observed (Kant 1963:B xix), we can never transcend the limits of possible experience. For empiricists, things-in-themselves belong to absolute reality and are unknowable as such since we are unable to take a "God's eye view" which would enable us to compare mind-independent entities with our experience of them; we are denied a point of view that would enable us to see the world unfiltered. We are, nonetheless, able to think of objects of experience as things in themselves "otherwise we should be landed in the absurd conclusion that there can be appearance without anything that appears" (Kant 1963:B xxvi). Empiricists can thus accept that there are mind-independent entities that act as the causes of our experiences even though these entities are unknowable.

Newtonian physics and Euclidean geometry have been extremely successful in helping us understand and describe empirical reality (the reality of our experiences). When constructing buildings or bridges, when sending a man to the moon, we continue to use Newtonian mechanics because these activities are restricted to the medium-sized objects, distances, and velocities of our empirical reality. We may attempt to transcend the limits of experience, but continuing scientific advances, particularly those of quantum physics, separate our empirical world of experience further and further from the world as it may be in itself. Newtonian physics is not only a useful way of describing and coping with empirical reality, but is a reflection of the way the human cognitive capacity synthesizes the manifold of representations. Our experiential frame of reference conforms to Newtonian principles because, as Kant claimed, these "concepts of reason are not derived from nature; on the contrary, we interrogate nature in accordance with these ideas, and consider our knowledge as defective so long as it is not adequate to them" (Kant 1963:A645/B673). Donald Hoffman has recently presented arguments from natural selection for a similar position to that of Kant's:

> Nowhere in evolution, even among the most complex of organisms, will you find that natural selection drives truth to fixation, i.e., so that the predicates of perception (e.g., space, time, shape and color) approximate the predicates of the objective world (whatever they might be). (Hoffman 2009)

Unlike realist theories, neo-Kantian empiricism about empirical reality and things-as-they-are-in-themselves does not involve metaphysical claims, but only the epistemological claim that we have can have no knowledge of such mind-independent entities: we may accept that there are mind-independent entities which act as the causes of our experiences, but these entities are unknowable.

2.2.2 **Perception and the Senses**

Whether thinkers believe in the existence of an independent external world or not, they can agree that a person must be constituted in a way that makes experience possible, and they can also agree that having such a constitution is not *sufficient* for experience. While we can agree that an entity we experience as inanimate, a rock for instance, cannot itself have experience, we can also agree that, even though having the right constitution would provide the potential for experience, something more is required for *having* experience. Stimuli as well as sensitivity to stimuli are both required. Put simply, sensory information is also *necessary* for experience, whether such information is received through the senses of sight, hearing, touch, smell, and taste, or through bodily sensations like pain. There are both subjective and objective conditions for having experience of an "external" world.

Our interface with the world is provided by many different types of sensory receptors which are encoded by sensory transducers the output of which are used by the brain to build a model of the world. Sensory transducers convert input stimuli to electrochemical signals. Many different types of sensory transducers have evolved to respond selectively to different types of input (light waves, sound waves, chemical and mechanical stimuli, electric and magnetic fields, and so on.) Each sensory receptor transmits a discrete signal which is part of an input vector of signals constituting one sensation event. Different modalities can interfere with each other as demonstrated by, for instance, the McGurk Effect which shows very neatly how what we see affects what we report hearing. In addition, the brain is plastic and malleable: sensory input can physically affect brain organization, an example of which is the enlargement of the hippocampus of London taxi drivers after training involving extensive use of spatial memory (Maguire et al., 2000).

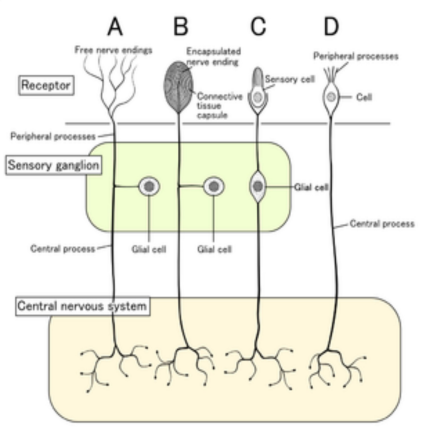

Illustration 1: Sensory Receptors (© wikipedia)

Visual modality is usually privileged in most discussions of sensation and perception which is not surprising given that sight is the pre-eminent sense required for human survival. However, even a cursory examination of the physical structure of the visual system will show that it is insufficient for an explanation of visual perceptions. What we "receive" in the brain are the results of large numbers of discrete signals resulting from the stimulation of many individual rods and cones in the retina of the eye. The brain then has to assemble

these electro-chemical signals into a meaningful image. Stimulation by individual photons is the only means of communication between the objects around us and our visual system[5]; yet, despite the lack of organization in the retinal stimulation, perceptions *are* organized. Furthermore, humans are able to adapt to visual distortions, thus implying that additional conditions are required for a coherent and complete account of experience.

Psychologists tend to favour either *bottom-up* (*data-driven*) or *top-down* perceptual processing. Bottom-up perceptual processing is unidirectional, starting at the retina and proceeding to the visual cortex with analysis of the input being increasingly complex at each stage in the visual pathway (see Gibson 1966). In top-down processing, contextual information is used in pattern recognition and, thus for this theory, perception is a constructive process (see Gregory 1970). Many philosophers and psychologists (e.g., Jerry Fodor and Stephen Pinker), as well as evolutionary psychologists (e.g., Leda Cosmides and John Tooby) reject the top-down approach by arguing that perception is impenetrable to a subject's background knowledge. Fodor argued that sensory transducers convert input stimuli to electro-chemical signals and produce output which is lawfully dependent on their input (Fodor 1983:40). The evidence he uses to demonstrate his modular view of perception is the Müller-Lyer illusion as shown in Illustration 2. Even if we know that the two horizontal lines in the illusion are equal, we, nevertheless, continue to see one line (the bottom line in this case) as longer than the other. This *informational encapsulation* of the perceptual modules limits the extent by which perception can be a constructive process (Fodor 1983) at least in the early processing stages. This position would, at first glance, appear to be contrary to Hoffman's theory of *visual intelligence* (Hoffman:1998) which he describes as a constructive process. Hoffman's v*isual intelligence* is, however, a rule-based process in which (as argued by supporters of modular theories such as Fodor's) a subject's theories, beliefs, and background knowledge play no part. For Hoffman, "visual worlds" are constructed from ambiguous figures such as those in and below, in conformance to visual rules not background knowledge (Hoffman 1998:24). Hoffman

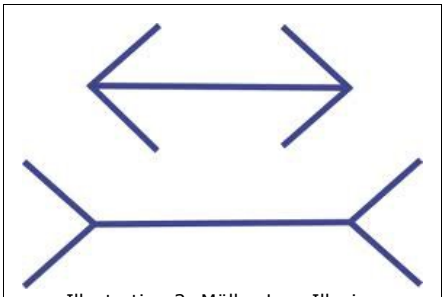

Illustration 2: Müller-Lyer Illusion

[5] This is not completely accurate because the visual system can be activated by converting signals of other modalities into a form similar to normal visual input: that is, the visual cortex may be recruited for identifying sound-encoded objects. For instance, Dr Peter Meijer (senior scientist at Philips Research Laboratories in the Netherlands) developed The vOICe system (the three middle letters standing for "Oh I See") which works by translating images from a camera on-the-fly into highly complex soundscapes, which are then transmitted to the user over headphones. After considerable practise, it is possible for some subjects to *see* complete images with depth and texture.

notes that there is a fundamental problem of vision namely that the "image at the eye [the image cast on the retina] has countless possible interpretations" (Hoffman 1998:13) but the Muller-Lyer illusion indicates that the brain rapidly and unconsciously processes information about length and size before any other stages in the perceptual process. This is not to say that all stages/modules in the perceptual process are informationally-encapsulated but only those (early) stages which construct images in conformity to visual rules and not to the subject's background knowledge.

In the preceding discussion, it must be emphasized that there is a difference between perception of an object and perception of a fact. Fact perception involves knowledge and beliefs, and therefore involves the whole cognitive apparatus, whereas object perception does not – knowledge, beliefs, and theories are, however, required for rending the percept "meaningful". Philosophical theories vary on whether we apprehend, in object perception, external objects or internal objects (representations of the objects or ideas in the mind), but, whether or not the senses are input systems or input modules, it is doubtful that any theory of perception proposes that inputs (such as photons) are directly sent to the mind without some form of preprocessing.

2.2.3 **Scientific Theories and Empiricism**

Advances in physics and technology have vastly expanded our understanding of reality (or, rather, what reality might be like to account for our experiences of it) but this does not go hand-in-hand with a more direct access to the entities which science postulates as real. In fact quantum mechanics, as it is currently formulated, supports the view that we can never form a true picture of the world-as-it-is-in-itself because of the actual role that measurement and observation play in the theory's description of fundamental reality — because of the role the observer plays, quantum theory cannot provide a picture of reality that is absolutely independent of the observer.

Further, we experience acceleration as a detectable change in velocity in a three-dimensional space, not as a curved line in spacetime as defined in Einstein's Special Theory of Relativity and Minkowski's Spacetime Theory. It is likely that we evolved three-dimensional space detection modules because the greatest acceleration that we could experience in the environment of evolutionary adaptation, consistently enough to affect evolution, was acceleration due to gravity which produces a radius of curvature in spacetime so large relative to human dimensions as to be undetectable without the use of advanced technology. There is,

apparently, no evolutionary advantage for humans to be able to experience curves in space-time, any more than there was an evolutionary advantage for humans to be able to hear infra-sound or see ultraviolet waves.

We experience time as absolute and uni-directional — we evolved in such a way as to experience proper time, or time in an inertial frame. Our cognitive functions have evolved such that we experience only the three-dimensional cross-section of spacetime. No matter how our technology advances, our realm of possible experience is restricted to empirical reality; technology cannot provide us with unmediated experience of the world as it is in itself. The external world of possible direct human experience has four dimensions (three spatial and one temporal) and contains medium sized objects, medium length distances, and medium velocities — in other words, the human cognitive faculties have evolved in such a way as to render the infinitesimally small objects and distances postulated by Quantum Mechanics and the extremely large objects, distances, and velocities postulated by General Relativity outside their frame of reference.

Perception can be defined as the active process of selecting, organizing, and interpreting sensory information; it is the method by which sensations (sensory information, sense-data) are interpreted. A physical account would be that sense organs absorb energy from stimuli in the environment and sensory receptors convert this energy into neural impulses which are then transmitted to the brain. Perception is the result of the brain's organizing the sensory information and translating it into something meaningful, however "meaningful" may be defined.

2.3 Part II: Why an Internalist Semantics

The traditional conception of meaning assumes that to know the meaning of a lexical item, sentence, predicate, or such like, is to be in a particular psychological state (e.g., perception, belief, memory) which represents entities in the world and properties of those entities. There is a relationship between this conception of meaning and that of certain philosophers of science who, in discussions of the observational-theoretical division, refer to the the truth or falsity of observation sentences and the role that sense contents play in their tokening (see Maxwell 1962:3-15). These philosophers of science thereby hold that sense contents are conceptual in nature, that is, that they have a representational content which could act as a premise or conclusion in an inference. "Observation statements" are public entities by definition and presuppose (and are expressed in the language of) some

possible fallible theory or theories. Perceptions, on the other hand, are not public but are directly accessible to the observer – access to percepts is *privileged*. The scientific observation statement "the Earth goes around the Sun" requires a theory which contradicts common-sense perceptions – we still talk of the Sun setting or rising even if we accept the solar-centric theory. Based solely on unaided sensory input, our *mental representation* of the Sun is of something that moves, but a scientific theory can modify interpretations of the input.

A representational mental state is "about" something – it has *intentional* content. In answering the question of how intentional states (specifically *de dicto* intentional mental states[6]) come to have the particular content that they do, contemporary philosophers tend to favour one of two main positions: *internalism*, whereby mental contents are sufficient to fix intentional content; and *externalism*, whereby mental contents are not sufficient and that "meaning ain't in the head" or, at least, is not completely "in the head" (Putnam 1975). Externalism is a thesis for which the property of a particular intentional mental state (e.g., a belief, a desire, an intention) is individuated by how the person having that property is related to the external world. In other words, the content of a person's thoughts and the meanings of the terms he uses to express them are dependent on his external environment. Internalism, in contrast, is a thesis for which a person's having the property does not depend on any relation to the external world. It should be emphasized that internalism does not deny that factors in the external world may causally influence whether or not the person has mental states with the property, but only holds that properties of intentional states are not *individuated* by relationships between an person and the person's external environment – identical factors in the external world could causally influence two different individuals to token intentional states with different properties depending on their individual intrinsic properties – both may experience a flash of black and yellow, for instance, which causes one to token an intentional state with the content 'tiger" but in the other an intentional state with the content 'sunflower'. Putnam's 1975 paper gave rise to an externalist semantics which is widely accepted in the cognitive science community, especially by many philosophers of mind and psychologists. Putnam arrived at this position by, in part, extending Saul Kripke's causal theory of reference (see Kripke 1971, 1972/80).

[6]Externalism and internalism disagree on *de dicto* intentional mental states which attribute propositional attitudes (mental states, like *belief,* held by the individual toward a proposition) in a manner that produces referentially opaque contexts (that is to say, substituting co-referring terms within their scope may change the truth-value of the sentence.) In English, the propositions are introduced with 'that-clauses'. 'John believes that George Sand was a Frenchman' is syntactically *de dicto:* 'John believes that Amandine-Aurore-Lucille Dudevant was a Frenchman' is not true despite the fact that 'George Sand' and 'Amandine-Aurore-Lucille Dudevant' are co-referential.

2.3.1 **The Causal Theory of Reference**

The causal theory of reference (hereafter the "K-P theory") is the theory that names refer to their referents by being causally connected to them in some appropriate fashion. A particular token of a name will belong to a causal chain of such tokenings which originated in an "initial baptismal" event (frequently of "ostension") when the referent was present. Putnam does, however, extend this to the possibility of the "baptismal event" taking place in the absence of the referent. In either case, names refer "rigidly" to their referents in that they refer to this referent in every possible world in which that referent exists.

K-P theory in itself only becomes tendentious when extended beyond a theory about the referent of a proper name to a claim that the *meaning* of a lexical item is its reference. The causal chain linking the proper names and their referents is mediated by the linguistic community but there is, in fact, rarely a single, homogeneous community to ensure that there is one and only one causal chain linking a specific name with a specific referent. There is no ultimate linguistic authority. A simple example is that of 'billion' which has different referents in UK and American English (a *million million* and a *thousand million* respectively). We can, of course, just rely on calling the UK and American 'billion' terms different, homophonic terms but, by the same token, we can just call 'Londres' of Kripke's Pierre[7] a homophone of 'Londres' in the linguistic community of Kripke's translator because 'Londres' and 'London' are not co-referring in Pierre's idiolect. Referentialist theories have a tendency to require proliferation of homophones to account for the variations in reference for a term. This is discussed further below in regards to, for instance, natural-kind terms and natural-kind terms *qua* natural kind terms (e.g., 'water' referring to water which happens to be a natural kind and 'water' as a natural-kind term); and to terms which have changed meaning/reference over time (e.g., 'meat' which originally referred to food generally but now refers to edible animal flesh, and 'fish' which was originally used to refer to all "creatures that live in the sea" but was co-opted by biologists to refer to a subset of sea creatures.)

[7]Kripke (1979) proposed a thought experiment in which a mono-lingual speaker of French, Pierre, believes to be pretty a certain city which is called 'London' by English-speakers and which he calls 'Londres'. After seeing many pictures of the city, he sincerely assents to the sentence "Londres est jolie". Pierre moves to London and learns English by the direct method of language acquisition. He becomes a normal speaker of English and one of the things he learns is that the city in which he is now living is called 'London'. Unfortunately, he moved to an unattractive part of the city and comes to assent to the sentence "London is not pretty". That is to say, he does not, upon reflection, sincerely assent to 'London is pretty'; he, in fact, sincerely and reflectively assents to 'London is not pretty'. Pierre's holds seemingly contradictory beliefs ("Londres est jolie" and "London is not pretty") without any logical deficiency on his part. Kripke acknowledges that, while it is possible to describe Pierre's situation coherently, it is not possible to answer satisfactorily the question: "Does Pierre believe that London is pretty?"

2.3.2 **Internalist and Externalist Semantics**

In their analyses of language, most analytic philosophers view natural language as an object to be investigated without reference to the mental processes of language users. Just as Chomsky (1986) differentiated between *E-Language* (natural language viewed as an external artifact with properties that are part of the external world) and *I-Language* (language viewed as a body of knowledge within the minds or brains of speakers), theories of meaning can be divided into externalist and internalist theories. Externalist and internalist theories can both be further divided into those concerned with the meaning of natural language (NL) lexical items and those concerned with the meaning of concepts which may be expressible by NL lexical items. Treating meaning as a relation between lexical items (or lexical concepts) and "mind-independent" entities in the world is, I contend, mistaken.

2.3.3 **Externalist Theories of Meaning**

Using, in part, the theory of rigid designation advanced by Saul Kripke (1972/1980), Putnam claimed that the *intension* of a linguistic expression is determined, mostly, by external factors. Both his and Kripke's externalist position on how the meaning of terms is fixed had considerable influence on the philosophy of language and on other areas of cognitive science. In this section I analyze, compare, and criticize two examples of externalist theories of meaning, namely Putnam's linguistic theory, concerned with the meaning of lexical items, and Fodor's psychological theory, concerned with the meaning of lexical concepts. These two theories both accept a form of the Kripkean causal theory of reference. They hence both propose an externalist semantics.

2.3.4 **A linguistic theory of meaning**

Putnam rejected the traditional view that the meaning of terms with different extensions requires different narrow psychological states. That is to say, he rejected the view that intension determines extension. The conclusion of his famous "Twin Earth" thought experiment is that "[E]xtension is not determined by psychological state" (Putnam, 1975:222). His thought experiment is intended to show that it is possible for two "English" speakers ($Oscar_1$ on Earth and $Oscar_2$ on "Twin" Earth) to be in the same psychological state with respect to a term 'T' even though the term refers to distinct extensions in each of the

two Earths: in 1750, prior to the development of modern chemistry in either of the two Earths, $Oscar_1$ and his "Twin" Earth counterpart $Oscar_2$ have identical experiences of and beliefs about water – they are in the same narrow psychological state – even though the extension of 'water' in each of the Earths is different. Hence, "the extension of the term 'water' . . . is not a function of the psychological state of the speaker by itself." Putnam's point is that it is possible to know the meaning of a term without being able to fix the extension of the term. Determining the extension of many terms must, Putnam proposes, be left up to experts; for such terms, there is a linguistic "division of labor" whereby their meaning is a function of a social interaction between average language users and experts (such as physicists and chemists).

The K-P theory is based on an extension of the theory of proper names but there are many cases when the original "baptismal" event occurred in the absence of the referent or even when the referent did not exist. An example of the latter is 'Atlantis'. This is not a "meaningless" term and is used as a proper name, but there is no referent; hence its meaning cannot be the referent. An example of "premature" baptism is that of "the North-West Passage": explorers searched for "the North-West Passage" before they knew if there was just one or if there were many. They were not searching for *a* north-west passage but for *the* North-West Passage. Scientists often postulate the existence of entities before being able to prove if these entities actually exist. Yet their use of the names of these entities are still meaningful whether or not the referents actually exist: the term "phlogiston" did not cease to be meaningful as soon as it was proved to have no referent. If the meaning *is* the referent, then it is questionable as to where the meaning lies for these or any terms referring to mythical or fictional entities.

For Putnam and Kripke, meaning is construed linguistically. Concerning individuation of the same term with different senses (e.g, 'bank'), Putnam refers to the "standard" view that a term is actually an ordered pair consisting of the extension and the sense (1975:216). While Putnam accepted that extension is a principal component of the meaning of natural-kind terms, he rejected that sense is a part of meaning, and proposed that the linguistic meaning of *every* term in the language could be specified by a finite sequence of elements (a *vector*), where, for natural kind terms, at least, the extension is a component of its *meaning vector* (Putnam 1975:246). This meaning vector has several other components including: semantic indicators (that place the referent in particular general categories),

syntactic indicators (that govern the term's formal relationships in phrases or sentences), and a stereotype (a set of typical descriptions of the term).

A change in the reference of the term but not in its stereotype is, Putnam claims, the only legitimate way that the meaning of an expression can change, but this claim may be hard to substantiate in the case of many artifacts: for instance, the stereotype of a quill feather changes significantly when it is used as a pen – the meaning has changed but not the reference. When someone utters the statement "Bush is no Einstein", where the context tells us that 'Bush' refers to George W. Bush who served as the 43rd President of the U.S.A., what are the meanings of the terms 'Bush' and 'Einstein'? If we accept the K-P theory of proper names, then the meaning of 'Bush' just is the referent, the person also known as George W. Bush, but, in this example, the term-world relation does not work for the term 'Einstein'. The speaker has in mind something other than the human DNA, with a specific nitrogen base arrangement, tagged 'Einstein'; the speaker-meaning just is the stereotype. Nevertheless, in a specific case there may be no way of definitively determining whether it is its stereotype or its reference that has changed and hence the usage of other expressions of the language have also to be considered (Putnam 1975:269-270). Putnam's theory of meaning is, thus, a form of semantic holism.

Underlying Putnam's analysis are two assumptions: 1) that the meaning of a term is its reference (extension); and, more significantly, 2) that terms themselves have meaning. The latter assumption further assumes an externalist perspective that (natural) language is an artifact (what Chomsky calls 'E-Language') that can be analyzed independently of language users in much the same way that we analyze the mathematical notation of a theorem without regard to the psychological states of the mathematicians who developed or use it – this is in contrast to the way we treat musical scores for which the composer's intention (the composer-meaning rather than the notation-meaning) is paramount. I contend that treating (natural) language as an external artifact gives rise to linguistic paradoxes such as Frege's Puzzle, Mates' Cases, and even Putnam's Twin case[8].

2.3.5 **A psychological theory of meaning**

Like Putnam, Fodor supports the view that meaning is denotation: "there is nothing to the meaning of a name except its bearer and nothing to the meaning of a predicate except the

[8] Linguistic puzzles such as these arise from making two assumptions: 1) The meaning of a term is given by its reference; and 2) The meaning of a sentence is given by its parts. The puzzles are discussed further in the section 2.3.9 "Externalist Semantics and Linguistic Puzzles".

property that it expresses" (Fodor 1990:161). Also like Putnam, Fodor proposes a causal theory of meaning but, rather than a theory of linguistic meaning, he supports a causal theory of *representation*, an informational theory of content using the notion of information to explain *naturalistically* how intentional states come to have content. On this account, the contents of mental representations are determined by their causes: for example, the mental representation (MR) CAT is about cats (has 'cat' as its content) because cats cause it. That is, an MR receives its content only through the mind's causal connections with the environment. Tokenings of propositional attitudes (such as *believing, fearing, hoping*) involves tokening an MR which, for Fodor, is a symbol in the language of thought (LOT). The syntax of a symbol is generally a determiner of its causal role: symbol tokens interact causally in virtue of their syntactic structures. Two symbols can be transformed one to the other if and only if they stand in semantic relations to each other (the semantic relations that, for instance, premises have to their conclusions in a valid argument). A symbol's truth condition is the state of affairs in the world which, it it obtains, would make the symbol true.

Unlike those who support a linguistic theory of meaning, supporters of theories of meaning with psychological components are likely to consider artifacts to be mind-dependent:

> What with one thing and another, I've been pushing pretty hard the notion that properties like being a doorknob are mind-dependent. ... the Standard Argument shows that only non-primitive concepts can be learned inductively. And it's been the main burden of this whole book that all the evidence—philosophical, psychological, and linguistic—suggests that DOORKNOB is primitive (unstructured); and, for that matter, that so too is practically everything else. (Fodor 1998a:147)

Fodor's position needs, however, some way of distinguishing between a *mind-independent* property of *being a P* and a *mind-dependent* property of *being a P.* "Suppose, in particular, that being a doorknob is just having the property that minds like ours reliably lock to in consequence of experience with typical doorknobs" (Fodor 1998a:148). But, if *being a dog*, for instance, is a mind-independent property and *being a doorknob* is not, what is the difference in the two locking mechanisms? Fodor (1998a:88n) states that he uses both 'stereotype' and 'prototype' to refer to mental representations of certain kinds of properties. On this view, 'the dog stereotype' and 'the dog prototype' both designate a concept (e.g., "BEING A DOMESTIC ANIMAL WHICH BARKS, HAS A TAIL WHICH IT WAGS WHEN IT IS PLEASED,... etc.") A stereotype for Fodor is a type of referent and can be either a particular entity (often termed an *exemplar*) or a set of properties (often termed a *prototype*). Studies

indicate that pre-verbal children can distinguish the use of the definite (the particular) from the indefinite (one-of-a-set article) with word shaping occurring almost always at the "Base Level" (Markman 1989). This would indicate that pre-verbal children can lock to the mind-independent property of *being a dog* without any scientific-cum-metaphysical theory but rather that being a dog "is just having the property that minds like ours reliably lock to in consequence of experience with typical" dogs.

The K-P theory provides no discussion of how we understand word meanings other than a brief mention that when it comes to individual competence in making the term/extension link, "concepts have a lot to do with meaning" (Putnam 1975: 246-7). While Fodor supports a form of semantic externalism, for him meaning is construed *psychologically;* he is concerned with psychological explanations rather than a theory of meaning for natural language. Linguistic expressions of a natural language are conventionally used to express intentional mental states and many philosophers (Fodor 1978, Searle 1983, and others), in contrast to Putnam for instance, claim that their semantic properties are inherited from those mental states. Fodor's theory of meaning is concerned with *underived* meaning, and is a theory in which the meanings of thoughts are *prior* to the meanings of any symbols used for communication or for expressing thoughts – the meaning of symbols is derived from the meaning of the thoughts being expressed (Fodor, 1987).

2.3.6 **Determination of Extension**

One of the first problems of meaning is that of how extension is determined (Putnam 1975:246). Putnam defines extension as the set of all things that the term is true of: for instance "'rabbit' ... is true of all and only rabbits". What the mechanism for determining an extension is depends, in part, on whether the extension is mind-independent or mind-dependent. The Kripke-Putnam account is concerned with mind-independent extensions, with those extensions which are determined by the world. On this view, "[T]he world consists of some fixed totality of mind-independent objects" (Putnam 1981:49) and, to paraphrase Plato (*Phaedrus* 265d-266a), the world is "carved by nature at its joints". There are clearly some metaphysical and epistemological assumptions underlying this view.

Underlying the K-P theory of reference is the idea of a world consisting of a fixed totality of mind-independent objects and properties; of a world that is, hence, independent of any discourse about it. Along with this assumption of metaphysical realism are also assump-

tions of a correspondence theory of truth and that there can be one true, complete description of reality:

> There is exactly one true and complete description of 'the way the world is.' Truth involves some sort of correspondence relation between words or thought signs and external things and sets of things (Putnam 1981:49)

A further assumption is that science will be able to resolve, at least in principle, all philosophical problems. It should be noted, however, that this does not reflect Putnam's later position in which he abandoned metaphysical realism along with its commitment to the existence of mind-independent entities. In his later writings, he advocates a type of pragmatic realism ("internal realism") motivated in part by the failure of scientific realism to provide an account of intentionality. He supports an ontology relativized to conceptual schemes:

> 'Objects' do not exist independently of conceptual schemes. We cut up the world into objects when we introduce one or another scheme of description (1981: 52). If, as I maintain, 'objects' themselves are *as much made as discovered*, as much products of our conceptual *invention* as of the 'objective' factor in experience, the factor independent of our will, then of course objects intrinsically belong under certain labels because those labels are just the *tools we use to construct a version* of the world with such objects in the first place. (Putnam 1981:54 italics in original)

Fodor too assumes metaphysical realism: "the centrality of perceptual mechanisms in mediating the meaning-making laws is also just a fact about the world, and not a fact about the metaphysics of content" (Fodor 1998a:79). On Fodor's view, having a natural-kind concept requires locking to the natural kind and having a natural-kind *qua* natural-kind concept[9] requires locking to the natural kind via a scientific-cum-metaphysical theory (Fodor 1998a:157). Fodor argues for a folk psychology within a physicalist framework and for an account of concepts which is solely an exercise in metaphysics – epistemology is not relevant (Fodor 1998a:5ff).

2.3.7 **Determination of Reference**

There is a distinction between how the meaning of a word is determined and how the reference of the word is determined. The Kripke-Putnam account of reference has been criticized recently for giving rise to what has been called the "*qua* problem" (Devitt & Sterelny, 1978). According to the Kripke-Putnam view, 'gold' involves the act of ostension,

[9] It is not clear how Putnam would distinguish between a natural-kind term and a natural-kind *qua* natural-kind term; between, that is, the theoretic term and the non-theoretic (non-theory-mediated) natural-kind *simpliciter*.

of pointing to samples of the substance gold which is the element with the atomic number 79 (Kripke 1972/1980:116). As Locke noted, however, natural kind terms can group objects in different categories depending on the properties under consideration:

> The same Convenience that made Men express several parcels of yellow Matter coming from Guiny and Peru, under one name, sets them also upon making of one name, that may comprehend both Gold, and Silver . . . under the name Metal (Locke, 1690/1959, III vi, 32).

Thus, when a sample of gold is pointed to, it is not clear which natural kind the sample instantiates. There is still the problem of fixing the reference of 'gold' to gold rather than to metals or elements; of gold *qua* gold, or *qua* metal, or *qua* element (Stanford & Kitcher 2000). To resolve this problem, Dewitt and Sterelny add a descriptive component to the causal theory of reference:

> Something about the mental state of the grounder must determine which putative nature of the sample is the one relevant to the grounding, and should it have no such nature the grounding will fail. It is very difficult to say exactly what determines the relevant nature.
>
> People group samples together into natural kinds on the basis of the samples' observed characteristics. They observe what the samples look like, feel like, and so on. They observe how they behave and infer that they have certain causal powers. At some level, then, people "think of" the samples under certain descriptions and as a result apply the natural kind term to them. It is this mental activity that determines which underlying nature of the samples is the relevant one to a grounding. The relevant nature is the one that is, as a matter of fact, responsible for the properties picked out by the descriptions associated with the term in the grounding. If the sample does not have these properties – if, for example, the alleged witch does not have the power to cast spells – then there will be no relevant nature and the groundings will fail (Dewitt & Sterelny 1987:73–74)

As Stanford and Kitcher emphasize, the descriptive component "must include all and only those features that are causally relevant to the production of the observable qualities in question."

One problem is to identify which scientific theory is of consideration when determining reference. A physicist and a chemist may have different interests in, and are thus concerned with, different underlying properties of the same substance. As stated previously, a physicist may be concerned with electron configuration and the chemist with chemical composition. Kripke claims that water is necessarily H_2O but if what is important about water is not that its molecules are composed of hydrogen and oxygen, but rather that its molecules

have a particular electron configuration, then it would be possible that Putnam's Twin Earth XYZ water could be theoretically identical to the water on Earth as long as it has the same electron configuration.

While determination of reference for (natural-kind) terms may be a matter of the knowledge of the linguistic community as a whole, as Kripke and Putnam maintain, how the term is used may differ from speaker to speaker. There is a distinction between semantic reference and speaker reference:

> In a given idiolect, the semantic referent of a designator (without indexicals) is given by a *general* intention of the speaker to refer to a certain object whenever the designator is used. The speaker's referent is given by a *specific* intention, on a given occasion, to refer to a certain object. (Kripke 1977)

Kripke (1977) differentiates between the "simple" case in which a speaker has the *specific* intention of referring to the semantic referent (that is to say, his specific intention is simply his *general* semantic intention) and the "complex" case in which a speaker has a *specific* intention which he believes, as a matter of fact, determines the same object as that determined by his *general* intention. While, in the "simple" case, the speaker's referent and the semantic referent coincide, they only do so in the "complex" case when the speaker's belief is correct.

2.3.8 **Meaning of Natural-kind terms**

For Putnam and Kripke, 'meaning'[10] is to be defined as the reference of an expression (its denotation[11]) as opposed to something that one wishes to convey, especially by language, or the sense of an expression (its connotation). Both Putnam and Kripke based many of their arguments on the idea that that proper names are *rigid designators* and extended the theory to apply to natural kind terms: "terms for natural kinds are much closer to proper names than is ordinarily supposed" (Kripke 1972/1980:127). Putnam states that, in order for most people to know the *meaning* of a natural-kind term like 'gold' or 'tiger', they have to defer to experts who are able to determine the micro-structure of the natural kind to which the term refers. In other words, without deference to experts, we would not know the *real* meaning of such terms. This, I find highly questionable. There is a difference between

[10] Putnam (1975:223-4) states that for his theory of meaning: the verb 'means' sometimes means 'has an extension' but "the nominalization 'meaning' *never* means 'extension'".
[11] Kripke (1972/1980:25n3) "Perhaps it would have been less misleading to use a technical term, such as 'denote' rather than 'refer'. My use of 'refer' is such as to satisfy the schema, 'The referent of "X" is X', where 'X' is replaceable by any name or description."

applying a natural-kind term incorrectly and not knowing the *meaning* of a term. The term 'gold' was used long before anyone knew the atomic composition of gold. It seems strange to claim, as Putnam does, that the meaning of a natural-kind term does not change but that users, prior to the discovery of the chemical composition of the term's referent, users of the term were unaware of its meaning. They were obviously not aware of the modern scientific usage of the term but were, nevertheless, able to use the term to express their thoughts. How the term is used depends on the interests of the user: a physicist may use 'gold' as another name for Au or for atomic number 79, while a lay person is more likely to use 'gold' to refer to a valuable, yellow[12] metal that will never rust or tarnish and that is often used in ornamentation or as a currency or standard for global currencies. In other words, the usage of the term is interest-relative. Even for the scientist using a natural-kind term, usage is determined by the scientific theories "in his head". When we defer to an expert, we are not usually in the position of the expert's giving us an ostensive definition – pointing to an electron, or a quark is likely to be problematic. Instead, the expert usually has to try to convey the relevant and current *theory*.

Even if we accept that a relevant scientific community maintains the *meaning* of a scientific term, the *reference* of the term is likely to change over time. According to Kripke, when a biologist discovered that 'whales are mammals, not fish', his denial that whales are fish does not show that his 'concept of fishhood' is different from that of the layman; "he simply corrects the layman" (Kripke 1972/1980:138). Nevertheless, in the Bible (Ezekiel 38:20), reference is made to "the fishes of the sea". The reference for the initial 'baptism', the initial tagging of 'the fishes of the sea', is not what we now-a-days would call 'fish' since the original extension of 'the fishes of the sea' would have included mammals such as whales and dolphins. Revisions in the reference of 'fish' were made for scientific reasons but not because of any empirical discovery that some creatures originally classified as 'fish' were mammals and that the writers of the Bible were in need of correction. By 'fishes', the writers of the Bible (and laymen at the time) meant *all* creatures that live in the sea – for them, the creatures of the world were delineated into "the fishes of the sea", "the fowls of the heaven", "the beasts of the field", and "all creeping things that creep upon the earth". The relation between 'fish' and its referents has undergone, and continues to undergo, revi-sion – in this case, the reference of 'fish' is now fixed by what is most useful to the biolo-

[12]The English term 'gold' comes from the Old English word for yellow, 'geolu'.

gists and, since 'fish' currently refers to a paraphyletic group[13], the reference is likely to undergo many more revisions.

A pertinent example of a change of reference is that of 'water': when Lavoisier discovered that the chemical composition of water is two-parts hydrogen and one-part oxygen, the term 'water' was co-opted by chemists as a name for the category of all substances composed of H_2O molecules (they could also have used 'ice' or 'steam'); the discovery of deuterium oxide ($H_2{}^{18}O$), however, led to the creation of a new term, namely 'heavy water' which, despite having the same chemical composition as water, has very different physical properties from those of ordinary water[14]. It should be noted that even "chemically pure" water is a mixture of isotopic species. In fact, many discussions of water and H_2O fail to distinguish between the two: both water and heavy water have the same chemical composition (two-parts hydrogen and one-part oxygen) but have different micro-structures (H_2O and $H_2{}^{18}O$ respectively).

2.3.9 **Externalist Semantics and Linguistic Puzzles**

Kripke claims that notions concerning the belief and content of a proposition are insufficiently clear "to draw any conclusion, positive or negative, about substitutivity" (Kripke 1979:269). This is a rather unsatisfactory conclusion. The problem that Kripke has in "drawing any conclusion" about why this should be the case arises, I contend, because his theory is non-cognitive; his theory relates solely to the "meaning" of names where "meaning" just is reference and hence names have no semantic properties. If, as 'conventional judgement' holds, 'Londres' and 'London' are referentially *opaque* **for** Pierre[15], then Pierre must hold contradictory beliefs, but if they are referentially *transparent*, then they identify the same referent for Kripke (and his readers) but *not* necessarily for Pierre. The puzzle arises only if we choose an opaque reading. Even though there is nothing in the expressions themselves that indicate which reading is correct in such situations, there is no puzzle in such cases if we reject the 'conventional judgement' that belief contexts are 'referentially opaque'.

A theory of concepts requires an account of a mechanism to explain the correlation of external entities and internal (mental) physical states so that mental states m*ean* (are

[13] Paraphyletic groups, unlike monophyletic groups, do not include **all** descendants of a single common ancestor – certain subsets of descendants are artificially ignored for practical reasons.
[14] For instance, heavy water melts at nearly 4°C higher than does ordinary water at ordinary atmospheric pressure.
[15] See footnote 7 on page 14 for a description of Kripke's "Pierre" thought experiment.

about) the external entities. If we accept CTM and hence that psychological processes are computational, then we have to be able to reconcile broad intentional content with "mechanisms that (contingently) insure that linguistic paradoxes Twin cases and Frege cases don't occur (very often)" (Fodor 1994:27). That is, we have to be able to explain how psychological laws can be intentional. Fodor (1994:57) presents the following constraint on concept identity:

C: Concepts that carry the same information are always coextensive

Where, by 'information', he means *intentional content* (the way the world is represented) whereby "if meaning is information, then coreferential representations must be synonyms" (Fodor 1998a:12). This constraint works for the MORNING STAR and EVENING STAR concepts and also, he claims, for the WATER and H2O concepts because each pair of concepts is informationally equivalent and informationally equivalent concepts are semantically equivalent and "semantically equivalent expressions must apply to the same things" (according to Informational Atomism). If, however, there are examples where Condition C fails, then his claim that "informationally equivalent concepts are semantically equivalent" may be in trouble. There are, Fodor states, only two options for an externalist semantics to distinguish between expressions referring to locally instantiated properties: (1) if they are not co-extensive, then an externalist semantics resorts to the use of counterfactuals; and (2) if symbols that are necessarily co-extensive and are co-instantiated as a matter of conceptual or metaphysical necessity, then an externalist semantics has to distinguish them by their *syntax*.

Fodor proposes that psychological laws are *broad* but questions whether broad content is the only kind of content required by psychological explanations. He considers Frege's example of Oedipus and Jocasta: if content is construed broadly then to know that Jocasta = Jocasta is also to know that Jocasta = Mother, but Oedipus believed the first but not the second identity. To explain this, Fodor proposes that propositional attitudes are three-place relations: relations between the creature, a proposition, and a mode of presentation, where modes of presentation are sentences (of Mentalese). Propositional attitudes, like sentences can thus be individuated by their syntax as well as their propositional content. Fodor himself does not seem to be satisfied with this solution, however, given that Oedipus' marrying his mother is a counter-example to the psychological generalization that people generally wish to avoid marrying their mothers. His answer seems to be that perception and cognition ensure that examples such as these do not proliferate (Fodor 1994:49)!

Frege and Mate's cases arise when the linguistic expressions are analysed strictly in accordance with a pure semantic theory of meaning combined with the rules of logic. There is no such difficulty if the expressions are analysed using a pragmatic theory of meaning. Oedipus may use the referent of 'Jocasta' as its meaning (in accordance with the K-P theory). However, while "Oedipus' mother" is a *contingent* property of Jocasta according to Kripke, it is *necessary* if taken as the *transparent* description of a referent by Oedipus. Jocasta was "tagged" as 'Oedipus' mother' when she gave birth to Oedipus and thus a new (albeit related) causal-historical chain was initiated. Rather than taking 'Oedipus' mother' as an opaque description, if we take it as a transparent description, we have a similar situation to Kripke's "Pierre" problem which is a problem of epistemology: Oedipus was unaware that 'Oedipus' mother' and 'Jocasta' had the same referent, just as Pierre was unaware the 'Londres' and 'London' have the same referent. These are not puzzles of logic but tell us something about the nature of beliefs; they are more a concern for psychology than for linguistics. Fodor's fifth thesis in his version of RTM is: "Whatever distinguishes coextensive concepts is ipso facto 'in the head'" (Fodor 1998a:15), but, in order to solve co-extension/co-reference problems while still maintaining an externalist semantics, Fodor needs to postulate a "something else", namely "Modes of Presentation", for individuating co-extensive but distinct concepts. Theories of concepts, such as Fodor's Informational Atomism that postulate an *externalist* theory of meaning, that is, theories of concepts that postulate that contents have a *broad* content, have difficulty explaining how:

- Distinct concepts in a single mind/brain can have the same referent (have the same *broad* or intentional content) — Frege's Puzzle and Mates' Cases;
- An identical concept (identical psychological state; identical narrow content) in two different minds/brains can have different referents (one concept could have different *broad* or intentional contents)— Putnam's "Twin Earth" problem; or
- One concept in a single mind/brain can have more than one referent (have the same *broad* or intentional content) — "Jade" problem.

Fodor's solution is to postulate an internal "something", namely modes of presentation (MOPs), as a solution to Frege's Puzzle and Mates' Cases. A solution to the other two would require different causal-cum-nomological relations between a single concept and different referents.

In particular reference to Frege's Puzzle, he contends that beliefs about the Morning Star and beliefs about the Evening Star have the same conditions of semantic evaluation, but they are different beliefs with different causal powers; they involve the tokening of different syntactic objects (Fodor 1990a:39). This allows synonymy to be defined at the level of mental representations: two expressions are synonymous just in case they are represented by the same mental representation[16]. His MOP solution cannot be used for the second and third problems however, because neither *being nephrite* and *being jadeite* nor *being* H_2O and *being* ***XYZ*** are synonymous. Fodor's solution to these two problems is somewhat problematic. Since conceptual (*broad*) content is constituted by a type of nomic, mind-world relation (Fodor 1998a:121), it is unclear how two distinct referents could token the same concept unless the they had identical properties. Jadeite and nephrite, for instance, would both have to exhibit the same property of *being jade.* He would be constrained to claim that locking to *being jadeite* or *being nephrite* through a chemical-cum-metaphysical theory that specifies their essences using a different mechanism of semantic access from the theory in place prior to finding out that jade has two different chemical compositions (mutatis mutandis for *being* H_2O and *being* ***XYZ)***. While this solution might be acceptable, it does seem to imply that reference is not the only determiner of broad content. *Being jade* is *being jadeite or nephrite,* and *being water* for Putnam's Twins is *being* H_2O *or XYZ*. Indeed, Putnam claims in his Twin Earth thought experiment that it is possible for two concepts to differ in their extensions but ascribe the same properties. This leads to the position that conceptual content is determined by both reference and property ascription which does not seem compatible with an atomic theory of concepts.

Many examples of co-extensive concepts (and co-referring terms) include context and/or additional information which results in their being non-substitutable in doxastic[17] contexts. A causal theory of reference, as presented by Kripke and Putnam, has no mechanism for including such additional information. Further, we need to separate discussions of entities (both objects and events) in the world from the names or terms we use to refer to them. As Bas van Fraassen has pointed out, scientific theoretical terms are introduced or

[16] How a connectionist model of concepts handles synonomy and linguistic puzzles is discussed in section 5.2.4 "Concepts and Categorization".

[17] 'Doxastic' relates to beliefs and similar attitudes such as 'think', 'hope', desire', want'. It is possible for a rational person to believe (think/hope, etc.) that Cicero is a good man and that Tully is not even though Tully and Cicero are the same person.

modified as required for theory construction (van Fraassen 1980:23-25), but this position can be extended beyond the requirements of scientific theories to all theories or beliefs about mind-independent entities.

2.3.10 An Internalist Theory of Meaning

The theories of meaning discussed in the previous sections take a literal view of meaning; they relate linguistic expressions to entities in the world (the K-P theory of meaning) or they relate denotation to the bearer of the name or the meaning of a predicate to the property it expresses (Fodor's informational theory of meaning). An internalist theory of meaning is, I contend, not only more psychological plausible, but also avoids the paradoxes of co-reference and co-extension which plague referential theories of meaning, can explain how metaphors can have meaning, and so on.

2.3.11 Reference and Interpretation

Kripke's development of his theory of reference applied to proper names. The question is whether the theory can be extended to other entities. Illustration 3 and Illustration 4 show two famous examples of ambiguous figures. Each of them has a "proper name", "Jastrow's Duck-Rabbit" and "Necker Cube" respectively, each of them can be "seen as" in more than one way. Ambiguous figures such as these highlight the problem of holding that sensory stimuli alone can determine perception: the same stimuli can result in more than one interpretation. Processes of interpretation are required for sensory stimulations to have meaning. Both illustrations are merely two-dimensional drawings. The *reference* of the DUCK concept and the *reference* of the RABBIT are both the two-dimensional drawing of Illustration 3. Both concepts have the same reference but, in this case, there is no *duck* nor *rabbit* in the external world. What appears to be happening is that the ambiguous image is linked to more than one concept but, since consciousness is sequential, only one meaningful image can enter consciousness at a time. This type of process is discussed in

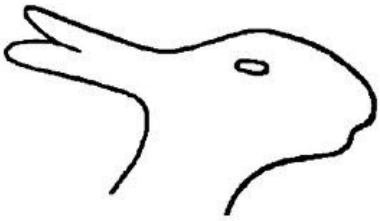

Illustration 3: Jastrow's Duck-Rabbit

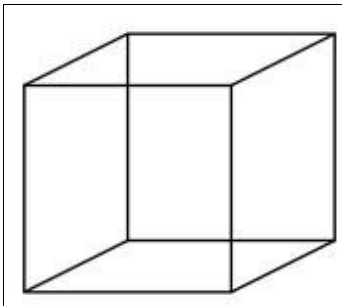

Illustration 4: Necker Cube

more detail later in the dissertation, but, very briefly, the process appears to be that, unconsciously, input becomes meaningful by being linked to a concept if a relevant one is available, or even to more than one, as, for example, in the case of ambiguous figures; this process takes place in a massively-parallel system; the meaningful images become input to the conscious mind which, being a sequential system, can only process one image at a time.

Perception of the world hence appears to depend, at least in part, on how the percipient conceptualizes the world. We construct visual representations of the perceived world. This is clearly the case when it comes to works of art such as that shown in Illustration 5: the reference may be an oil-painted canvas but how each viewer views it depends on his or her own "theory" of aesthetics. As Magritte himself commented:

> The famous pipe. How people reproached me for it! And yet, could you stuff my pipe? No, it's just a representation, is it not? So if I had written on my picture 'This is a pipe,' I'd have been lying! (cited in Harry Torczyner, *Magritte: Ideas and Images*, p. 71.)

Illustration 5: René Magritte's "The Treachery of Images", 1928-29

Other common examples of interpretation are the phonemic restoration effect[18] and completion of fragmented pictures. Both of these examples demonstrate the use of background knowledge, of prior beliefs, and of expectations.

When it comes to language interpretation, Daniel Dennett claims that even the most obvious text is interpretable only with the aid of rather obvious assumptions. For instance, a non-English speaker will interpret Illustration 6 below as having the same symbol in the middle of each group, whereas an English speaker will find it effortless — to the point of not noticing — to see "the cat" as the probable interpretation of this symbol string (Dennett 1990). Again, we are left with the question of what the actual reference might be.

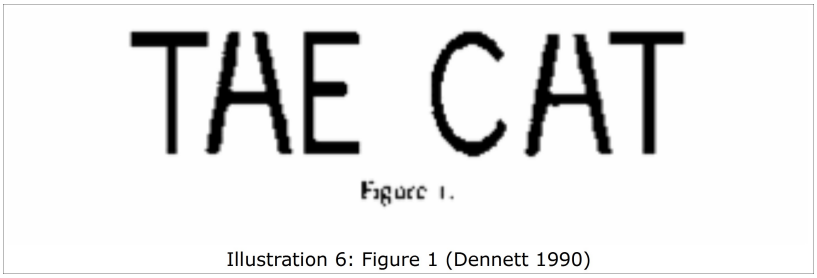

Illustration 6: Figure 1 (Dennett 1990)

[18] Richard M. Warren (1970) conducted experiments in which one phoneme of a word was replaced with a cough-like sound. His subjects had no difficulty in "restoring" the missing speech sound perceptually. Interestingly, they were not able to identify accurately which phoneme had been replaced.

An extreme view of this interpretationalist position (see, e.g., Kuhn 1962; Goodman 1968/1976) holds that two different people experiencing the same phenomenon would "see" radically different entities; what they "see" would depend on what background theories they have about the world, on their "web" of beliefs. Such a radical view would, however, require that, to share a concept, it would be necessary to share the same background theories. There is, however, considerable empirical evidence to support the universality of many concepts. B. Berlin and P. Kay (1969) proposed a theory of meanings for basic colour terms based on their observation of semantic universals and evolutionary regularity in the naming of colours across languages, despite the differences in cultures and background theories, and also the ability to identify colour differences even in the absence of colour terms. This suggests that many concepts (at least sensory concepts) may be innate and thus require no background theories for their acquisition – this is discussed further in Chapter 3.

The phenomenon known as "blindsight" provides evidence that conscious awareness is not necessary for reacting appropriately to visual stimuli. People with blindsight may report losing part or all of their visual field. These patients do however *perceive* visual stimuli even though they have no conscious awareness of them. The patient suffering from blindsight may have damage to visual pathway V1 or visual pathway V2 which are shown in Illustration 7. In Type 1 blindsight, subjects have no awareness of any visual stimuli whatsoever, but are nonetheless able to predict, at levels that are significantly greater than chance, visual aspects such as location, or type of movement. In Type 2 blindsight, subjects have some awareness of visual aspects such as movement within the blind area, but have no visual percept. David Armstrong (1968) suggested that "In perception the brain scans the environment. In awareness of the perception another process in the brain scans the scanning." It is possible to react appropriately to the stimuli without any conscious awareness. The lack of the need for conscious awareness to process perceptual input is discussed in more detail later in this dissertation.

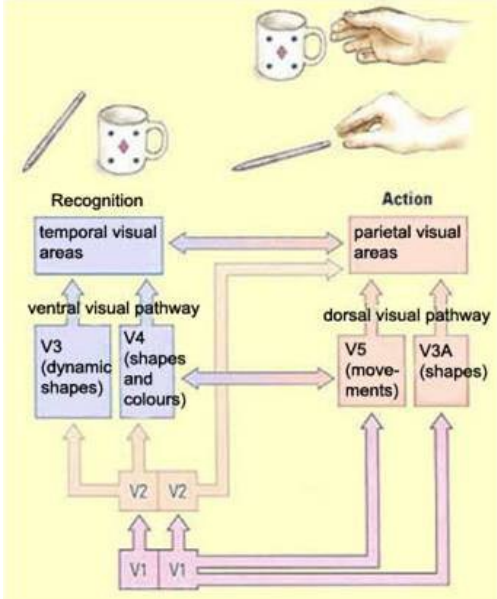

Illustration 7: Visual Recognition and Action (© McGill University)

2.3.12 **Metaphors and Figurative Language**

In developing his causal theory of reference, Kripke used the theory of proper names, but proper names can be used in many ways. A proper name may be used when referring to a person (an a*ct* of referring), to call the person, to question the person, and so on. The different usages are indicated by differences in intonation but the name-reference link is clearly insufficient to distinguish between the usages. In "Homer was a talented man", 'Homer' could be referring to the person "baptized" Homer or, as is more likely, to whosoever wrote the "Odyssey" and the "Iliad". 'Homer' in this case might be used in the same manner as 'anonymous' is used – the name is an abbreviated description. It is unclear how externalist theories of meaning can distinguish between a proper name *qua* proper name and a proper name that is an abbreviated description other than to deny that proper names are ever abbreviated descriptions. Externalist theories of meaning also have difficulty with metaphors, and like figurative language: it is unclear what the reference or denotation of 'sun', or the property expressed by the predicate "is the sun", in a metaphor such as the following might be:

(1) Juliet is the sun (*Romeo and Juliet*, Act. II, sc. 2)

The sentence (1) is, presumably, not a statement of identity. Even if we treat "the sun" as a simile ("Juliet is like the sun"), it is not clear what property is being expressed, and yet any speaker of English is able to understand the sentence and what is being expressed without difficulty even if he cannot express it explicitly. That we cannot express explicitly what is being communicated by many metaphors, seems to preclude, in such cases, that reference is a "social phenomenon" as Putnam contends (even though, 'sun' is a natural-kind term.) Audience members in Shakespeare's time were able to understand the large number of his newly-minted metaphors without any social interaction, without any causal history following the original "baptismal event". It is doubtful that audience members in Shakespeare's time would have attached the same meaning to 'the sun' as today's audience members would – most people in Shakespeare's time would have believed in a geocentric universe. It is also doubtful that many individual audience members in any age would attach exactly the same meaning to the phrase, but that is not crucial to understanding what Shakespeare was trying to convey – the individual "meanings" only had to be close enough.

According to connectionist theories of concepts[19], when networks of different structures are trained on the same task, they develop activation patterns which are strongly similar, thereby suggesting that it might be possible to produce empirically well-defined measures of similarity of concepts and thoughts between different individuals as is required for concepts to be shared. On this view, semantic identity is not based on causal connections to the external world, but is based on an *internalist* account of sameness and similarity (Churchland 2007:134). Rather than having a semantic and syntactic structure similar to that of a natural language, as some other theories of concepts posit, internal representations are sub-patterns that include the micro-features that are specific to the context. This would help explain variations between and within groups of language users. For instance, one person's internal representation of DOG might include the micro-feature +DANGEROUS and yet would be able to use the term 'dog' with other people whose internal representations of DOG do not include +DANGEROUS. All that would be required is that the sub-patterns of their DOG representations be similar enough for the same term to be applied without confusion.

2.3.13 Concepts, Categories, and Meaning

Relevant cognitive processes retrieve the same body of knowledge, the same mental representation of an entity, from long-term memory whether we categorize the entity, draw an inductive inference about the extension it belongs to, draw an analogy between it and some other entity or entities, or even understand a linguistic expression in which a term referring to the entity appears. In other words, a concept is copied (unconsciously) from a class of concepts stored in semantic memory to short-term memory, perhaps along with other non-conceptual knowledge of the item in question, for use in a cognitive process (e.g., deduction, induction, categorization). The concepts form semantic networks and, in conjunction with the "relevant cognitive processes" are, I claim, what makes a mental representation meaningful.

A semantic theory has to be able to account for a person's implicit knowledge, that is, for the knowledge which the subject may not be able to articulate but which, nevertheless, is an authentic propositional attitude. This implicit knowledge is different from tacit knowledge where the information-bearing state to which tacit knowledge refers is not an authentic propositional attitude. Language users are not always able to articulate what a word means for them: they may be able to provide an ostensive definition for many objects

[19]Support for connectionist theories of concepts is presented in Chapters 3 and 5.

but Quine's (1960) indeterminacy of translation problem arises for ostensive definitions. As discussed in Chapters 3 and 5, only the connectionist model has an account of implicit knowledge (non-conceptual content), that is, of the knowledge which the subject may not be able to articulate but which, nevertheless, is an authentic propositional attitude. Expert judgement, at least, is not naturally modelled as the *sequential interpretation of a linguistically formalized procedure,* posited by the Computational Theory of Mind, which takes place at the conceptual level. Intuition, for example, must be formalized at the sub-conceptual level (Smolensky 1988:§6).

2.3.14 The Term-Reference Link

Much is made in reference theories of meaning of an introducing event of a name that involves ostension. The reference of a term is determined by society (through division of linguistic labour). By his claim that "reference is a social phenomenon", Putnam means that individual users of a term do not have to know how to distinguish reliably the token referents of the term; they do not have to know how to distinguish elms from beeches, or how to distinguish aluminium, for example, because they can always rely on experts to do it for them. But how is this accomplished? If an expert chemist points to a nugget of gold and say "gold" or even "Au", how is the non-expert to understand to what 'gold' or 'Au' refers? As stated earlier, Locke noted that natural kind terms can group objects in different categories depending on the properties under consideration. Thus, when a sample of gold is pointed to, it is not clear which natural kind the sample instantiates. There is still the problem of fixing the reference of 'gold' to gold rather than to metals or elements, or even to nugget or yellow or shiny.

In the case of proper names, if someone points to a person and says "John" to a young child, how is the child to understand the referent? The child could take the the reference to be to a person, or adult, or man, or even "undetached-human-parts" (to paraphrase Quine). As Quine (1960) pointed out in his discussion of the indeterminacy of reference, no unique interpretation of a word is possible, because the meaning of a word varies with context. Further, I contend that it is not possible to learn the unique reference of a term merely through ostension unless the hearer already has the concept which the term is used to express. When a child points to a dog and asks what it is, (that is, asks for the term that is used to refer to that entity), the child already has a DOG concept no matter how incomplete. When a child learns the term 'dog' through an ostensive definition in the presence of

one token of *dog*, how is the child then able to apply the term correctly to other dog-tokens given the vast difference in the phenomenological properties of different breeds of dogs? Merely providing the child with a term-reference link will not provide a child with the meaning of 'dog'. If the dog-token that was used to introduce the child to the 'dog' term was a poodle, it is hard to explain how the child would then be able to apply the term correctly to a Great Dane – there is a "poverty of the stimulus". The explanation has to include a psychological component. The idea that children have an essentialist disposition is questionable. Evidence suggests that children tend to acquire base concepts like DOG before concepts at higher or lower levels (e.g., ANIMAL or POODLE) but this is not a justification for the "essentialist" position that a reference theory of meaning would require; it can be accounted for using Fodor's theory of concept acquisition, namely that concepts are acquired ("locked to") in virtue of experiences of typical instances of stereotypes. This "locking to" normally occurs prior to learning the term which expresses the concept. In other words, the term-reference link occurs after the concept is acquired. Such a process also requires some innate mechanisms: there is, for instance, neurological evidence that there are distinct neural mechanisms dedicated to the processing of different word categories such as animal, vegetable, and even furniture, as well as grammatical distinctions between nouns, verbs, and adjectives. Studies like these provide a basis for biologically-motivated theories of language processing; for supporting the view that there are neural structures or processes for the categorical organization of lexical knowledge (Caramazza, Hillis, Leek, & Miozzo in Hirschfeld & Gelman 1994:68-84). Studies of patients exhibiting category-specific impairment provide evidence that, no matter how many times the link between a lexical term and a referent is pointed out to these patients, they will not grasp the *meaning* if the referent is a token of the impaired category. The mechanism for processing the word categories has to be "in the head"; the meaning of the term is not mind-independent even if the term-referent link is.

2.4 PART III: CONCLUSIONS

As argued in more detail later in this dissertation, certain thought processes may be encoded in a "broadcast" form – that is, "broadcast" to the conscious mind. This process is consistent with Fodor's claim that, for any utterance within its domain, the language system module would provide a type-individuated representation specifying the linguistic and (possibly) the logical form of the utterance as input to the central, general purpose cognitive system (Fodor 1983:90-91). If the long-term memory contains relevant lexical concepts

and, depending on the thought or input, relevant sentential concepts, along with, perhaps, relevant encyclodaedic knowledge, the thoughts are encoded as a semantic representation and combined with a linguistic (usually phonetic) representation to produce a linguistically-encoded *icon*[20] in the conscious mind that is "tagged" as *meaningful* (see Jackendoff 2012 for a similar position.) If there are no lexical concepts but there are other relevant perceptual concepts in long-term memory to link to visual image input, the image that is "broadcast" to the conscious mind is a *meaningful* but non-verbal icon. Icons in our conscious mind may appear as *meaningful*, even in the absence of lexical and sentential concepts. All icons that appear in the conscious minds of non-human animals and pre-linguistic children can still be meaningful but are not linguistically encoded. This "broadcasting" of images to our conscious mind is an *internal* process. Visual or aural input itself has no "meaningful" tag. The link between the "meaningless" input and the meaningful conscious image is created "in the head". That is, the unconscious mind creates a link between the input or thought with semantic information stored in long-term memory to render the input or thought meaningful. *Meaning* does not arrive attached to a sequence of phonemes or to a viewed printed text. Without the relevant conceptual knowledge, the input is meaningless regardless of whether there is a socially-maintained link between a term and a reference.

I contend that an externalist theory of meaning, whether linguistic or psychological, can explain, at best, how technical, theoretical terms/concepts can fix and maintain reference. When children point to an item and ask for its name, they obviously have the concept. The word itself is merely an *epistemic artifact,* to use an expression of Clark (2003); it is a tool, like a pencil or a computer, but, like those other artifacts, has no meaning in itself, just as the pixels on a television screen have no meaning in themselves until they are interpreted by the viewer. While a linguistic community (or an expert subset of the linguistic community) may, as Putnam (1997) argues, maintain the link between a linguistic item and a referent or extension, it is the link between the referent and the name which is being maintained, not the meaning. The linguistic item is used as a tool to express the mental representation; to link the mental representation of the speaker with the appropriate (or close enough) mental representation of the hearer/reader.

The attempt to base a theory of meaning on the causal theory of reference may have some success for singular terms and for natural kind terms treated as singular terms, but is

[20] The use of the term "icon" here is owed to Hoffman's analogy of icons in the interface of a PC (Hoffman 2009) in that icons hide the underlying complexity.

less successful when applied to general and abstract terms, particularly when applied to members of sets that have no common, mind-independent, essential nature. It is even questionable that *any* objects of our experience, even natural-kind entities, are truly mind-independent. As Putnam now maintains, objects in the world are "*as much made as discovered*, as much products of our conceptual *invention* as of the "objective" factor in experience" (1981:54 italics in original).

Such being the case, then, a theory of meaning has to reflect conceptual schemes. What we *mean* when we use the term 'water' is not what we *mean* when we use the term 'H_2O' even if the two terms do have the same reference. What we mean in each case is interest-relative and depends on the concepts that are linked to the thought before we use the terms to express that thought. Different people's hearing (or seeing) the two terms will each result in different conscious linguistically-encoded icons simply because the concepts (and encyclopaedic knowledge) that are used in creating the semantic representation of each of them will be different. My water-thoughts may not be identical with your water-thoughts and, in fact, are unlikely to be so. All that the two sets of thoughts have to be is "close enough" so that when I express my water-thoughts by saying "water", the semantic representation produced in your (unconscious) mind is close enough to mine that our resulting behaviour appears appropriate to each of us. On this view, the only way that co-extensive concepts can be individuated is "in the head". We are able to "share" our thoughts and the related concepts, by expressing them linguistically, but the language we use cannot be analyzed using the rules of logic as though it were a mind-independent object.

Natural language, whether spoken, signed or written, comprises symbols to which meanings are associated These symbols may be combined syntactically and can be used as a form of communication both with oneself and with others. Meanings are associated with symbols through social interaction of a linguistic community, just as Putnam told us, but this activity requires intentionality and the recognition of intentionality in others. In other words, the use of a natural language to communicate with others requires that we take an intentional stance (Dennett, 1971); that we assume that, when a con-specific produces a particular signal, he associates the same, or a similar-enough meaning to the signal as we do. When I encountered a dog for the first time, the resulting DOG concept could have referred to the category of dogs, or mammals, or animals, or pets, etc. So, when I uttered the term 'dog' its extension was not necessarily that of my mother's DOG concept. When I say

'water', am I referring to the theoretic or Natural-Kind *qua* Natural Kind, and the non-theoretic (non-theory-mediated) Natural-Kind *simpliciter?*

Just as there is no sound in the external world — sound being the result of pressure waves impacting an auditory system — there is no meaning external to a cognitive system. No sound without audition; no meaning without cognition.

---

3. COGNITIVE ARCHITECTURES

3.1 INTRODUCTION

A fundamental assumption of cognitive science is that certain behavioural regularities can be attributed to different mental representations and to symbol-manipulation processes operating over these representations. Connectionists counter this assumption by proposing that these behavioural regularities can be attributed to simple processing units and connections between them. In this section I examine these alternate cognitive architectures and their fitness in explaining many of the crucial aspects of cognition especially those related to modularity of mind, propositional attitudes, the Language of Thought hypothesis, natural language processing, learning, abductive reasoning, and the problems of co-extension, co-reference and synonymy. Arguments are offered in support of connectionist cognitive architecture models as offering better explanations of such aspects of cognition.

Chapter 3 is divided into three parts:

- Part I provides an analysis of various views concerning metal representations and an overview of the Representational Theory of Mind which is accepted by supporters of several competing models of cognitive architecture;
- Part II presents the two main competing models of cognitive architecture, namely the Computational Theory of Mind and Connectionism;
- Part III compares how these models claim to handle crucial aspects of cognition;
- Part IV presents conclusions and support for the view that connectionism provides the best model of cognitive architecture.

3.2 PART I: REPRESENTATIONALISM

If the hypothesis that the brain is a computer is to make a definite claim, then the word 'computer' must have a precise meaning. The traditional position is that a computer is a universal symbol system; it is a a device that processes *representations* consisting of strings of symbols; it processes representations consisting of compositional, recursively-structured, quasi-linguistic strings. The mind can construct symbols and manipulate them in various cognitive processes. The mind can relate the symbols to something in the world, as when verifying a description; or to something not in the world, such as an image of a hypothetical

state of affairs. Perception leads to the construction of mental symbols representing the world. On this view, there is a causal link from the world to an internal representation. Representation evolved as a result of natural selection: it is required for safe navigation around the world; it provides the processes governing actions with information about what is where; and it is recovered over several stages of visual processing – it is not, for instance, explicit in the patterns of light falling on the retina (Johnson-Laird 1988:34-35).

A *concept* is usually taken to be a type of mental representation that aids in organizing experience. Without a capacity to organize, experience would otherwise be what William James called the baby's impression of the world, namely "one great blooming, buzzing confusion" (James 1890:462). Much of the perceptual information we receive is ambiguous and so, without an ability to categorize, experiences would comprise an infinite number of unrelated objects, properties, sensations, and events. An example such as that given in Illustration 8 can demonstrate the power of the mind to organize perceptions based on an existing concept. The illustration can be viewed as a collection of random dots or can be organized into a picture of a Dalmatian dog in an uneven landscape. Once the pattern of dots has been so organized, it is difficult, if not impossible, not to "see" a dog in the picture.


Illustration 8: Example of "seeing as"

The results of research into cognitive development show that conceptual categories are formed by children from earliest infancy and that these categories are very similar to those of adults (Gelman 1999; see Gelman 1996 for review). Even though there is widespread agreement among cognitive scientists that concepts pick out categories, that concepts *refer* to categories, there is not such an agreement on just what concepts might be. Psychologists (for example, Solomon, Medin, and Lynch 1999; Markman 1999; Carey 2009) and many philosophers of mind (Fodor 1987; Dretske 1988; and others), take concepts to be a subset of mental representations which are used in various cognitive processes such as inference and induction, as well as categorization. The term *mental representation* is defined slightly differently when used in philosophy, psychology, and cognitive science. In philosophy, mental representations are viewed as mental objects with semantic properties such as content, reference, and truth-conditions. They are "mentalia" (thoughts, concepts,

percepts, ideas, notions, schemas, images, et cetera.) In cognitive science, mental representations are often viewed as information-bearing structures, and cognitive states are constituted by the occurrence, transformation, and storage of these structures. Susan Carey, along with most psychologists, assumes that "representations are states of the nervous system that have content, that refer to concrete or abstract entities (or even fictional entities), properties, and events" (Carey 2009:5).

3.2.1 **Mental Representations**

Any claim that mental states have content leads to a metaphysical commitment to mental representations. Among those who support the existence of mental representations, there is by and large agreement with David Marr's (1982) definition of mental representations as "a formal system for making explicit certain entities or types of information, together with a specification of how the system does this." Such mental representations are viewed as a formal system combining hypothetical internal cognitive symbols (such as thoughts) and the mental processes (such as inferences) that use such symbols.

Some thinkers have presented cognitive theories that assume no commitment to mental representations. B. F. Skinner (1938, 1950, 1990) and other radical behaviourists held that inferred, or unobservable entities had no place in a scientific psychology, and there are many who deny that there are mental representations of any kind (e.g., Dennett 1987; Gibson 1966, 1979, Stich 1983). These "eliminativists" about mental representations contend that psychological theorizing should be couched in neurological or behavioural terms rather than in terms related to representational mental states. As C. R. Gallistel (2001) has pointed out, mental representations are not neurobiologically transparent; that is to say, neurobiology, in its current state, is unable to describe how the hypothetical internal cognitive symbols and the mental processes that use such symbols might be realized neurobiologically. It is not surprising, therefore that those who attempt to eliminate mental representations from psychological theorizing (for example, Edelman and Tononi 2000, Hull, 1930, 1952, Rumelhart and McClelland 1986) ground their psychological theories in neurobiology (Gallistel 2001:9691). This leads some eliminativists to claim that many "folk psychology" concepts, such as belief or desire, are not well-defined, that common-sense understanding of the mind (*folk psychology*) is false, and that, in consequence, they have no coherent neurobiological basis — it will never be possible to explain such concepts in terms of lower-level neurobiological processes. This argument in rejection of folk psychological concepts

seems to be self-refuting given that folk psychology as a theory has had considerable success in enabling an everyday communication with few words and that such efficiency is not achievable using a complex, neuroscientific terminology (Fodor 1987, 1999).

Grounding their psychological theories in neurobiology does present some difficulties for Eliminativists in describing what a cognitive architecture might consist in, what would make a particular phenomenon *cognitive*. For representationalists, a cognitive level is one in which mental states encode features of the world, and a cognitive architecture is thus an architecture of representational states and processes. While some supporters of a connectionist cognitive architecture, e.g., Paul and Patricia Churchland, argue that folk psychology should be eliminated in favour of a more scientific theory based on neuroscience, they nevertheless are still representationalists. Both the Classical and Connectionist cognitive architectures (such as that supported by the Churchlands) are based on the reality of representational states, although there are extreme eliminativist versions of both types of architectures (see, for instance, Ramsey, Stich, and Garon, 1990).

A main impetus for postulating the existence of mental representations is finding an explanation of common-sense, dispositional psychological states (such as thoughts, beliefs, desires, and perceptions) and processes (such as inference, categorization), and of how such states are transformed and stored in the mind/brain. According to Representational Theories of Mind (RTM), these psychological states have *intentionality*—they are *about* or *refer* to entities—and their intentionality can be explained in terms of the semantic properties of mental representations. Cognitive mental processes (thinking, reasoning, and such like) are hence taken to be sequences of intentional mental states. While there may not be unanimity among those involved in cognitive science on the actual types of structures and processes of intentional mental states, there is wide agreement that they are nevertheless explainable in terms of mental representations.

Mental representations (thoughts) express the propositions that are the objects of propositional attitudes; and propositional attitudes are both productive and systematic. As constituents of thoughts, and of each other, concepts play a particular role in explaining this productivity and systematicity, and it is the compositionality of concepts and thoughts that provides the explanation (Fodor 1998a:34-36). Compositionality is therefore essential to any theory of concepts, where compositionality is defined as the property that a system of representation has when:

a) it contains both primitive symbols and symbols that are syntactically and semantically complex; and

b) the latter inherit their syntactic/semantic properties from the former.

Fodor's "non-negotiable" condition number 3 states that "concepts are the constituents of thoughts and, in indefinitely many cases, of one another. Mental Representations inherit their contents from the contents of their constituents" (Fodor 1998a:34). Some concepts are complex and the syntax and the content of the constituents of a complex concept determines its syntax and content (Fodor 1998a:104). That is, complex concepts are composed of primitive concepts. The argument that concepts compose is primarily that they are productive and systematic, which allows Fodor to stipulate that:

> "the claim that concepts compose is true only if the syntax and content of complex concepts is derived from the syntax and content of their constituents *in a way that explains their productivity and systematicity*" (italics in original, Fodor 1998a:104).

Fodor and Lepore enlarge on this in "The Red Herring and the Pet Fish" (Fodor & Lepore 2002) in which they present the following argument:

P1. Concepts are productive because there are infinitely many mental representations;

P2. There are infinitely many mental representations because new, relatively complex mental representations can be constructed by using old, relatively primitive ones as their constituents;

P3. That mental representations have constituent structure is thus essential to explaining the compositionality of concepts;

Conclusion: Compositionality serves to specify the constituency relations that mental representations enter into.

Here, compositionality is a function that "maps a finite basis of simple mental representations onto an infinity of complex mental representations together with their structural descriptions" along with an interpretive function which maps arbitrary mental representations, simple or complex, into their semantic interpretations (whatever they may be). Their interpretive function specifies a relation between mental representations and things in the world (for example, between representations and their extensions).

3.2.2 **Empirical Evidence for Mental Representations**

There is empirical evidence from studies of both human and animal behaviour that provide support for underlying mental representations some of which are likely innate. Studies by Andre Meltzoff and colleagues, for example, demonstrate that infants (even neonates) who are given pacifiers of specific shape and texture are able to recognize the correspondence between the tactile and visual presentations of those shapes and textures (Meltzoff & Borton 1979), and that neonates will imitate facial expressions (Meltzoff & Moore 1977, 1997). Results of the latter studies are sufficient to demonstrate that there are innate supramodal representations of the correspondence between what one's face feels like and what another looks like. Illustration 9 provides the conceptual schematic of the 'active intermodal mapping' hypothesis that Meltzoff and Moore (1997) propose as the basis of early facial imitation.

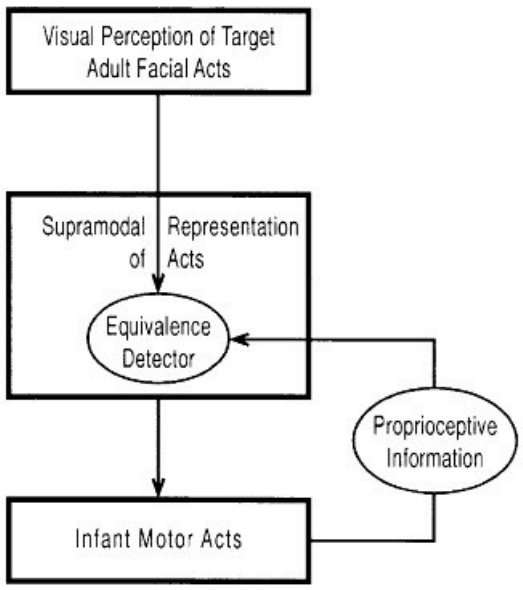

Illustration 9: A conceptual schematic of the active intermodal mapping hypothesis (Meltzoff & Moore 1997: Figure 1)

Experiments conducted by Richard Herrnstein and colleagues show that at least one non-human species, pigeons in this case, are selectively sensitive to stimuli. In these experiments, pigeons were presented with several photographs some of which depicted trees and others depicted similar-looking objects such as the tops of celery stalks. Despite the fact that the photographs were taken from a variety of different perspectives, the pigeons proved to be very skilled in sorting the photographs that depicted trees from those that did not. Furthermore, the pigeons proved to be as skilled when presented with several other categories such as *flower*, *fish*, *human*, and even *automobile* (Herrnstein 1979, 1984). It would appear that pigeons are causally responsive to groupings of objects that have similar extensions to those of certain categories of human cognition (to use a modularity of mind thesis, both humans and pigeons have visual modules that process pattern-recognition routines) and, since *automobile* was included in the Herrnstein studies, it also appears that the pattern recognition by pigeons are not limited to naturally-occurring objects. This is not to say that pigeons have the same concepts as humans, but it does seem to imply that they

might have proto-concepts that can be employed in categorization. David Premack (1988) has demonstrated that chimpanzees without language training are able to categorize by matching to sample, and to recognize similarities of proportions. There is also empirical evidence for the existence of cognitive maps (allocentric representations of physical space) in dogs, cats, and chimpanzees (Bressard 1988).

Studies indicate that pre-verbal children can distinguish the use of the definite (the particular) from the indefinite (one-of-a-set article) with word shaping occurring almost always at the "Base Level" (Markman 1989). The following table shows the levels of "word-shaping" studied:

Levels of Word-shaping by Children Based on Diagram in *Categorisation and Naming in Children* by E. Markman (1989)		
Word Shaping Level	**Description**	**Examples**
Superordinate	Broad—established by verbal consensus	Animals, Vehicles, Furniture
Base Level	Narrow—class criteria that are directly perceivable	Cat/Dog, Car/Truck, Chair/Table
Subordinate	Narrower still	Siamese/Terrier Volvo/Tipper

Not only do categories embody knowledge, they are also a means of extending it, and, since categories capture information, such as inferring that tigers are carnivorous animals from their jaws and claws, categorization functions to support inductive inferences (Markman 1989:10). While categorization at the superordinate level is established by verbal consensus according to Markman, categorization at the base level requires no linguistic knowledge or capacity as demonstrated by Herrnstein's pigeon studies. What does seem to be required at the superordinate level is some form of compositionality: either union of extensions by consensus, or definitions, or some mechanism for creating complex categories from more uniform ones. Presumably, Markman limits "verbal consensus" to consensus among users of a natural language. Many non-linguistic species appear (by behaviour and vocalization repertoires) to group several other species into what would conform to a superordinate human category of, for instance, *prey,* or, more specifically, *air-borne-threat*, *ground-based-threat*, and so on. What would determine membership in such categories is, of course, species-specific. PREY, PREDATOR, and such-like species-specific concepts could be innate

as some ethological studies suggest (categorization by non-human species is discussed further below).

3.2.3 **What are Mental Representations?**

Mental states are deemed, by philosophers of mind, to be either representational or qualitative. Only the first is of interest when discussing concepts. Theories of meaning tend to follow the two main theories used to explain the nature of concepts, namely: the *Representational Theory of Mind*, according to which concepts are mental representations, and the semantic theory of concepts, according to which concepts are abstract objects — an abstract idea or a mental symbol sometimes defined as a "unit of knowledge," built from other units which act as a concept's characteristics. Not all mental representations are concepts: mental representations are often viewed as a continuum from the sensory/perceptual at one end and the conceptual at the other, where sensory representations are of the phenomenal aspects of the world (what external objects look, smell, feel, sound like, and so on); whereas conceptual representations are either constructed from sensory perceptions and are *about* or *refer* to external entities, or are about abstract entities for which no sensory evidence exists. A mental representation can therefore be defined as any internal state that mediates or plays a mediating role between a system's inputs and outputs by virtue of that state's semantic content; where *semantic content* may be defined in terms of information that is causally efficacious and in terms of whatever it is that that information is used for. In summary, mental representations mediate between environmental stimuli and the behavioural output by being causal at the appropriate level of analysis (see Hofstadter 1979, Marr 1982, Pylyshyn 1984, *et al*).

The term *mental representation* is defined slightly differently when used in philosophy and in cognitive science. In philosophy, mental representations are viewed as mental objects with semantic properties such as content, reference, and truth-conditions. They are "mentalia" (that is, thoughts, concepts, percepts, ideas, notions, schemas, images, et cetera.) In cognitive science, mental representations are often viewed as information-bearing structures; cognitive states are constituted by the occurrence, transformation, and storage of these structures. I will follow normal philosophical usage except where explicitly identified.

3.2.4 **Representationalism**

A fundamental assumption of cognitive science is that certain behavioural regularities can be attributed to different representations and symbol-manipulation processes operating over these representations (Pylyshyn, 1999). A *Representational Theory of Mind* (RTM) holds that thinking occurs within an internal system of representation. It is the thesis that intentional states (such as beliefs and desires) are relations between a thinker and symbolic representations of the content of those states. RTM holds that mental states have intentionality (are about the world) in virtue of a representational relationship holding between the mental state and the object or property in the world.

A *mental representation* is taken to be either a hypothetical internal cognitive symbol that represents external reality, or a mental process using such a symbol. As previously noted, Marr defines a *representation* as "a formal system for making explicit certain entities or types of information, together with a specification of how the system does this" (Marr, 1982:20). A "mental representation" is a basic concept of the theory that cognitive states and processes are constituted by the occurrence in the mind or brain of information-bearing structures or representations. A representation is something that can stand for concrete objects, sets, properties, events, and states of affairs in this world, in possible worlds, and in fictional worlds; as well as abstract objects such as universals and numbers. It can represent both an object (in and of itself) and an aspect of that object (or both extension and intension), and can represent both correctly and incorrectly (von Eckardt 1993:527).

RTM postulates the actual existence of mental representations that mediate between the observer and external entities (objects, events, and so on) – mental representations *represent* to the mind objects or events in the external world. Hence, for most supporters of RTM, it is a form of indirect realism. The proviso "most" is needed because some supporters of RTM also support some form of direct realism according to which mental representations are not required for perception (see Gibson 1966, Reid 1983). Dretske, for one, distinguishes between *perception of facts* and *perception of objects* claiming that, while the former requires "conceptual skills needed to classify and sort perceptual objects" (Dretske 1995:332), the latter may be considered as direct and non-mediated because:

> One doesn't have to know, let alone know for certain …, that there are physical objects in order to see (sense perception) physical objects. Such knowledge is only required for the perception of the fact that there are such objects … it may turn out that we see ordinary physical objects … every moment of our waking life

> without ever being able to know (if the sceptic is right) that this is what we are seeing. (Dretske 1995:338)

The distinction that Dretske proposes would require two different perceptual systems in the brain: perception of facts would require a system for conceptual categorization, and perception of objects would require a system designed solely for objects. The problem, for Dretske, arises in determining the nature of sensory representations; in determining whether they are "what philosophers and psychologists used to call sensations (raw, uninterpreted, sensory givens)" or more what are now called percepts, that is "cognitively enriched (more fully interpreted) experiences of the object" (Dretske 1995:345). It is not clear why Dretske limits perceptual processing to only two systems. It is also not clear how a dedicated system for direct perception of objects could handle problems of object constancy; that is, how an object could be recognized (perceived as an object) in varying viewing conditions such as differences in lighting or orientation, and object variability including size, colour, and so on. An example of interpretation by the human visual system is shown in Illustration 10: in general, the human visual system interprets lighting as coming from above, and so the spheres whose upper part is brighter are interpreted as emerging from the image, while the spheres whose upper part is darker are interpreted as receding. To account for variations in viewing conditions the object perception system would have to be able to find what is common to the object description from varying viewpoints and in the retinal descriptions (Humphreys & Quinlan 1987). It is not plausible that an object could be picked out "directly" in a visual field such as that presented in Illustration 8 on page 39 without some more complex form of object recognition process such as the use of structural information as proposed by viewpoint-invariant theorists (Peterson & Rhodes 2003); or the matching of 3-D model representations of the visual object with 3-D model representations stored in memory (Marr and Nishihara 1978). Visual illusions, such as the Müller-Lyer illusion discussed previously, are evidence that direct realism, even Dretske's that is restricted to objects, is not possible. In contrast to direct realists such as Reid, Gibson, and Dretske, Fodor contends that direct perception of distal objects is impossible and that what is required is a necessarily-mediated and indirect, inferential theory of perception. At the

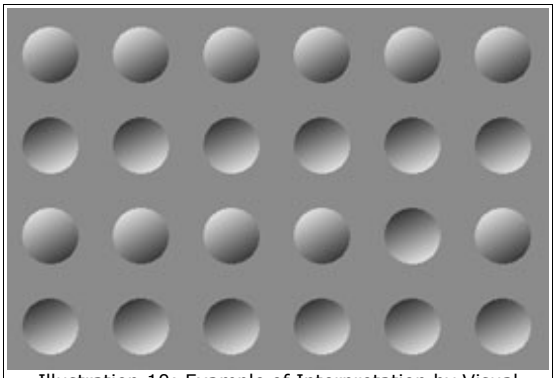

Illustration 10: Example of Interpretation by Visual System

base of the inferential processes, he claims, is a computational mechanism using representations encoded by the transducers (Fodor, 1983:42).

For those committed to a their existence, mental representations are assumed to be semantically-evaluable objects (having content, reference, truth-conditions, and so on); they may be construed as *mental objects with semantic properties*. Consequently, mentalia (thoughts, concepts, ideas, percepts, ideas, etc.) exist as semantically-evaluable objects. According to this view, propositional attitudes (beliefs and desires, for instance) are token mental representations with semantic properties. To support RTM is to be functionalist about mental states, where mental states are classified (or typed) by their function; and to be realist about propositional attitudes, where propositional attitudes are real mental entities that enter into causal explanations of behaviour. Mental states in general represent reality and have information-providing functions which, along with their structural differences, are used by *representationalists* to account for the intuitive differences between conceptual representations (thoughts, judgements, beliefs) and sensory representations (experiences, sensations, feelings); they are used to distinguish "seeing and hearing from believing and knowing" (Dretske 1995:9). The distinction between experiences and thoughts is based on the origin and nature of their functions: sensory representations are states with *systemic,* phylogenetic, indicator functions, whereas conceptual representations are states with *acquired* indicator functions (Dretske 1995:19). Dretske (in common with Fodor, Millikan, Kripke, *et al.*) proposes an *externalist* account of linguistic meaning and of the content of belief[21] (Dretske 1995).

To summarise, *RTM* is the theory that thinking occurs within an internal system of representation and propositional attitudes are token mental representations. In accordance with the terminology used by Fodor (1998a:25) 'thoughts' are the mental representations which express propositions and are analogous to closed sentences; 'concepts' are the constituents of thoughts and are analogous to the corresponding open sentences; and 'propositions' are the objects of propositional attitudes. Propositional attitudes are taken to be relations between an agent and a mental representation (Fodor 1987). For r*epresentationalists,* the immediate object of knowledge is a mental particular rather than the external object which caused the perception.

[21] Dretske extends his externalist account to non-propositional and non-conceptual mental content. He supports this strong externalism by claiming that being an externalist with regard to propositional content, should lead one to be an externalist also with regard to sensations and qualia.

If we accept RTM, then we have a means for distinguishing *content* identity from *concept* identity. This is because RTM "says that to each tokening of a mental state with the content *so-and-so* there corresponds a tokening of a mental representation with the content *so-and-so*" (Fodor 1998a:41). Hence, it might be possible to have different mental representations that correspond to the same content: for example, it would be possible to entertain the mental representation MORNING STAR without entertaining the mental representation VENUS, and, even though the Morning Star and the Evening Star refer to the same object, RTM allows that MORNING STAR and EVENING STAR may be different concepts. While beliefs about the Morning Star and Evening Star have the same conditions of semantic evaluation, they do not involve tokens of the same syntactic objects; they are different beliefs and hence have different causal powers (Fodor 1998a:39).

3.3 Part II: Rival Cognitive Architecture Models

In attempts to understand cognitive processes, whether human or non-human (machines included) researchers in cognitive science have sought means of implementing these processes in some kind of working system. There is no unanimity, however, on the architecture of such systems. Proposed models generally fall into one of two categories of cognitive architecture: symbolic computational systems and connectionist systems. Even within each of these two basic models, there are many different proposals for how different aspects of cognition may be implemented. In this section, the two principle rival architecture categories, Computational Theory of Mind and Connectionism, are described. Their relative strengths and weaknesses are analyzed on Part III.

3.3.1 The Computational Theory of Mind

The Associationists (such as David Hume), held that mental particulars have both semantic and causal properties, and mental processes operate by the association of one state with its successor states. In 1936, Alan Turing formalized the idea of computation, and, in 1950, he published "Computing Machinery and Intelligence", which proposed that: mental particulars are symbols which are syntactically structured; and the causal relations among mental particulars are computational rather than associative. The *Computational Theory of Mind* (CTM) combines an account of reasoning with the *Representational Theory of Mind* (RTM). CTM is the thesis that the mind functions as a computer or, more specifically, as a representation-oriented symbol manipulator in which basic syntactic structures possess a transparent compositional semantics. In other words, the mind has a structure similar to

that of a digital computer where thoughts are viewed as higher level brain states with a syntactic structure; this syntactic structure is isomorphic to the semantic interpretation; and thought processes are computational operations defined over such structures. Computational operations are sensitive only to the formal, syntactic properties of the structures. A modern digital computer can carry out any finitely specifiable task and operates on the basis of internal representations, thereby providing a fundamental analogy for systems of internal representation. Zenon Pylyshyn, a leading supporter of the strong symbol system hypothesis (the conjecture that *only* universal symbol systems are capable of thought) contends that "The idea that mental processing is computation is indeed a serious empirical hypothesis" (Pylyshyn 1984:55), that the mind is "continually engaged in rapid, largely unconscious searching, remembering, and reasoning, and generally in manipulating knowledge" (*ibid*:193), and that this knowledge is "encoded by properties of the brain in the same general way the semantic contents of the computer's representation are encoded — by physically instantiated symbol structures" (*ibid*:258).

Fodor (1998b) finds remarkable the fact that it is possible to tell "just by looking at it" that any sentence of the syntactic form 'P and Q' (such as 'John swims and Mary drinks', for instance) is true only if both P and Q are true, where "to tell just by looking" means to see that the entailments hold even without knowing the meaning of either P or Q, or, for that matter, without knowing anything about the non-linguistic world at all. For Fodor, this really is remarkable since the combination of what P or Q means plus the state of the non-linguistic world is what decides whether 'P and Q' is true. Fodor states that this line of thought is often summarized as follows: "some inferences are rational in virtue of the syntax of the sentences that enter into them; metaphorically, some inferences are rational in virtue of the 'shapes' of these sentences." (Fodor 1998b)

Turing noted that a machine can be made to execute any inference that is formalized in this sense because machines can be constructed to detect and respond to syntactic relations between sentences. If a computer is presented with an argument which depends solely on the syntax of its sentences, then it will accept the argument if and only if it is valid. Such a machine could be said to be rational. Similarly, to a certain extent, what makes minds rational is the ability to perform computations on thoughts, if we assume that thoughts are syntactically structured like sentences and that "computations" means Turing-type formal operations. This, Fodor states, is the theory that forms the basis of Steven Pinker's claim that "thinking is a kind of computation" (Fodor 1998b; Pinker 1999) and, Fodor continues,

> It has proved to be a simply terrific idea. Like Truth, Beauty and Virtue, rationality is a normative notion; the computational theory of mind is the first time in all of intellectual history that a science has been made out of one of those. (Fodor 1998b)

Nevertheless, Fodor (1998b) points out that Turing's account of computation is *local* in at least a couple of respects: it does not look beyond a sentence's syntax to its semantics; and it assumes that the internal syntactic structure of a thought is the sole determinant of its role in a mental process. Yet not all rational processes are likely to be local in either respect, and Turing-type computational rationality is of little use in providing explanations of the semantic and global features of mental processes. Rational beliefs are often formed by *abductive inference*, by 'inferences to the best explanation'. If we are given only what perception presents to us as currently the fact and what beliefs are currently available to us in memory, we have the cognitive problem of finding and adopting new beliefs that are *best confirmed on balance* which Fodor takes to mean something such as "the strongest and simplest relevant beliefs that are consistent with as many of one's prior epistemic commitments as possible" (Fodor 1998b). Nevertheless, such properties as relevance, strength, simplicity, and centrality apply, not to single sentences, but to belief systems as a whole, and, as Fodor states, we have no reason for supposing that such *global* properties of belief systems are syntactic.

A further question arises as to whether all cognitive processes can be said to be *computable* at all. One definition of a *computable function* is given by the Church-Turing thesis which states that every effective (or mechanical) computation can be carried out by a Turing machine given unlimited amounts of time and storage space. The *Stanford Encyclopedia of Philosophy* provides the following statement of the thesis:

> A method, or procedure, M, for achieving some desired result is called 'effective' or 'mechanical' just in case
>
> 1. M is set out in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols);
>
> 2. M will, if carried out without error, produce the desired result in a finite number of steps;
>
> 3. M can (in practice or in principle) be carried out by a human being unaided by any machinery save paper and pencil;

4. M demands no insight or ingenuity on the part of the human being carrying it out.

According to this thesis, any function for which a self-terminating algorithm can be provided is computable, *algorithm* being understood as a sequence of steps that a person could perform given sufficient time, stamina, pencils and paper. There is, however, no universally-accepted definition of “algorithm”. Marvin Minsky asserted that “algorithm” is synonymous with "effective procedure" (Minsky 1967:311), an effective procedure being "...a set of rules which tell us, from moment to moment, precisely how to behave", although, as he points out, this definition may be subject to "... the criticism that the interpretation of the rules is left to depend on some person or agent" (Minsky 1967:106). In order to avoid having to provide, for each statement of the rules, mechanisms for interpreting them, he sought to identify a "reasonably uniform family of rule-obeying mechanisms" formulated thus:

> “(1) a *language* in which sets of behavioral rules are to be expressed, and
>
> “(2) a *single* machine which can interpret statements in the language and thus carry out the steps of each specified process." (italics in original, Minsky 1967:107)

He noted that "there remains a subjective aspect to the matter. Different people may not agree on whether a certain procedure should be called effective" (Minsky 1967: 107), but, nevertheless, introduced his "Turing's Analysis of Computation Process": “[A]ny process which could naturally be called an effective procedure can be realized by a Turing machine" (Minsky 1967:108).

Michael Sipser provided three levels of description of Turing machine algorithms (Sipser 2006:157):

> *High-level description*: "wherein we use ... prose to describe an algorithm, ignoring the implementation details. At this level we do not need to mention how the machine manages its tape or head."
>
> *Implementation description*: "in which we use ... prose to describe the way that the Turing machine moves its head and the way that it stores data on its tape. At this level we do not give details of states or transition function."
>
> *Formal description*: "... the lowest, most detailed, level of description... that spells out in full the Turing machine's states, transition function, and so on."

Daniel Dennett provides a less formal description of an algorithm: “[T]he standard textbook analogy notes that algorithms are recipes of sorts, designed to be followed by novice

cooks", which, if executed correctly, always produces the same results – "[A]n algorithm is a foolproof recipe."(Dennett 1995:51).

It is nonetheless questionable what proportion of mental processes are algorithmic; what proportion are effective procedures in the same manner as a Turing computation process is said to be. As Paul Smolensky argued:

> it is sound to formalize knowledge in linguistically expressed laws and rules for the purpose of science (or any other public knowledge), but it does not follow that knowledge in an individual's mind is best formalized by such rules. (Smolensky 1988:§10)

Smolensky notes, nonetheless, that novices "consciously and sequentially follow rules", and thus it is natural to model the cognitive processing of novices "as the sequential interpretation of a linguistically formalized procedure" (Smolensky 1988:§2.1). This is discussed further in paragraph 3.4.5 below.

Steven Pinker wrote that the key idea of the Computational Theory of Mind (CTM) can be expressed in one sentence:

> The mind is a system of organs of computation, designed by natural selection to solve the kinds of problems our ancestors faced in their foraging way of life; in particular, [the problems of] understanding and outmaneuvering objects, animals, plants and other people. (Pinker 1997:)

While Fodor claims that CTM is "the only game in town" for giving an account of the natural mechanisms underlying thought and cognition, it is only part of the of the truth about cognition (1975).

Fodor has pointed out that Turing-type computational rationality cannot help to explain the semantic and global features of mental processes and that there is no reason to suppose that *global* properties of belief systems (such properties as relevance, strength, simplicity, and centrality) are syntactic (Fodor 1998b). Thus the computational theory of mind can provide explanations for only a subset of mental processes. On the other hand, while the version of RTM that Fodor proposes endorses an account of *thinking* that is computational, it endorses an account of the *semantics* of mental representations that is *non-computational.* The metaphysical view that computational relations determine the semantic properties of their constituent mental representations is thereby rejected along with the idea "that the notion of a computation is *prior* to the notion of a symbol" (italics in original, Fodor 1998a:20). While Fodor accepts Turing's idea that thought is a kind of computation,

he does not accept that *all* causal processes are computational. For Fodor, a mental representation has its content in virtue of its causal-cum-nomological relations to the entities to which it refers; "for example, what bestows upon a mental representation the content *dog* is something about its tokenings being caused by dogs" (italics in original, Fodor 1998a:21). In other words, he combines a computational theory of mind with a causal account of mental content. This leads to Fodor's informational semantics for which:

> *A representation* **R** *expresses the property* **P** *in virtue of its being a law that things that are* **P** *cause tokenings of* **R** *(in, say, some still-to-be-specified circumstances* **C***).* (italics in original, Fodor 1998a:21)

Which is to say that a mental representation **R** means ***X*** as long as tokenings of **R** are reliably caused by **X**s: as long as the tokenings of the concept COW, for example, are reliably caused by cows and tokenings of WATER by H_2O. Given, however, Fodor's claim that conceptual content is information, he cannot then agree that the content of the concept H_2O is different from the content of the concept WATER, especially as he accepts that they are different concepts. Hence, content individuation is insufficient for concept individuation (Fodor 1998a:24). What is required for concept individuation is an adequate account of content that includes a "narrow", psychological component (encoded in syntax) and a "broad" component (which determines reference). This line of reasoning clearly leads to Fodor's position that most concepts are atomic.

The basis of cognitive science is that cognitive mental processes are operations defined on syntactically structured mental representations which are similar to sentences; that is, that mental representations are language-like. The idea is to explain the productive and systematic nature of cognitive states and the mostly truth-preserving nature of cognitive processes. As Fodor (1995:15, 1998a:35, and elsewhere) has pointed out, the systematicity and productivity features of natural languages can be used to illustrate the systematicity and productivity of thought. A natural language is said to be *productive* because there is no upper bound to the number of well-formed sentences or expressions it can contain. A natural language is said to be *systematic* in that the ability to produce/understand/think some sentences implies the ability to produce/understand/think others of related content: a mind that can grasp the proposition that John loves Mary can "in point of empirical fact" also understand the proposition that Mary loves John. That is to say that the systematicity and productivity of thought arise from the compositionality of mental representations, which in turn depends on their constituent syntactic structure. The

tendency of mental processes to be truth-preserving is explained by the hypothesis that they are computations which are, by stipulation, causal processes that are syntactically driven. The *form* of a belief is thus independent of its content.

Pinker (1999) defended the theory that the human mind is a naturally-selected system of computational modules. For Pinker, the computational theory of mind is a key element of cognitive science but Fodor (2000:24-25) argues that there are serious problems with viewing cognition as computational:

Premise 1: Because mental processes are computational, they are sensitive only to the syntax of mental representations;

Premise 2: Because the syntactic properties of any representation are essential, syntactic properties of mental representations are essential;

Conclusion: Mental processes are insensitive to the context-dependent properties of mental representations.

This conclusion seems to be false to Fodor and he offers some counter-examples to demonstrate this falsity:

Simplicity: The complexity of a thought is not an intrinsic property—it is context-dependent—but the syntax of a representation is one of its essential properties and hence does not change when the representation is transported from one context to another. Nevertheless, the computational theory of mind requires that the simplicity of a thought supervene on its syntax.

Abductive reasoning[22]: Abductive inferences can only be computational at the price of a "ruinous holism"; that is, only if the units of thought were much larger than in fact they could possibly be. The problem is "how to reconcile a local notion of mental computation with ... the fact that information that is relevant to the optimal solution of an abductive problem can, in principle, come from anywhere in the network of one's prior epistemic commitments" (Fodor 2000:42).

Conservatism: Estimates of which beliefs count as significantly relevant and which do not when determining the conservatism of a change of theory are context sensitive, but (pursuant to Premises 1 and 2 above) the syntactic properties of representations are not theory sensitive and cannot change with context.

[22] Abductive reasoning is a form of reasoning which relies on the formation and evaluation of hypotheses given the best available information. It is performed through *local* approximations of *global* processes. Each problem has to be solved case-by-case.

Referring to the "persistent failure" to produce artificially intelligent machines, Fodor asserts that "[t]he failure of our AI is, in effect, the failure of the Classical Computational Theory of Mind to perform well in practice" (Fodor, 2000:38). Fodor claims that the pre-theoretical, 'folk' taxonomy of mental states conflates intrinsically intentional natural kinds (beliefs, desires, and the like) with the intrinsically conscious ones (sensations, feelings and the like). He also claims that a main result of the attempt to fit the facts of human cognition to the Classical Turing account of computation is that a comparably fundamental dichotomy among mental processes is also needed; that is, a dichotomy is needed between local and non-local mental representations. Further, there is a characteristic cluster of properties that typical examples of local mental processes reliably share with one another but do not share with typical instances of global ones. The features which Fodor finds most pertinent for his purpose are:

- that local mental processes appear to fit Turing's theory that thinking is computation;
- that they appear to be largely modular; and
- that much of their architecture, and their knowledge about their proprietary domains of application, appear to be innately specified.

As defined by the Classical Turing account of cognition, mental processes are causally sensitive to, and only to, the syntax of the mental representations that they are defined over; in particular, mental processes are not sensitive to the *meaning* of mental representations. Also, mental processes are sensitive only to the local syntactic properties of mental representations; in particular, to the identity and arrangement of their constituents. The constituent structure of a mental representation, and whatever can be defined in terms of its constituent structure, is all that is "visible" to a Classical computational "machine" when it views an object in its computational domain. But, Fodor claims, not all the syntactic properties of a symbol are local in that sense; many types of relational syntactic properties are not: in particular, properties such as being the simplest of the available solutions to a computational problem are "global" not local (Fodor, 2005:26).

Fodor thinks that the appeal to the compositionality of mental representations to explain the productivity and systematicity of mental states has been, by and large, highly successful but that the attempt to reduce thought to computation has been much less so. In his later work, he took this scepticism further stating that, while computational nativism is "clearly the best theory of the cognitive mind that anyone has thought of so far (vastly bet-

ter than, for example, the associationistic empiricism that is the main alternative)", it is "nonetheless quite plausible that computational nativism is, in large part, not true" (Fodor, 2005). It should be noted that Fodor was not trying to convince anyone that computational models of cognition do not, or will not work but rather to point out that a (perhaps increasingly substantial) portion of the cognitive science community is already worried that "something has gone badly wrong with computational cognitive psychology and wonders what it might be and how it might be fixed." He contends that he was offering a diagnosis of a pattern of failures and to suggest "some ways in which intrinsic features of the 'Classical' computational model of mind might fail to capture crucial aspects of cognition even though Turing's account of computation underlies it, and there's a sense in which 'Turing machines can do anything'" (Fodor, 2005:25).

One of the major failures of computational cognitive psychology that Fodor identifies is its inability to provide a convincing account of abduction[23] which he suggests is more analogous to the (non-computational) cumulative accomplishments of the scientific community over millennia. A possible solution to the "globality-cum-complex-sensitivity" requirements of abduction and other similar cognitive processes might lie in the choice of cognitive *architecture*, but Fodor concludes that:

> The substantive problem is to understand, even to a first approximation, *what sort* of architecture cognitive science ought to switch to insofar as the goal is to accommodate abduction. As far as I know, however, nobody has the slightest idea. (Fodor 2000:47)

Whether or not abductive reasoning is amenable to algorithmic processing has yet to be definitively proved one way or the other. Being global and context-sensitive is a claim made for parallel distributive processing (PDP) architectures, and these "connectionist" models may be able to provide non-algorithmic accounts of global, context-sensitive cognitive processes.

3.3.2 **Connectionism**

The division between CTM and connectionist theories reflects the division of cognitive science into *computational cognitive science* and *neural cognitive science* (Rumelhart & McClelland, 1986). As previously discussed, the basis of *computational* cognitive science is

[23]In abductiive reasoning, perceptual input and whatever beliefs are stored in memory are used as data to solve the cognitive problem of finding and adopting new beliefs that are consistent with as many of one's prior epistemic commitments as possible (see Fodor 1998 for a discussion).

that cognitive mental processes are operations defined on syntactically-structured mental representations which are similar to sentences; that is, that mental representations are language-like. According to this view, the brain is a biological symbol-manipulator. *Neural* cognitive science, in contrast, rejects the mind/computer analogy which is the basis of *computational* cognitive science, holding instead that behaviour and cognitive capacities should be interpreted through the use of theoretical models. These models, which use the physical structure and processes of the nervous system as their inspiration, are "neural networks" composed of large sets of neuron-like units which interact locally through "synaptic-like" connections imitating the interactions found in the human brain. Each of the approximately 10 billion neurons in the human brain has approximately 10,000 connections (*synapses*) to other neurons. The input signals from these synapses is combined by the neuron in a manner that determines if and when to transmit a signal to other neurons. The input signals are modulated by the synapses before combination by the neuron, and by changes in modulation at each synapse the system "learns".

Andy Clark (1993) emphasizes three key features of connectionism:

> 1) *Superposition*: the ability to represent more than one thing with the same structure. The same neural network can be "trained" to recognize many different items by changing the weights of its connections;
>
> 2) *Context sensitivity:* because weights can encode multiple items, the "representation" of an item is automatically context-sensitive; and
>
> 3) *Representational change*: the ability both to create new representations and to acquire new representational capacities.

The symbols of the classical, symbol-system hypothesis do not change with the context in which they are located; context is, instead, expressed through relationships between symbols. For neural networks, on the other hand, representational context is embodied – context is internally expressed. While the models of the classical, symbol-system hypothesis have the ability to create new representations by combining symbolic expressions to create new symbolic expressions, this is performed through the combination of pre-existing, *internal* expressions, in contrast to a connectionist model which "learns" through training by an external environment.

For connectionism, the mind is neither a symbol manipulator nor a computational system and can be modelled as an artificial neural network consisting solely of quantitative processes. Through the use of complex networks of neurons (together with weights that measure/modulate the strength of connections between them), connectionism can manage without symbolic computation. That is to say, connectionism views the brain as a parallel distributed processing (PDP) device. Information processing in such networks begins with an INPUT pattern and ends with an OUTPUT pattern. This type of architecture is very well suited for perception-like tasks requiring object recognition. It is uncontentious that the brain is a vast collection of neurons, that humans can think, and that there have been some experimental successes[24] in modelling skills in such areas as facial recognition using PDP networks. There have also been some successes in other areas such as: a neural network trained to produce the past tenses of English verbs (James McLelland and David Rumelhart 1986); a network, called NETtalk (Sejnowski and Rosenberg 1987), which takes, as inputs, vector codings for seven-letter segments representing printed words, and produces vector codings for phonemes as outputs which can be then be used directly as input to a sound synthesizer thereby producing audible sounds. Another model (based on experiments by Gorman & Sejnowski, 1988) has been able to distinguish between sonar echoes returned from explosive mines, and the solar echoes returned from rocks of comparable sizes.

A main difference between the classical, symbol-system hypothesis position and that of connectionism is the nature of the internal representations that they each posit. The classical position is that internal representations have a semantic and syntactic structure analogous to that of a natural language: "classical theories – but not connectionist theories – postulate a 'language of thought'" (Fodor and Pylyshyn 1988:316). The classical theories posit internal representations that are similar to a sentence of a natural language in so far as they are composed of constituents and syntactic rules which, in combination, determine the meanings of the strings in which they appear. Further, these constituents are in some measure isomorphic to the lexical items of the natural-language sentences in which the thoughts are reported. To have the thought that "John loves the girl", then, is to be in some relation to a complex internal token the constituents of which have the context-independent meanings of 'John, 'loves', 'the', and 'girl'. Thus, the classical position proposes a conceptual-level compositional semantics:

[24] While these results are impressive, humans also respond to environmental features and taking these into account has proven difficult. For example, the vowel sound for the letter "a" in the English word "rain", may not be counted or recognized when voiced in isolation; what makes something an /a/ or an /e/ or an /ej/ is in part a matter of the entire linguistic surround (Churchland (1989).

> Classical theories ... take mental representations to have a combinatorial syntax and semantics, in which (a) there is a distinction between structurally-atomic and structurally-molecular representations; (b) structurally-molecular representations have syntactic constituents that are themselves either structurally-molecular or are structurally-atomic; and (c) the semantic content of a (molecular) representation is a function of the semantic contents of its syntactic parts, together with its constituent structure. (Fodor and Pylyshyn 1988:316)

It is because the constituents ('John', 'loves', 'the', and 'girl') make the same semantic contribution to the sentence "John loves the girl" as they make to the sentence "the girl loves John" that understanding the one sentence implies understanding the other. If, on the other hand, sentences of English were atomic, then there would be no reason why understanding "John loves the girl" would imply understanding "the girl loves John" any "more than understanding 'rabbit' implies understanding 'tree'" (Fodor and Pylyshyn 1988:331).

Both the classical and most connectionist theories are *representationalist*[25], but they differ in cognitive architecture where, for the classicists but not for the connectionists, mental representations are formed by a combinatorial syntax and mental processes that are sensitive to the combinatorial structure of mental representations (Fodor and Pylyshyn 1988:316). For most connectionist models, on the other hand, representations are distributed in a cognitive architecture that is implemented in a type of network. Unlike classical computers, a connectionist network's behaviour does not result from manipulating symbols in accordance with an algorithm. In distributed representations there is nothing analogous to the logical operations of taking an element from one complex representation and combining it with another (Fodor and Pylyshyn 1988) and, although some connectionist techniques have been developed which allow for structure-sensitive processing, such logical operations result in *functional* compositional structure rather than the *concatenative* compositional structure[26] of the classical cognitive architecture (van Gelder, 1990).

Some connectionist models appear to reject the representationalist position (e.g., Brooks 1991) but, as Clark & Eliasmith (2002) state, this is not so much a division between representational and non-representational models but is rather a division between a view of

[25] As opposed to *eliminativists* for whom the appropriate level of psychological theorizing is neurological and the semantic notion of representation is, therefore, not required.

[26] A complex representation is said to have a *concatenative* structure if the individual constitutive elements are embedded in it, and can be retrieved from it, without alteration; and to have a *functional* compositional structure if unaltered tokens of these elements are not embedded in the complex expression even though they are still usable or retrievable (Clark & Eliasmith 2002).

mental representations as memory-intensive and all-purpose forms of internal representation (as favoured by the standard computational theories) and that of internal representations as sparse and action-oriented forms which exploit stimuli from both the body and the external world to produce a response from which is built the representation itself. This connectionist position is of "just-in-time representation" (Ballard *et al.* 1997) based on the notion that "The world is its own best model" (Brooks 1991). Such connectionist models many still accept the idea of mental representations as internal content-bearing states but reject the view of them as rich and action-neutral forms of internal representation. While the latter may enable flexible processing, they, nevertheless, require additional computational effort to produce a behavioural response.

3.4 PART III: STRENGTHS AND WEAKNESSES

3.4.1 Modularity

Chomsky (1980) introduced the notion of domain-specific cognition by proposing that humans are born with domain-specific systems of knowledge. Among candidates for such systems are:

> *Knowledge of language*: neonates have the ability to recognize speech sounds at birth; to differentiate sounds of the mother tongue from unrelated languages as early as four days after birth (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984); and to learn languages amazingly quickly, despite the impoverished input;
>
> *Knowledge of physical objects*: infants, starting at around three months of age, react with surprise when two objects appear to occupy the same location; and, starting around four months of age, register surprise if one solid appears to pass through another; and
>
> *Knowledge of number*: Around six months of age, infants are able to discriminate between large sets of objects on the basis of numerosity (provided that the sets to be discriminated differ by a large enough ratio); and this capacity appears to develop prior to language acquisition and the ability to count symbolically (Xu & Spelke 1999).

Expanding on the notion of domain-specific cognition, Fodor (1983) presented a famous argument for modularity of mind. The basic idea is that the mind consists of a number of

functionally specialized mechanisms which evolved in response to recurring adaptive problems. His theory is not that all mental systems are modular (the "Massively Modular Mind Hypothesis") but rather that only peripheral subsystems are modular — high-level perception and cognitive systems are, on his view, non-modular. Fodor presents, therefore, a mental architecture that is composed of a number of modules that receive and process inputs received from the transducer systems[27], and produce outputs as representations that are then processed by a non-modular central system. In contrast, evolutionary psychologists, who have the goal of identifying and comprehending the species-specific architecture of the human mind and brain, tend to support the Massively Modular Mind Hypothesis (Tooby & Cosmides, 1992, Cosmides & Tooby 1997). Evolutionary psychology is:

> informed by the additional knowledge that evolutionary biology has to offer, in the expectation that understanding the process that designed the human mind will advance the discovery of its architecture (Cosmides, Tooby, & Barkow, 1992:3).

Evolutionary psychologists, such as Cosmides and Tooby, tend to be strongly adaptivist: they propose that the modular structure of the mind results from evolutionary pressures; and that it is even possible to identify which particular evolutionary pressures resulted in which type of module. Whether they accept the the Massively Modular Mind (MMM) hypothesis or a more limited version of modularity such as that proposed by Fodor, many (computational) cognitive scientists, psychologists, anthropologists, psycholinguists, and philosophers are now willing to accept that at least some cognitive abilities are domain-specific. Computational cognitive science tends toward the view that the mind is computational; that it is composed of distinct modules each of which specializes in the processing of distinct types of information, has specialized functions, and is informationally and functionally encapsulated (Chomsky, 1980; Fodor, 1983). In contrast to the *cognitivist* position of evolutionary psychologists and many others who support at least some form of the modular mind hypothesis, connectionists tend to be non-cognitivist and anti-innatist, holding instead that only a general capacity for learning is genetically inherited and that all cognitive capacities are the result of learning and experience.

Analysis of transducers and of Noam Chomsky's idea that the "poverty of the stimulus" argument are used to support the existence of an innate language "faculty" led Fodor to postulate the existence of perceptual modules. He used the examples of illusions, such as

[27] Transducers are peripheral sensory systems that convert input stimuli to electrochemical signals which output is in a computationally usable form and is lawfully dependent on their input stimuli (Fodor 1983:40).

the Müller-Lyer, to support his hypothesis that certain mental processes are informationally-encapsulated by which he meant that that they are not cognitively accessible; that they are not penetrable by our belief systems. Even if we are cognitively aware of the illusion we are unable to change our percept of it. He further postulated that these modules are domain-specific, mandatory, and fast. His hypothesis can be supported using adaptionist arguments for, as Immanuel Kant wrote, rational processing would be too slow and unreliable in dangerous situations that are very likely to arise in natural environments:

> In the natural constitution of an organic being—that is, of one contrived for the purpose of life—let us take it as a principle that in it no organ is to be found for any end unless it is also the most appropriate to that end and the best fitted for it. (Kant 1785/1956:395)

Kant's position is echoed in Hoffman's principle of perceptual categorization:

> **Principle of Satisficing Categories**: *Each perceptual category of an organism, to the extent that the category is shaped by natural selection, is a satisficing solution to adaptive problems.* (Hoffman 2009)

Hoffman argues that this principle aids in understanding the origins and purpose of perceptual categories. Satisficing categories are, Hoffman contends, those optimally suited to resolving common problems faced by all organisms, although specific solutions will vary since the actual form of the problems vary from niche to niche.

Fodor does not support the idea that the mind itself is modular but only that there are some, mostly perceptual, modules in the mind. Modules, on Fodor's account, decompose proximal stimuli in a series of stages which are *encapsulated*: they are not accessible to thought. He accepted Chomsky's idea of the innate language "faculty" but contended that this was an rationalist epistemological position (that is, that certain *knowledge* — certain sets of representations, or body of information — has to be innate) and as such could not be used to justify the postulation of cognitive modules, often termed "Darwinian/Chomsky" modules. Fodor describes human decision-making as *intentional psychology* and points out that belief formation and intentional action use are not characteristic of perceptual input systems which are domain-specific and do not involve inferential mechanisms. In order to compute the best hypothesis about the state of the world, the processing of cognitive systems is always subject to correction in the light of the organism's body of beliefs and perceptual input and thus these *belief-fixation* mechanisms cannot be domain-specific. The modular solution is suited to *informationally-bounded* problems, but is not suited to *inform-*

ationally-open decision problems. Fodor proposes, therefore, a mental architecture that is a hybrid of specialised cognitive modules and a central, domain-general processor.

Fodor's thesis was adopted by, in particular, evolutionary psychologists, such as Cosmides, Tooby, and many others, to explain many of their empirical findings. Cook and Mineka (1990), for example, postulated the existence of a module specific to the fear of snakes in rhesus monkeys whereby this "module" has all the properties of Fodorian perceptual modules, namely innate, fast, domain-specific, mandatory, and relatively informationally-encapsulated. Experimental ethologists have reported similar findings in other species such as the Tinbergen/Lorentz hypothesis (discussed in more detail in Chapter 5) that appears to demonstrate that certain species of birds have an innate fear of over-head predators: young, experimentally-naïve turkeys raised in batteries will panic if a particular shape is moved in one direction but show no response when the same shape is moved in the opposite direction — the shape moving in one direction looked similar to that of a hawk, whereas the same shape moved in the opposite direction looked similar to that of a goose. In addition, species other than humans are susceptible to certain illusions: bees also have been shown to react to the illusionary white rectangle appearing between appropriately positioned "pacmen", as shown in Illustration 11, suggesting that edge detection in both bees and humans may involve similar mechanisms that lead to the percept of illusory contours (van Hateren, Srinivasan, & Wait 1990). All such results seem to support the idea that the mind has considerably more modules than Fodor postulates, but these examples could still be deemed to be "perceptual' and peripheral and evidence of modularity at those levels cannot be used to justify postulating modularity at higher-levels.

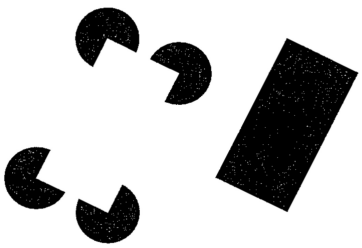

Illustration 11: Rectangle Illusion (© Isac & Reiss, I-Language, 2008)

Cosmides and Tooby, as well as other evolutionary psychologists, postulate the existence of *cognitive* modules and base these claims on empirical findings. The flagship example used to support their claim is the "Wason Selection Task" which Peter Wason developed to determine if humans use scientific hypothetico-deductive reasoning. Evolutionary psychologists use the task to demonstrate that subjects will perform better (that is, in conformance with the rules of the propositional calculus) if the task is worded in the form of a social contract. They claim that, if the task is so worded, a "cheater-detection" module (CDM) is triggered and that such a module would have been a beneficial adaptation in the

environment of evolutionary adaptation (EEA)[28]. Since the situations in which the CDM is being currently used is significantly different from that of the EEA, the issue of proper and actual domains has to be taken into account. Dan Sperber (2002), Murray Clarke (2004), and Stephen Pinker (1999), among others, claim that, if the gap between the actual and proper domains is sufficiently small, such a module will perform reliably.

Some supporters of the "Massively Modular Mind" hypothesis (MMM) contend that *most* but not all of the mind is composed of Darwinian/Chomsky modules, and some, like Sperber, claim that the mind is completely modular. Fodor, and others such as Kim Sterelny, have criticized both versions of MMM. Sterelny claims that MM could not explain the considerable flexibility and creativity of the human mind. Defenders of MMM, on the other hand, claim that, if central systems were non-modular, domain-general, and content neutral, then computation could become intractable: the possible domain of any reasoning process would include all epistemic commitments thereby giving rise to the "frame problem"[29]. In support of MMM, Cosmides and Tooby state that:

> If there is an adaptive problem that can be solved either by a domain general or a domain specific mechanism, which design is the better engineering solution and, therefore, the design more likely to have been naturally selected for? (Cosmides & Tooby 1994:89)

The idea being that, functionally specialized mechanisms can be quickly, efficiently, and reliably fine-tuned for processing specific types of information, but general-purpose mechanisms are amenable to fine-tuning for a single task as long as there is little or no impact on its performance of other tasks. The claim is that, if two adaptive problems have different or incompatible solutions, then two functionally-specialized mechanisms are likely to outperform one general-purpose mechanism, and therefore natural selection will tend to favour functionally-specialized mechanisms.

Almost all supporters of MMM accept the Computational Theory of Mind (CTM) — the thesis that reasoning is local computation on syntactically-structured symbols. Fodor contends that, because such computation is local, CTM cannot account for global reasoning such as abduction (inference to the best explanation). Supporters of MMM claim that the modularity thesis can account for abduction since computation within each module is local.

[28] The EEA is usually assumed to be the latter part of Pleistocene when humans were hunter-gatherers.
[29] The "frame problem" according to Dennett, concerns how "a cognitive creature … with many beliefs about the world" can update those beliefs when performing an action so that the beliefs remain "roughly faithful to the world" (Dennett 1978:125). Fodor defines it as 'the problem of putting a "frame" around the set of beliefs that may need to be revised in light of specified newly available information' (Fodor 1983:112-3)

Fodor's argument that MMM cannot account for abduction is based on the combination of the claim that the input to an abductive process is the whole of one's epistemic commitment and the claim that modules are informationally-encapsulated. This latter claim is contested by supporters of MMM because their theory does not entail that all modules be encapsulated; in fact, their theory is completely consistent with the existence of non-encapsulated modular processors (Darwinian modules) as well as Chomsky modules. This position prompted another criticism from Fodor, namely that there would have to be some mechanism whereby input to these modules is limited to only relevant information in order to avoid the frame problem (Fodor 2001). The same criticism could, however, be levelled at Fodor's position: that there would have to be a mechanism available for determining which information is relevant to enable domain-general, non-modular, content-neutral systems to process any particular global problem without information-overload and without compromising the existing set of beliefs. Hence both the massively modular hypothesis favoured by evolutionary psychologists and the limited modular hypothesis favoured by Fodor, for instance, suffer from a frame problem.

Gigerenzer and Selten (2001) have offered a possible solution to the frame problem that would still be compatible with the computational theory of cognition. They point out that human rationality is not optimal and that humans often reason using "fast and frugal" heuristics. The main problem with the heuristics solution to the problem of abduction, according to Fodor, is that it simply "won't do." The reasons that it "won't do" are that: 1) to be reliable, abduction may require information from anywhere in the network of prior epistemic commitments; 2) to be feasible, abduction requires that, in practice, only a small subset of even relevant background beliefs be actually considered; and, even more crucially, 3) abductive inference itself is often involved in choosing which problem-solving heuristic to use. Postulation of global cognitive processes is not an option within the Turing framework and "bona fide abductive inferences are nonlocal, hence noncomputational by definition" (Fodor 2000:44) .

In addition, defenders of MMM may be taking insufficient account of how humans actually reason. First, the Wason Selection Task can be interpreted in many ways without having to postulate the existence of a Cheater Detection Module and, secondly, that inference to the best explanation, the way it is conducted by humans (including scientists) may not have to entail global processing. In the tests performed by Cosmides and Tooby (1987) the predicted selection of "benefit taken" and "cost not paid" corresponds to the truth conditions

of conditionals of the propositional calculus leading them to conjecture that by wording the Wason Selection Task in the form of a social contract, that is, in a more socially relevant format, subjects are more likely to select the logically correct answer of P and NOT-Q. Gerd Gigerenzer and Klaus Hug (1992), on the other hand, claimed that the key to these tests is that cheater detection is pragmatic and depends on perspective, whereas logic does not. To test their claim, they used a rule in the form of a "switched social contract". In this case, the logically correct answer would be P and NOT-Q, as previously, but detection of a rule violation would require selecting the NOT-P and Q cards thereby violating the rules of propositional calculus. The "illogical" choice would detect cheaters, whereas the formally logical choice would not. The result appears to be that we do not use logical reasoning when attempting to detect violation of a social contract rule. In addition, certain social contracts, such as "if an employee works at the weekend, he takes a day off the following week" has two possible answers depending whether the person being asked is an employee or an employer (Gigerenzer & Hug 1992). Additionally, empirical evidence may contradict Cosmides and Tooby's claim that the purported Cheater Detection Module evolved during the hunter-gatherer phase of human existence: studies of capuchin monkeys, for example, show that emotional reactions to perceived unfairness evolved considerably earlier (Brosnan & de Waal 2002)[30]. These reactions support an early evolutionary origin of inequity aversion. During the evolution of cooperative behaviour it may have become critical for individuals to compare their own efforts and rewards with those of others. Negative reactions may occur when expectations are violated. In fact, the cheater detection effects may only be the result of activation of part of the amygdalae during reasoning: that is, that the subject has to have an emotional investment even when reasoning counterfactually and a relentlessly logical person would not be able to detect cheaters reliably.

One of the characteristics of Fodorian modules is that they are realized in dedicated neural architecture and there are empirical studies of the effect of focal brain lesions that lend some support to the claim that mental faculties are localized. Localization is also supported by some neuroimaging studies which attempt to identify the areas of the brain that are active when subjects perform specific mental tasks. Such studies also show, however, that many fundamental mental tasks appear to involve large areas of the cortex suggesting the existence of large-scale networks rather than small localized regions. There is, for ex-

[30] Brosnan & de Waal (2001) report that monkeys refused to participate if they witnessed a conspecific obtain a more attractive reward for equal effort. The reactions was even greater if another monkey received such a reward without any effort at all. Many of the "working" monkeys became so enraged that they often threw their rewards at the researchers.

ample, disagreement about the precise location of even Broca's area which has been highly studied and is purported to be the centre of language production (Poeppel, 1996). Every lobe of the brain seems to be involved in some measure in language production (Pulvermüller, 1999). These latter results lend credence to the connectionist claim that mental processes are distributed and massively parallel. In addition, the brain structure includes many different types of neurons and there appears to be a correlation between a neuron's type and its specific function with differently-shaped neurons predominating in different areas of the brain. As J. R. Anderson points out, distributed representations:

> assume that a memory record does not reside in any single neural element but rather is represented by a pattern of activation over a set of neural elements (Anderson 1995:224-5)

That said, there is no reason to suppose that being modular and domain-specific, and being distributed and massively parallel are mutually exclusive. It is possible that a domain-specific system could be multi-modal and composed of sub-systems.

3.4.2 **Propositional Attitudes**

Failure of substitution of co-referring expressions in propositional attitude statements creates a problem for any semantic theory. Just what constitutes a propositional attitude is a problem for any cognitive architecture. Folk psychology (or common-sense psychology) concerning the prediction and explanation of our own and each other's behaviour is, as Fodor (1987) noted, both successful and indispensable. It provides a common-sense conceptual framework for mental phenomena which provides:

> a simple and unifying organization to most of the major topics in the philosophy of mind, including the explanation and prediction of behavior, the semantics of mental predicates, action theory, the other-minds problem, the intentionality of mental states, the nature of introspection, and the mind-body problem. Any view that can pull this lot together deserves careful consideration. (Churchland 1981:68)

The content of a person's beliefs, desires, fears, doubts, and other such states, is a reliable indicator of what a person is likely to do. In order to make sense of each other's behaviour, we ascribe these states to them and generalize accordingly. By so doing, we commit ourselves to the basic truth of folk psychology and to an ontology that includes these states. Fodor writes,

> It does seem relatively clear what we want from a philosophical account of the propositional attitudes. At a minimum, we want to explain how it is that propositional attitudes have semantic properties, and we want an explanation of the opacity of propositional attitudes; all this within a framework sufficiently Realistic to tolerate the ascription of causal roles to beliefs and desires. (Fodor 1981:18)

Propositional Attitudes (PAs) refer to folk psychological attitudes (e.g., belief, desire, fear) taken toward a proposition (e.g., "I believe that the Sun revolves around the Earth"); a PA is a mental state that expresses a relationship between an individual (the attitude holder) and a proposition; PAs imply that a person can have different mental postures toward the same proposition (e.g., believing that P; hoping that P); and PAs are "about" or refer to something – they imply *intentionality*.

For Fodor, and many other *intentional realists* (e.g., Dretske 1988), propositional attitudes can be viewed as computational relations to mental representations; they are semantically interpretable and have a causal role. Intentional states, such as beliefs and desires, are relations between a thinker and symbolic representations of the content of the states; to have a thought is to stand in a certain functional relation to a token mental representation composed of concepts; and to think is to perform content-respecting computational operations, à la Turing, over such representations. Propositional attitudes are susceptible of semantic evaluation (of evaluation in such terms as satisfied/unsatisfied, true/false, and so on), and they have intentional content through which they influence behaviour, belief fixation, and so on. They are also functionally discrete in so far as they can play causal roles individually.

As regards the role of propositional attitudes and belief-fixation, there are two main theories according to which: 1) Perceptual aspects are privileged and the fixation of beliefs is dependent on perception, that is, on the relation with the external world; and 2) The inferential relation between old and new beliefs is privileged, such that the fixation of beliefs is substantially an internal process. Fodor (1983) rejects the latter, *interpretationalist* theory citing as evidence the persistence of perceptual illusions (e.g., the Müller-Lyer illusion) that clearly demonstrate the impenetrability of perception to background knowledge. Fodor is here emphasizing the distinction between observation, which is, for the most part, independent of a subject's background beliefs, and the perceptual fixation of beliefs, which involves the holistic relation between the subject's beliefs.

Direct perception of distal objects is impossible and hence it is necessary, Fodor contends, to adopt a necessarily mediated and indirect inferential theory of perception. At the base of these inferential processes is a computational mechanism using representations encoded by the transducers (Fodor, 1983:42). *Input systems*[31] constitute, for Fodor, a natural kind, which he defines as "a class of phenomena that have many scientifically interesting properties over and above whatever properties define the class" (Fodor, 1983:46), and *input processing* makes non-demonstrative inferences from sensory data to produce hypotheses about the objects in the external world. These hypotheses are transmitted to central systems for the purpose of the fixation of beliefs (or any other propositional attitude). That belief-fixation is a central system process is something Fodor insists upon:

> *the operation of the input systems should not be identified with the fixation of belief.* What we *believe* depends on the evaluation of how things look, or are said to be, *in light of background information* about (inter alia) how good the seeing is or how trustworthy the source. (Fodor 1983:46, italics in original)

and, as discussed in the preceding section on modularity, perceptual modules for Fodor are domain-specific, mandatory, and informationally-encapsulated, whereas central systems are non-modular, domain-general, and content neutral.

Now, Fodor contends that only peripheral modules are computational. Central systems are not modular, on his account, but are responsible for belief-fixation. These central systems are not modular because, he contends, belief-fixation is *isotropic* and *Quinean*. That is to say, belief-fixation is isotropic in so far as it involves the epistemic interconnectedness of beliefs in the same way that in confirming a scientific hypothesis:

> everything that the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect" (Fodor, 1983, p. 105).i

By stating that confirmation is also 'Quinean', Fodor means that:

> [T]he degree of confirmation assigned to any given hypothesis is sensitive to properties of the entire belief system … [S]implicity, plausibility, and conservatism are properties that theories have in virtue of their relation to the whole structure of scientific beliefs *taken collectively*. A measure of conservatism or simplicity would be a metric over *global* properties of belief systems (Fodor 1983:107–108).

[31] An *input system* is a computational mechanism that "presents the world to thought" by processing the outputs of sensory transducers (Fodor 1983:40).

The epistemic interconnectedness of beliefs required for belief-fixation and the requirement for confirmation to be sensitive to properties of an entire belief system mean that isotropic and Quinean processes cannot be performed by informationally-encapsulated systems. Unfortunately, "[t]he more global (e.g., the more isotropic) a cognitive process is, the less anybody understands it" (Fodor 1983:107). Despite this, Fodor goes on to state that scientific, isotropic confirmation is able to provide a model for the non-modular fixation of beliefs; and that Quinean confirmation metrics (simplicity, plausibility, and the like), being global properties of belief systems, constrain confirmation non-arbitrarily[32] (Fodor 1983:111).

Fodor's view of the brain is that of a combination of a "hard-wired", stable neural architecture (associated perception and language) and a central system (associated with thought) composed of Quinean/isotropic systems with connections between them that are *unstable* and *instantaneous* (Fodor 1983:118). The distinction between input analysis and belief-fixation is a division between "relatively local and relatively global computations" (Fodor 1983:126). What is still needed is to determine the structure of an entire belief system such that it enables individual instances of belief fixation and this is problematic given that we have "no computational formalisms that show us how to do this, and we have no idea how such formalisms might be developed" (Fodor 1983:129). Perhaps neuroscience will be able to develop such formalisms. This remains to be seen. That we do not yet have the formalisms that Fodor seeks is, of course, not an argument against the view of the mind as computational.

Proponents of the Computational Theory of Mind (CTM) claim that it provides an account of the semantic and intentional properties of mental states and also a framework for psychological explanation such that intentional state ascriptions can figure in such explanations. As Fodor notes:

> To a first approximation, symbols and mental states both have representational content. And nothing else does that belongs to the causal order: not rocks, or worms or trees or spiral nebulae. (Fodor 1987: xi)

and so:

> It would, therefore, be no great surprise if the theory of mind and the theory of symbols were some day to converge. (*ibid*)

[32] Quinean metrics can be used to support any hypothesis whatsoever if relevance of beliefs is non-isotropic, if the subset of beliefs is *arbitrarily delimited* (Fodor 1983:111).

Given that mental representations have semantic properties and the semantic properties of propositional attitudes are inherited from those of mental representations, it would, therefore seem, to Fodor at least, that "the semantic properties of the formulae of natural languages are inherited from those of the propositional attitudes that they are used to express" (Fodor 1981:31). Whatever computational formalisms might be involved in individual instances of belief fixation, propositional attitudes themselves are syntactically structured like sentences (Fodor 1998b; Pinker 1999) and thinking is done in a Language of Thought (Fodor 1975, 2008). The Language of Thought (LOT) hypothesis is discussed in paragraph 3.4.3 below.

Connectionism tends to a view of the mind as non-modular: information in neural networks is represented by distributed patterns of activation which are transformed into other activation patterns through modifications of the connection weights of the communication links between the network's units. Neural network models are not in general divided into any type of module other than what is required to distinguish between input units, output units, and layers of hidden (or internal) units. Many connectionists, such as the Churchlands, have taken a different tack from Fodor on belief-fixation, namely that there is increasing evidence supporting the view that very little human knowledge has to do with propositional attitudes and indeed it is possible to find examples of computational systems in which there are many symbol manipulations that have no obvious description at the level of propositional attitudes — for example, we may use inference rules like modus ponens, the disjunctive syllogism, and so on, to reason, without necessarily representing the rules explicitly (see also examples discussed by Fodor 1987:23–6; Dennett 1998:107). A more extreme view is that of Ramsey, Stich, and Garon (1990) who have argued that:

> If connectionist hypotheses of the sort we will sketch turn out to be right, so too will eliminativism about propositional attitudes. (Ramsey, Stich, & Garon 1990:500)

What they mean here is not just that reference to the propositional attitudes will be eliminated from folk, or common-sense, psychology, but that propositional attitudes will turn out not to exist at all. In other words, they contend that common-sense psychology and connectionism are incompatible. The justification they provide is that, because information is encoded holistically in distributed connectionist models, connectionism cannot

encode the necessary functionally-discrete states[33]. The connectionist models they studied are cognitive-level, superpositional[34] networks which encode information in a widely distributed and sub-symbolic manner. There is nothing in these models, they contend, with which propositional attitudes can plausibly be identified[35] (Ramsey, Stich, & Garon 1990:520) and hence, since distributed, superpositional connectionist networks will be the future paradigm for cognitive science (or so they appear to believe) propositional attitudes do not exist. The arguments that Ramsey, Stich, and Garon present are based on their claim that propositional attitudes (of folk psychology) are defined by a cluster of three features that they name "propositional modularity": propositional attitudes, on their view, are *semantically interpretable*; have a *causal role*; and are *functionally discrete* (Ramsey, Stich, & Garon 1990:504). Their position on the non-existence of propositional attitudes thus rests on three claims: that propositional attitudes are propositionally modular; such propositional modularity cannot be implemented in a distributed, superpositional connectionist network; and distributed, superpositional connectionist networks are the future of cognitive science.

Some philosophers (e.g., Bogdan 1993; Clark 1995) have questioned whether propositional attitudes are modular at all and some claim that distributed, superpositional connectionist networks can, in any case, be propositionally modular in the way that Ramsey, Stich, and Garon claim they are (e.g., Clark 1995; Smolensky 1995). Clark, in particular, suggests three ways of overcoming their eliminativist arguments:

> We might just deny that the folk care about propositional modularity. . . . Or we might try to show that propositional modularity is safe whatever turns up in the head. . . . Finally, we might argue that distributed, sub-symbolic, superpositional connectionist models are actually more structured than RS&G think, and hence visibly compatible with the requirements of propositional modularity. (Clark 1995:345)

Ramsey, Stich, and Garon claim that being *functionally discrete* is a necessary condition for being a propositional attitude. It should be noted, however, that there is more than one type of functional discreteness: *dispositional and occurrent* are prime examples. A dispositional belief is one that is about something that is likely to occur and endures (is stored in memory); an occurrent belief is one that is about something that is occurring and is

[33] In Classical Turing architectures, local syntax determines the causal powers of a mental representation; and representations that are syntactically distinct are hence type-distinct which means that local syntactic properties are essential.
[34] Superpositionality, the ability to represent two different items using the same structure, is discussed in § 3.3.2 .
[35] RS&H claim that these models encode the "intentional object" of a propositional attitude but not the attitude itself. Fodor uses the "dodge" invented by Stephen Schiffer: "for each episode of believing that P, there is a corresponding episode of having 'in one's belief box', a mental representation which means that P" (Fodor 1998a:8). For RS&H then, there is no way for the connectionist models they studied to encode "belief boxes".

fleeting. Both dispositional and occurrent beliefs can be implemented by connectionist models including the specific, superpositional types discussed by Ramsey, Stich, and Garon (see, for instance, Clark 1995; Smolensky 1995). For any supporter of the state space semantics approach (see, for instance, Churchland 2007), o*ccurrent* representations are points, regions, or trajectories in an activation-vector space, and longer-term representations (e.g., *dispositional* representations) are points in the weight space.

3.4.3 **Language of Thought**

Underlying discussions of propositional attitudes is the claim that propositions can be viewed as terms in a formal representational language. Fodor (1975) claims that, because direct perception of distal objects is impossible, it is necessary to adopt an inferential theory of perception. The transducers provide the required computational mechanisms and these mechanisms produce a representation of the distal stimulus on the basis of the properties of the proximal stimulus. But, as Fodor states, "representation presupposes a medium of representation, and there is no symbolization without symbols. In particular, there is no internal representation without an internal language" (Fodor 1975:55). That the language of thought has to be *internal* rather *public* is because we would otherwise have to deny the ability to think to non-verbal animals and pre-linguistic infants and, as Fodor says:

> there are homogeneities between the mental capacities of infraverbal organisms and those of fluent human beings which, so far as anyone knows, are inexplicable except on the assumption that infraverbal psychology is relevantly homogeneous with our psychology (Fodor 1975:57)

Fodor notes that both humans and infraverbal organisms typically find disjunctive concepts (e.g., 'red or blue') difficult to master (see Fodor, Garrett, & Brill 1975) and says that this can be explained on the assumption that both humans and nonverbal organisms employ relevantly similar representational systems (Fodor 1975:57).

Pylyshyn used Fodor's LOT hypothesis to extend his claim that knowledge is encoded by physically instantiated symbol structures by also claiming that mental activity is basically manipulation of *sentence-like symbolic expressions* and that human thought is the manipulation of "sentence analogues" in an internal mental language called "mentalese" or the "language of thought" (Pylyshyn 1984:194). Some thoughts are intrinsically connected with other thoughts. If a normal cognitive agent lacked some thoughts, he would also lack certain other thoughts. This intrinsic systematicity can be explained by a semantically and syn-

tactically combinatorial language of thought (Fodor & Pylyshyn, 1988). The classical treatment of mental processes rests on two main ideas:

> 1) It is possible to construct languages such that certain features of the syntactic structures of formulae correspond systematically to certain of their semantic features; and
>
> 2) It is possible to devise machines whose function is the transformation of symbols, and whose operations are sensitive to the syntactical structure of the symbols on which they operate

and

> Such a machine would be just what's required for a mechanical model of the semantic coherence of thought; correspondingly, the idea that the brain is such a machine is the foundational hypothesis of classical cognitive science. (Fodor & Pylyshyn 1988)

Because semantic symbols and syntactic rules can easily be represented in a classical, von Neumann computer architecture, Language of Thought (LOT) is usually represented as implemented by "classical AI" (also known as GOFAI: "good, old-fashioned AI").

Although connectionism does not require symbols, representations can be symbolic. Connectionsts (e.g., Churchland 1989, Smolensky 1988) hold that there is a level of representation which lies beneath that of the sentential or propositional attitudes; and that there is a learning dynamic that operates primarily on sub-linguistic factors. According to this view, depicting knowledge as an immense set of individually stored sentences cannot explain how is it is possible to retrieve from the millions of sentences stored, the small number relevant to a current predictive or explanatory problem; nor can it explain how such retrieval can be accomplished in fractions of a second. Fodor (2000) himself has pointed out how one of the major failures of computational cognitive psychology lies in its inability to provide a convincing account of abduction.

Many connectionists propose a *state-space semantics* (SSS) approach which views concepts as "functionally salient *points, regions,* or *trajectories* in various neuronal activation spaces" in contrast to the LOT approach which views concepts as "functionally salient *wordlike* elements in a *languagelike* system of internal representations" (Churchland 2007:126). Fodor & Lepore (1992) present a theory of meaning for which the basic LOT concepts are unstructured atoms that are mutually independent, and where each atom has its content through a causal-cum-nomological relation with some aspect of the external

world. This theory of meaning contrasts starkly with Paul Churchland's SSS for which conceptual content is a *portrayal* of the world that has no automatic referential connection to the external world; even "primitive" concepts have a complex structure; and a creature's set of beliefs ("cognitive portrayals" of the world) are constitutive of meaning (Churchland 2007:135). On this view, a concept "encompasses a substantial *range* of distinct but closely related cases, in that the mature network will have generated, in the course of its learning the concept, a proprietary *volume* within its activation space" (Churchland 2007:144). A framework of concepts constitutes a portrayal of some section or aspect of the external world — this portrayal is a complex physical structure composed of internal relations that mirror the family of relations comprising the target external domain[36]. In other words, according to this state space semantics account, "there exists a relation-preserving mapping from the external domain to the acquired structure of the relevant neuronal-activation space" (Churchland 2007:159).

3.4.4 **Natural-language and Cognitive Architecture**

Chomsky (1957) claimed that humans are born with a strong biological predisposition for language and emphasized that the language input that a child receives under-specifies his/her ultimate knowledge of language (the "Poverty of the Stimulus argument"). He further claimed that humans have an innate "language faculty" which is configured for language acquisition:

> An engineer faced with the problem of designing a device for meeting the given input-output conditions would naturally conclude that the basic properties of the output are a consequence of the design of the device. Nor is there any plausible alternative to this assumption as far as I can see. (Chomsky, 1968)

Fodor agrees with Chomsky that there must be something innate to account for humans' natural language ability but his analysis of the conditions required for learning a natural language lead him to claim a relationship between natural languages and the language of thought: "not all the languages one knows are languages one has learned, and that at least one of the languages which one knows without learning is as powerful as any language that one can ever learn" (Fodor 1975:82). In other words, the ability to learn a natural language

[36] The "target external domain" is whatever gave rise to the peripheral input to the neuronal transducers, the output of which is used by the brain to build a model of the world. An example provided by Churchland (2007:203) is that of the mapping of the "peculiar and well-defined three-dimensional structure of the human phenomenological space" (the inner portrayal) with the "objective space of possible electromagnetic reflectance profiles displayed by material objects" (the target external domain) – that is, how we experience the colour of an object versus the subset of electromagnetic radiation reflected by the external object.

in the first place is dependent upon having an innate language of thought. Fodor does not claim that having an LOT is a sufficient condition for being able to learn a natural language but does claim that it is necessary. That some other non-human species might have some type of LOT appears reasonable given the similarity of brain structures.

Fodor integrated the Chomskyan idea of the language facility as a cognitive module with his own computational theory of mind. According to the Fodor-Chomsky model, the language module is complete and inviolate in the brain; one that performs all and only the functions of language without external influence. According to Chomsky, it is not possible to learn language through a general-purpose mechanism and, although language experience is essential for language acquisition, linguistic experience only initializes the "language organ". Much of the knowledge which language requires is innate; moreover, it is encapsulated in so far as one's beliefs about the world play no part in the ability to construct grammatical sentences in one's natural language—innate "Universal Grammar" is not open to introspection.

Fodor holds that language is an "eccentric stimulus domain." A stimulus domain is considered to be "eccentric" if there is a wide experience/data gap; if it is difficult to develop the right concepts to describe the data from experience alone. Language appears to be a paradigm example of an eccentric stimulus domain: there is nothing in the spectrograph of an utterance that marks out the salient features needed for linguistic interpretation. Indeed, how we extract out from all the background noises just those which constitute an utterance and how we parse the spectrograph of the utterance into a phonological representation is not at all obvious. Language is certainly cognitively demanding and its usage involves complex multi-tasking such as keeping track of what one has said; what others have said; determining other's attention, understanding, or dissension, and so on.

There have been attempts to apply connectionist principles to the area of language acquisition. Models have been designed to test grammatical tasks such as the production of English past tense verbs (Rumelhart & McClelland 1986), or recognition of gender of French nouns (Sokolik & Smith 1992). Parallel Distributed Processing may be able to provide a more plausible explanation than rule-driven language systems of why children learn a second language more readily than adults when neurological constraints on language learning and the parallel processing which characterizes brain function are taken into account (Sokolik 1990).

3.4.5 **Levels of Cognitive Architecture**

One of the difficulties inherent in performing a comparison of descriptions of cognitive architectures is that they are not always determined at the same cognitive level. Notably, Fodor & Pylyshyn (1988) accused connectionists of confusing the level of psychological explanation with the level of implementation. Cognitive architecture, they claim:

> consists of the set of basic operations, resources, functions, principles, etc. ... whose domain and range are the *representational states* of the organism. (Fodor & Pylyshyn 1988:10)

On this account, the operations, resources, and so on, consist of rule-governed symbol manipulation; of processes that are sensitive to the structure of the symbols that are conceptual-level representations. The level of cognitive analysis adopted by the sub-symbolic paradigm, favoured by connectionists, on the other hand, is lower than the level traditionally adopted by the symbolic paradigm. As Paul Smolensky (1988:§1.3) explains, the preferred level of the symbolic paradigm is the *conceptual level*; and the preferred level of the sub-symbolic paradigm is the *sub-conceptual level* which lies between the neural and conceptual levels. Smolensky notes that there are both semantic and syntactic distinctions between the symbolic and sub-symbolic paradigms:

> For the *Symbolic Paradigm*: entities are typically represented by symbols; symbols are operated upon by symbol manipulation; and operations consist of a single discrete operation;
>
> whereas
>
> For the *Sub-symbolic Paradigm*: entities are represented by a large number of sub-symbols; sub-symbols participate in computation; and operations often consist of a large number of finer-grained operations.

Differences between the two levels can be illustrated by comparing how novices and experts learn. (This is discussed in more detail in the section on learning below.) Novices are conscious of sequentially following linguistically-formalized rules. The intuitive knowledge of a novice can thus be easily modelled as a symbol-system manipulator (Smolensky 1988:§2.1) given that, for the symbolic paradigm, programs:

- Consist of linguistically formalized rules that are sequentially interpreted;

- Are composed of elements (symbols) referring to essentially the same concepts as the ones used to consciously conceptualize the task domain; and

- Have a syntax and semantics comparable to those running on the conscious rule interpreter.

This is not, however, the case for the intuitive knowledge of an expert. A study on expertise conducted by a Dutch psychologist, Adriaan de Groot, during the 1940's concluded that the advantage that master chess players have over weaker players is that their view of a chess game is organized into thousands of *chunks,* where a *chunk* is a familiar stimulus grouping which is stored in memory as a single unit. The analysis of the game into configurations of several chess pieces forms, according to de Groot, the basis for selecting the appropriate moves. The conclusion by de Groot came from a series of memory studies that compared experts' and novices' ability to recall pieces on a chessboard as it might appear 20 moves into a game. Experts were able to replace about 90% of the pieces correctly, compared to weaker players' 40%. A later experiment by de Groot showed that, when pieces were placed *randomly* on the board, thereby eliminating familiar groups or chunks, the master players scored no better than weaker players in reproducing the layout of the pieces (de Groot 1965). This can be explained by the requirement of the rule interpretation process to maintain the retrieved, linguistically-encoded rule in memory during the interpretation process and that, in consequence, it is essential that the activity pattern which represents the rule be stable for a relatively long time. On the other hand, no correspondingly stable pattern need be formed once the connections to perform the task directly have been developed. This provides a natural explanation of the loss of conscious phenomenology with expertise. The fact that the rule interpretation process is sequential is not a result of the cognitive architecture; but is, rather, the result of our limitation to following only one verbal instruction at a time. “Even if the memorized rules are assumed to be linguistically-encoded, there is no commitment as to the encoded form (phonological, orthographic, semantic, and so on)” (Smolensky 1988:§6.1). This observation will form an important part of the arguments for a hybrid cognitive architecture in Chapter 5.

Smolensky (1988:§2.2) claims that expert knowledge is not useful for cultural purposes because an expert’s individual knowledge does not posses the properties of cultural knowledge. He states that the method of formulating knowledge and drawing conclusions have the properties of:

- *Public Access*: the knowledge is accessibly to many people;

- *Reliability*: different people (or the same person at different times) can reliably check whether conclusions have been validly reached; and

- *Formality, bootstrapping, universality*: the inferential operations require very little experience with the domain to which the symbols refer.

But expert knowledge is not publicly accessible or completely reliable, and is completely dependent on ample experience.

3.4.6 **Learning**

H. L. Dreyfus and S. E. Dreyfus (1986) produced a five-stage typology of developing expertise and, according to this model, skills progress from the novice stage, with a rigid adherence to rules provided by an instructor; through three more levels of increasing competency and proficiency; to the expert stage at which there is no longer any reliance on rules, guidelines or maxims. The rate of play for expert chess players, they note, is 5 to 10 seconds per move and can be even faster without significantly degrading performance. Given the speed at which they are performing, expert chess players "must depend almost entirely on intuition and hardly at all on analysis and comparison of alternatives" (Dreyfus & Dreyfus 1986); that is, experts do not *consciously and sequentially follow rules*. This observation is not new: in the Platonic dialogue *Euthyphro*, Socrates attempts to discover the rules for recognizing the characteristics of piety by asking an expert on the subject. All the expert, Euthyphro, is able to provide, however, are examples of piety. He knows how to judge acts as pious or impious; but is unable to state the rules used for generating such judgements. Even though many philosophers and knowledge engineers accept the view that expertise is based on the application of sophisticated heuristics to a large volume of facts:

> [A]n expert's knowledge is often ill-specified or incomplete because the expert himself doesn't always know exactly what it is he knows about his domain. (from Edward Feigenbaum's book *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World*, as quoted in Dreyfus & Dreyfus 1986)

Expert judgement, at least, is not naturally modelled as the *sequential interpretation of a linguistically formalized procedure* and thus we can safely say that not all mental processes are necessarily algorithmic, *pace* Barbel Inhelder and Jean Piaget who maintained that "reasoning is nothing more than the propositional calculus itself" (Inhelder & Piaget 1958:305). Propositional calculus is concerned with the truth-value of the propositions, but not their content. In the preceding discussions of algorithms and rule-based procedures

(both those in this paragraph and in paragraph 3.3.1 on the Computational Theory of Mind above), there is no concern for the content of the propositions — computation is a causal relation between symbols that respects symbolic content. Algorithms, effective procedures, recipes, and such like, have, as John Searle remarked about computers, "a syntax but no semantics" (Searle 1980).

In the symbolic paradigm, both conscious rule application and intuition (i.e., both conscious and unconscious rule interpretation) are described at the conceptual level, but, in the sub-symbolic paradigm, conscious rule application can be formalized in the conceptual level but intuition must be formalized at the sub-conceptual level (Smolensky 1988:§6). The intuitive processor is presumably responsible for all of animal behaviour: perception, practised motor behaviour, fluent linguistic behaviour, intuition in problem-solving and game-playing — in short, practically all skilled behaviour. The transference of responsibility from the conscious rule interpreter to the intuitive processor during the acquisition of skill has been well studied by cognitive scientists (Anderson *et al*., 1981). In short, adherents of theories based on the symbolic paradigm and those of theories based on the connectionist, or sub-symbolic paradigm may often be talking at cross purposes.

3.4.7 **Comparison of Cognitive Architectures**

There are many crucial differences between the classic, symbol-manipulation, computational cognitive architecture and that of the connectionist model. There are also some aspects that both architectures have in common. The following table shows comparisons of some of the main aspects of the classical, symbolic and the connectionist views of cognitive architecture.

Cognitive Architectures — Comparison		
Cognitive Phenomena	**Classical Model**	**Connectionist Model**
Level of explanation	Symbolic	Sub-symbolic
Mental representation	Symbol structure with combinatorial syntax and semantics; Local	Activation patterns; Can be symbol (node) but not necessarily; Usually distributed
Concept	Linguistically-encoded — word-like in a language-like system	Points, regions, or trajectories in a neuronal activation-vector space — "prototype"
Conceptual content	External, real-world entity	Internal
"Fuzzy" and graded concepts	Difficult to implement	Easily implemented
Meaning	Reference or denotation; Beliefs irrelevant	"Portrayal of the world" with no automatic referential connection

Cognitive Architectures — Comparison		
Cognitive Phenomena	**Classical Model**	**Connectionist Model**
		to the external world; beliefs constitutive of meaning
Language of Thought[A]	Concepts are symbols in an LOT	Individual nodes (symbols) can be components of an LOT, but not typically. State-space semantics (Churchland 2007)
Modularity	Mind massively modular (see Cosmides & Tooby; Sperber, *et al*); or Modular perceptual peripheral systems with domain-neutral central system (see Fodor, Pylyshyn, *et al*)	mental processes distributed and massively parallel but correlation between a neuron's *type* and its specific function with differently-shaped neurons predominating in different areas of the brain
(Conscious) Behaviour	Result of a practical syllogism through process of symbol manipulation	Output of a neural net in response to a specific set of inputs
Categorical framework	Referential	Theory of how the world is structured
Fundamental theories	Propositional	Non-propositional
Processing	Sentential	Massively-parallel
Basic mode of operation	Rule-governed symbol manipulation	Vector transformation
Language learning	Pre-existing Universal Grammar and/or language of thought	Pre-existing fundamental mode of representation and learning
Perception (as opposed to peripheral transduction)	Inferential theory (necessarily mediated and indirect)	Ampliative and theory-laden activity
Transformation of representations (e.g., inference, computation)	Combinatorial syntax	Vector transformations
Fixation of beliefs	Realist: dependent on perception and relations to the external world; Isotropic and Quinean process (Fodor 1983)	Interpretationalist; Internal process
Frame problem	Accessing relevant information among vast store of linguistically-encoded structures problematic	Received information results in an almost instantaneous activation of a prototype, a cognitive system is able to access the relevant consequences of a change in the environment almost immediately, thereby avoiding the frame problem.[B]

[A] Not all supporters of the Classical Model support the Language of Thought Hypothesis.
[B] AI attempts to model symbol manipulation have often resulted in a version of the Frame Problem because of the

Cognitive Architectures — Comparison		
Cognitive Phenomena	**Classical Model**	**Connectionist Model**
difficulty of drawing on the relevant background knowledge.		

3.4.8 **Systematicity, Productivity, and Compositionality**

The view of the mind as a symbol-manipulator (the "classical model") holds that the systematicity (and productivity) of thought arise from the compositionality of mental representations, which in turn depends on their constituent syntactic structure. Further, the tendency of mental processes to be truth-preserving is explained by the hypothesis that they are computations which are syntactically-driven causal processes. This explanation of the systematicity of thought is not available to supporters of connectionism, so Fodor and Pylyshyn (1988) claim, and connectionism is not able provide an alternative.

As Fodor (1995, 1998a, and elsewhere) has pointed out, the systematicity and productivity features of natural languages can be used to illustrate the systematicity and productivity of thought (see the section 3.3.1 "The Computational Theory of Mind" above for a detailed discussion). Fodor and McLaughlin (1990) argue that such an explanation of the systematicity of thought is not available to connectionism: it is possible, for instance, for connectionist models to be trained to recognize "John loves Mary", but fail to recognize "'Mary loves John."' Hence, connectionism cannot not guarantee systematicity, and thus cannot provide an explanation of the pervasiveness of systematicity in human cognition. Although systematicity may exist serendipitously in connectionist architectures, the classical, symbol-system solution can offer an explanation because, in classical models, pervasive systematicity "comes for free". Attempts have been made by connectionists to rebut these criticisms. For instance, several connectionists have offered counter-examples to Fodor and Pylyshyn's (1988) "strong systematicity" position (e.g., Chalmers 1990; Pollack 1990, Smolensky 1996). Most of their arguments rest on the claim that Fodor and Pylyshyn have severely underestimated the power of distributed representations. In addition, Chalmers (1990) has pointed out that if Fodor and Pylyshyn were correct, then no neural nets, even those that implement a classical architecture, would exhibit systematicity, and, given that the brain is a neural net, systematicity would be impossible in human thought. Clearly such is not the case and, therefore, it should be possible, at least in principle, for systematicity to exist in a neural network (the human brain being a "proof of concept").

Fodor and Pylyshyn's (1988) argument goes something like:

P1: If thought is systematic, then internal representations are structured;

P2: Representations in connectionist networks are unstructured;

C: The architecture of thought is not connectionist.

For both the classical and connectionist models, thought is systematic but, as Andy Clark (1989:146) points out, Fodor and Pylyshyn hold that systematicity of thought is a *contingent*, *empirical* fact: "It is an empirical question whether the cognitive capacities of infraverbal organisms are often structured that way[37]" (Fodor & Pylyshyn 1988:41). For Clark, the fact that the mind is not composed of myriads of unrelated thoughts is not an *empirical* fact but is a *conceptual* fact because a "radically punctate mind is no mind at all". In addition, behavioural responses result from a network of thoughts; thought ascription is a process that is abstract, idealising, and holistic; and what needs *empirical* explanation is the systematicity of *behaviour*, not the systematicity of thought (Clark 1989:147). On Clark's account, it is the systematicity of *behaviour* that grounds thought ascription and, given the holistic nature of thought ascriptions, there is no reason to suppose that the recurrent and recombinable elements of such systematicity need have a conceptual-level semantics (*Ibid*). Hence, Clark and connectionists in general are holistic and behaviouristic about thought ascription, and deny that the way thoughts are described in natural languages can be used as the basis for describing the systematicity of "in-the-head" processing; in particular, they reject the Language of Thought hypothesis.

Clark (1989) has proposed that compositionality could be an emergent property of a network. His main argument is that Fodor & Pylyshyn assume that internal representations have a semantic and syntactic structure similar to that of a natural language, but the internal representations in a distributed connectionist network are not similar to parts of conceptual-level descriptions — they are, in contrast, sub-patterns that include the micro-features that are specific to the context. Internal representations, for Fodor & Pylyshyn, have a semantic and syntactic structure which is similar to that of natural language sentences — the basis for their claim that there is a "language of thought" (Fodor & Pylyshyn 1988:12). But what this suggests is that the thought that "Mary loves John" is related, in some way, to a

[37] It is an empirical question whether an infraverbal organism who has been taught the meanings of individual symbols for "Mary", "give", "banana", and "John", and responds correctly to the order "Mary give banana John", would also respond correctly to the command "John give banana Mary" without additional training. So far, the answer seems to be "no", but this may indicate only that they cannot use symbolic communication systematically rather than that their thought processes are not systematic.

complex internal representation which has the meanings of "Mary", "loves", and "John" as context-independent parts — the semantics of "Mary loves John" is inherited from the semantics of "John, "loves", and "Mary". Connectionists, in contrast, do not posit such context-independent items that are congruent with the parts of of conceptual-level descriptions. Hence an argument for the conceptual-level compositionality of internal representations cannot be used against connectionism (Clark 1989:148). Using the notion of a cluster of micro-features rather than semantic and syntactic structures might be better able to explain why the thought that "John loves London" does not function in the same way as "John loves Mary". If we use Fodor and Pylyshyn's LOT argument, then the thought that "Pierre loves London" is a complex internal representation which has the meanings of "Pierre", "loves", and "London" as context-independent parts and the semantics of "London loves Pierre" would be inherited from the semantics of "Pierre, "loves", and "London". For any reasonably competent speaker of English, however, "London loves Pierre" is a semantically incorrect proposition even though it is syntactically correct. The requirement that the (atomic) constituents be context-independent creates a problem for explaining the intuitive incongruity of "London loves Pierre" even though the LOT argument is correct in that we can at least think it. It is clearly possible for a natural language sentence to obey the language's rules of syntax and be composed of meaningful, context-independent parts, yet still be meaningless, as Chomsky's famous example "colorless green ideas sleep furiously" demonstrates.

3.4.9 **The Intentional Stance**

In explaining and, especially, predicting the behaviour of an object, Dennett (1987:43-68) defines three levels of abstraction:

> *The physical stance* which is the most concrete and is concerned with material composition (at the level of physics and chemistry);
>
> *The design stance* is concerned with purpose and function (at the level of biology and engineering); and
>
> *The intentional stance* is the most abstract and is concerned with belief, planning, etc. (at the level of mind and software).

The *intentional stance* is characteristically the attribution of beliefs and goals and is, therefore, the concern of cognitive science. Connectionists disagree with supporters of the classical model of cognition about the nature of cognitive processes, they nevertheless often agree that cognition can be usefully described through the use of theoretical notions like

beliefs, goals, knowledge structures, plans, and concepts (e.g., Smolensky 1991; Clark 1993).

3.4.10 Co-extension and Co-reference

When it comes to co-extensive concepts, connectionism offers a clear alternative to that of the supporters of the classical, symbol-processing model. For supporters of the classical model, such as Fodor and Lepore, Putnam, and many others, meaning is externalist; they support a view of meaning as reference or denotation. On the account of concepts as presented by Fodor, concepts have a content that is causally linked with an entity in the external world. In order to individuate co-extensive but nevertheless distinct concepts, a "something else" must be postulated and this "something else" is internal, is psychological. In contrast, conceptual content for connectionists like Paul Churchland, is a portrayal of the world that has no automatic referential connection to the external world, and what constitutes meaning is a creature's set of beliefs ("cognitive portrayals" of the world) (Churchland 2007:135). This connectionist theory of meaning is thus *internalist*. By making a creature's set of beliefs constitutive of meaning, connectionists are "meaning holists" and are constrained to explain how two creatures could be in the same psychological state; how they could "share" a concept. This is discussed further in the section on connectionist theories of concepts in Chapter 5, but briefly, the claim is that it is possible to define and use a notion of sameness and similarity of *configuration-in-activation-space* for concepts in the minds of two people where "sameness" of configuration would be taken to mean that they are in the same psychological state and "similarity" of configuration would mean that they had similar (but not identical) concepts (Churchland 2007:134).

3.5 PART IV: CONCLUSIONS

A study of the physiology of the eye reveals that no visual image of the world is transmitted to the brain; what is transmitted is primarily information about: colour, contrast and edges, and change signalling movement. Even this information is incomplete: while there are approximately 120 million rods and 6 million cones receiving stimuli, there are only about one million retinal ganglion cells to carry information to the brain. The mind is, nevertheless, able to construct images and categorize percepts based on very limited and often ambiguous input. The Computational Theory of Mind is loosely based on the digital computer analogy, but a more productive and accurate analogy would be that of a system of

pattern-recognition and pattern-completion. Such a system is more compatible with connectionism.

Connectionists reject the symbolic paradigm hypothesis concerning formalized knowledge because actual artificial intelligence (AI) systems built on the hypothesis seem too brittle, too inflexible, to model true human expertise; and, further, the process of articulating expert knowledge in rules seems impractical for many important domains (common-sense, for instance). Connectionists contend that the symbolic paradigm hypothesis has not contributed any insight into how knowledge is represented in the brain. In particular, only the connectionist model has an account of implicit knowledge (non-conceptual content), that is, of the knowledge which the subject may not be able to articulate but which, nevertheless, is an authentic propositional attitude. Expert judgement, at least, is not naturally modelled as the *sequential interpretation of a linguistically formalized procedure* posited by the Computational Theory of Mind and which takes place at the conceptual level. Intuition, for example, must be formalized at the sub-conceptual level (Smolensky 1988:§6).

Human knowledge (and possibly that of many other species) consists predominantly, according to connectionist theories, of a substantial set of "prototypes"[38] which are understood to be the clearest cases which have been learnt and which occupy a specific region (a "hot-spot") inside an activation space, and where a concept encompasses a range of related cases:

> The picture I am trying to evoke, of the cognitive lives of simple creatures, ascribes to them an organized 'library' of internal representations of various prototypical perceptual situations, situations to which prototypical behaviors are the computed output of the well trained network. (Churchland, 1989:207).

A relevant example is that of the mine/rock network that has been able to out-perform human operators in distinguishing between sonar echoes returned from explosive mines and the solar echoes returned from rocks of comparable sizes. As discussed further in Chapter 5, the Prototype Theory of concepts is well suited to implementation in connectionist models given that neural nets are capable of "learning"[39] the difference between subtle statistical patterns that would be very hard to implement in rule-based classical models. Further, Churchland suggests that the model he advocates is applicable to many types of "prototypes", including categorical, social, temporal, and motivational) suggesting that it can

[38] This view has much in common with the Prototype Theory of Concepts (see Rosch 1973a, 1973b, 1975, 1978) for which some members of a category are more typical than others forming the concept. A pigeon is usually viewed as a more typical bird that an ostrich. A PIGEON prototype would occupy a more central position in the "hotspot" of the BIRD activation space.

provide a unified account of a large part of explanatory understanding (Churchland 1989:212–8). Explanatory understanding, on this account, consists in the activation of a prototype vector whereby the incoming activation pattern is "amplified" because the prototype results from the previous complex processing of many examples during its learning stage (Churchland 1989:210-2). Hence, because the received information results in an almost instantaneous activation of a prototype, a cognitive system is able to access the relevant consequences of a change in the environment almost immediately (Churchland 1989:178), thereby avoiding the "frame problem" to which the classical, symbol-processing model is liable, whether in its Massively Modular or Fodorian domain-general central system incarnation.

Classical cognitive science uses symbolic representations which closely mirror the structure of propositionally-described knowledge: the Computational Theory of Mind is a combination of the notion that mental representations are syntactically-structured with the idea that mental processes are calculations which act only on the symbolic form of the mental representations. Relationships between a proposition's concepts are represented explicitly by symbolic structures in either a hierarchy or through the use of rules. Determining the relevance of what information needs to be represented in the hierarchy or input to a rule can quickly lead to a "computational explosion". As Fodor has stated, one of the major failures of computational cognitive psychology is its inability to provide a convincing account of abduction. A possible solution to the "globality-cum-complex-sensitivity" requirements of abduction and other similar cognitive processes might lie in the choice of cognitive *architecture*, but Fodor concludes that:

> The substantive problem is to understand, even to a first approximation, *what sort* of architecture cognitive science ought to switch to insofar as the goal is to accommodate abduction. As far as I know, however, nobody has the slightest idea. (Fodor 2000:47)

Proponents of the connectionist model claim to have just such an idea. It should be noted, however, that, in a connectionist model, there is no real distinction between processes and memory stores; in particular, there is no clear distinction between conceptual knowledge

[39] Connectionist models usually start as *tabula rasa* and "learn" through connection weight adjustment (where weighted connections represent synapses or groups of synapses). In a "trained" model, each input pattern generates an internal pattern (an abstract underlying representation) over the hidden units. This would be a little like a human learning a DOG concept through exposure to different instances of dogs. Clearly the prototypes for an Inuk exposed only to instances of huskies would be different from that of the Mexican exposed only to instances of chihuahuas. A human would have, according to nativists at least, some innate concepts such as ANIMATE, while a connectionist model would learn to categorize objects as ANIMATE or INANIMATE – they are able to generalize beyond the input patterns (see, for example, Daugherty and Seidenberg 1992).

and background knowledge. The difficulties with connectionism cited by Fodor & Pylyshyn, and others, notwithstanding, there has been a lot of recent progress in neurophysiology and the understanding of neural networks which has led to models that overcome many of the early problems with the result that neuroscientists tend to favour connectionism. That neuroscientists tend to favour connectionism[40] is not that surprising given its parallel distributed processing and the fact that the brain is massively parallel, while, in contrast symbol-manipulation models have no resemblance whatsoever to underlying brain structure.

A review of the relative strengths and weaknesses of the two competing cognitive architecture models (outlined in the Comparison table above), suggests that a hybrid architecture that combines the strengths of both, at the appropriate cognitive level, is needed. There is considerable empirical evidence that supports the view of the brain as a massively-parallel processor but how this would translate into a massively-parallel mind is questionable[41]. The problem is that the thought processes appear as sequential to the conscious mind and, as such, are amenable to being explained in symbol-manipulation terms, but much of our cognitive processes are not conscious and do not appear to be translatable into a language-like medium, as is demonstrated by the phenomenon of expert knowledge. As previously stated, there is evidence for a level of representation which lies beneath that of the sentential or propositional attitudes; and for a learning dynamic that operates primarily on sub-linguistic factors. What appears to be required is a link between the two architectures[42]. In Chapters 4 and 5, I present a possible direction that might provide a means for linking the two architectures using a variation of recent dual-process/dual-system theories.

---

[40]The best known proponents of connectionism, the neuro-philosophers Patricia and Paul Churchland, are both neuroscientists by training.

[41] Fodor and Pylyshyn, and other proponents of the Classical model, accept connectionism at the neural '(abstract neurological') implementation level and conclude that arguments for Connectionism "are coherent only on this interpretation" (Fodor & Pylyshyn 1988:1). They thus reject that the mind/brain architecture is Connectionist *at the cognitive level* even as part of a hybrid model.

[42] Clark (1989), for one, has attempted such a unification but, to date, with only limited success.

4. Dual-Process Theories of Mind

4.1 Introduction

Before offering a cognitive architecture model which differs somewhat from those that have been widely-accepted, I examine in this chapter several dual-process and dual-system theories and the attributes that these processes and systems are purported to exhibit. The results of this examination will be used in the following chapter.

Chapter 4 is divided into 4 parts:

- Part I is an overview of some of the most popular dual-process and dual-system theories found in current philosophical and psychology literature;
- Part II presents many of the problems and weaknesses of these theories;
- Part III presents a dual cognitive architecture model which compares how the two main cognitive architecture models (the classical symbolic, computational model and the connectionist model) relate to the attributes of the two system types common in many dual-process theories; and
- Part IV provides some conclusions following from the preceding discussions and observations.

4.2 Part I: Dual-process Theories – Standard View

A position common among many psychologists today, especially social psychologists[43], draws on the distinction between automatic and controlled processing. This distinction has given rise to various dual-process theories which emphasize contrasts between: heuristic processes and those that are systematic or analytic; between those that are intuitive and those that are analytic; between those that are associative and those that are rule-based; between those that are implicit and those that are explicit; between those that are experiential and those that are rational; or any combination of these. Two psychologists, Keith Stanovich and Richard West (see Stanovich 1999, Stanovich & West 1999, 2000, 2007, 2008a. 2008b), proposed that there are two cognitive systems in the mind (**System 1** and **System 2**) which underly the two process types. They proposed that **System 1** operates automatically and rapidly, with minimal, if any effort, and without any sense of

[43] Social psychology is concerned with how individual personality, attitudes, motivations, and behaviour influence and are influenced by social groups.

voluntary control; and that **System 2** allocates attention to demanding mental activities that include complex computations. The following table presents attributes associated with standard views of the differences between the two systems of a dual-process theory:

Dual-Process Theory – Differences between System 1 and System 2 (Evan 2008, Kahneman 2011)		
Cluster	**System 1**	**System 2**
I	Unconscious reasoning	Conscious reasoning
	Implicit	Explicit
	Automatic	Controlled
	Judgements based on intuition	Judgements based on critical examination
	Operates effortlessly and automatically	Operates with effort and control
	Processes information quickly	Processes information slowly
	Hypothetical reasoning	Logical reasoning
	Large capacity	Small capacity
	Default process, can be overridden by **System 2**	Inhibitory, used when **System 1** fails to form a logical/acceptable conclusion
	Unintentional thinking	Intentional thinking
	Holistic, Perceptual	Analytic, reflective
II	Prominent in animals and humans	Prominent only in humans
	Prominent since human origins	Developed over time
	Nonverbal	Linked to language
	Evolutionary rationality	Individual rationality
	Modular cognition	Fluid intelligence
III	Associative; includes recognition, perception, orientation, etc	Rule-based; includes comparisons, weighing of options, etc.
	Domain specific	Domain general
	Contextualized	Abstract
	Pragmatic	Logical
	Parallel	Sequential
	Stereotypical	Egalitarian
IV	Universal	Heritable
	Influenced by experiences, emotions, and memories	Influenced by facts, logic, and evidence
	Unrelated to working memory	Related to working memory
	Independent of general intelligence	Linked to general intelligence

For Kahneman (2011), the impressions and feelings that originate in **System 1** provide sources for the explicit beliefs and deliberate choices of **System 2**; and, while the automatic operations of **System 1** may generate complex patterns of ideas, logically progressive thought construction can only take place in the slower **System 2**. Not all dual-process theorists (J. Evans, for example) agree that, underlying the two process types, there are

two cognitive systems (Stanovich's "**System 1**" and "**System 2**"), holding instead that there are several cognitive systems (especially for unconscious processes) or even just one complex system. Nevertheless, what is common to all dual-process and dual-system theories is the demarcation between processes that are unconscious and those that are conscious (Cluster I in the previous table): one group of processes (or "**System 1**") consists of those that are fast, unconscious, and automatic; and the other of processes (or "**System 2**") that are slow, conscious, and controlled. "The highly diverse operations of **System 2** have one feature in common: they require attention and are disrupted when attention is drawn away" (Kahneman 2011). Whether dual-process or dual-system, these theories have in common the contrast between fast, automatic, or unconscious processes and those that are slow, effortful, and conscious (Samuels 2006). Following the terminology suggested by Evans (2008), I will use the "Type 1" and "Type 2" terminology in order to avoid any commitment to a two-system view. While I accept the *basic* view of different cognitive systems as presented by Kahneman and others, I will diverge significantly from their positions on the constitution of these systems.

4.3 Part II: Problems with the Standard View

In this section, I evaluate some of the features attributed to the standard or commonly accepted two process, or two system, accounts. While the list of attributes may vary slightly from one account to another, they generally follow those given in the previous table.

4.3.1 Conscious and Unconscious Processes

Most reasonably competent car drivers have experienced driving, sometimes long distances, without being conscious of doing so. Keeping to the speed limit, avoiding obstacles, following the desired route, and so on, can be accomplished without any conscious processing. There is a major difference between performing an action and being conscious of performing the action. There is also a major difference between experiencing an event and being aware of the experience: there are many anecdotes about athletes who are unaware that they have sustained serious injury during a sporting event and only become aware (experience pain) after the event has ended. Certain neurological disorders provide some evidence that the areas of the brain correlated with being conscious of performing at least some actions are distinct from those correlated with actually performing them. Visual perception involves many different cortical areas contributing to visual perception but visual information appears to be processed by two major cortical systems: 1) a ventral visual

pathway extended to the temporal lobe and associated with conscious perception, identification, and recognition of objects using the objects' intrinsic visual properties; and 2) a dorsal visual pathway extended to the parietal lobe and associated with the exercise of visual-motor control of objects[44]. Certain types of brain lesions in the ventral visual pathway can lead to "blindsight", a disorder discovered by the psychologist Larry Weiskrantz (1986). Patients suffering from "blindsight" report that they are unable to see objects, but are nevertheless able to guess at the location or even visual features of the objects with an accuracy significantly greater than chance. Weiskrantz hypothesized that the person suffering from blindsight perceives a light in their blind area but is unable to report the perception because it is not being monitored.

Dissociation of the two visual pathways can help explain the phenomenon of driving "unconsciously" and may, in fact, provide advantages when swift action is required. As Immanuel Kant argued, for any simple act of self-preservation that man performs, he would be far better served by instinct than by reason (GMM, 396). For dual-processing theorists, the dissociation of conscious and unconscious processing systems may help explain the difference between novice and expert performance: novice performance consists predominantly of conscious rule-following but expert performance is predominantly independent of any conscious rule-following. Novice versus expert learning and performance is discussed further in the "Implicit and Explicit Learning" paragraph below.

Experiments using electroencephalography performed by H. H. Kornhuber and I. Deeke (1964) demonstrated that, when subjects are asked to raise a finger, brain activity builds up as much as one and a half seconds before finger movement commences. Thus, anyone observing the brain activity would know that the subject would raise a finger about one second before the subject would have the subjective experience of freely initiating movement. The neuroscientist, Benjamin Libet, used electroencephalography (EEG) to study "freely voluntary" acts:

> Freely voluntary acts are preceded by a specific electrical change in the brain (the 'readiness potential', RP) that begins 550 ms before the act. Human subjects become aware of intention to act 350-400 ms after RP starts, but 200 ms before the motor act. (Libet 1999)

Libet's results led him to claim that the volitional process is initiated unconsciously. Susan J. Blackmore wrote that "Many philosophers and scientists have argued that free will is an

[44] The ventral and dorsal pathways are shown in the "Blindsight" illustration "Visual Recognition and Action (© McGill University)" on page 30.

illusion. Unlike all of them, Benjamin Libet found a way to test it" (Blackmore 2007). Libet did, however, also claim that the conscious function could still control the outcome by inhibiting (or "vetoing") the action, but the results of experiments, such as those conducted by Kornhuber, et al., and, later, by Libet himself, suggest that unconscious processes in the brain are the true initiators of voluntary acts thereby suggesting a form of neurophysiological determinism which, if true, means that decisions are the outcome of pre-existing causal factors (beliefs, needs, preferences, etc.) Incompatibilists claim that this is true and its truth is incompatible with the existence of free-will. Compatibilists, in contrast, claim that the truth of neurophysiological determinism is compatible with choices based on one's own beliefs, desires and inclinations. The question is what role consciousness plays in these choices.

Daniel Wegner (2002, 2003) hypothesized that there is no difference between the experience of having personally caused an action and the experience of cause and effect in general. He conducted several studies to test this hypothesis and reported that subjects believe that something causes something else if and only if what the subjects thought to be the causal event:

- occurs just before the putative effect (the *priority principle*);
- is consistent with the putative effect (the *consistency principle*); and
- is the only apparent cause of the putative effect (the *exclusivity principle*).

Wegner provides considerable evidence to support his view that we are largely ignorant of when and how we act and the ignorance is of two types:

- we are acting without realizing that we are so doing; and
- we are not acting, but think that we are.

He cites, as examples of the former type, the ouija board; facilitated communication; water divination; and hypnotism. An example of the second type of "ignorance" is provided by his 'I-Spy' study in which subjects incorrectly believe that they have selected a figure on a computer screen through "priming" – they were previously encouraged to think about the figure. In addition to these types of "ignorance", are the common problems of overestimating the effect that we have on objects and people around us, and of forgetting or even confabulating earlier actions in order that the actions appear in a better light. Michael Gazzinga (1988) has suggested that the brain has a monitoring system that often

"constructs hypotheses" and will often make guesses that are not always correct. Gazzinga gives the example of split-brain patients who often fabricate a false hypothesis to explain incorrect selections during experiments in which they are shown two different images – one visible only to left eye (controlled by the right hemisphere) and the other only to the right eye (controlled by the left hemisphere) – and are then asked to use both hands to pick two items that they most associate with what they have seen. In most patients the language centre is in the left hemisphere and, because the corpus collosum has been severed, they cannot report information available only to the right hemisphere. The patients will fabricate an explanation for the selection made by the left hand. Gazzinga suggest that it is not just in damaged brains that a monitor will fabricate or guess when needed, but that fabrication and guesswork is a normal part of its functioning. This fabrication and guesswork is performed unconsciously.

Wegner (2003) reports that in many cases of clinical anomalies, such as 'alien hand syndrome' and schizophrenia, the experience of free will is "fabricated". While we experience the conscious willing of an action, Wegner maintains that acts of will cannot be taken to be the uncaused causes of such action and that the experience of conscious will arises from inferring a causal path from thought to action (shown as a purple arrow in Illustration 12) while the actual causal paths (shown as a green arrow in Illustration 12) are not consciously present. Wegner (2003) concludes that "[T]he experience of conscious will is a marvelous trick of the mind, one that yields useful intuitions about our authorship – but it is not the foundation for an explanatory system that stands outside the paths of deterministic causation." As Voltaire put it: "When I can do what I want to do, there is my liberty for me ... but I cannot help what I do want." This is not to say that, given sufficient information about an agent's beliefs, desires, and inclinations, his behaviour can, at least in principle, always be correctly predicted. Chaos Theory, the study of nonlinear dynamics, demonstrates that, for any system, there will always be differences smaller than the smallest measurable difference (see Gleick 1988).

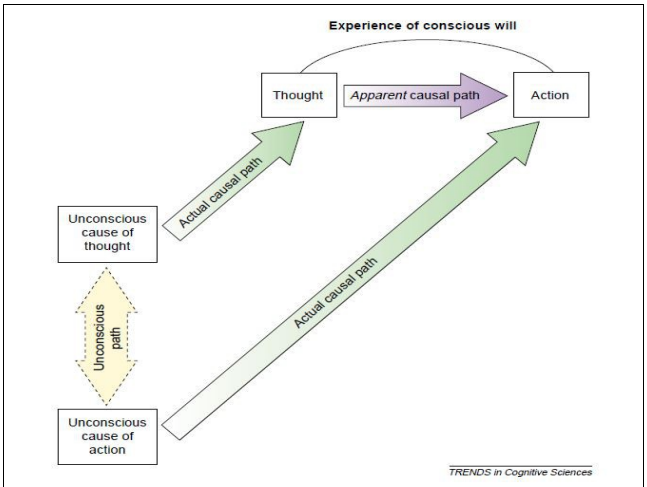

Illustration 12: Experiencing Conscious Free Will (from Wegner 2003 - modified from Wegner & Wheatley 1999)

The findings of the studies by Kornhuber et al., Libet, Wegner, and others lend support to the view that there are two cognitive processes but differ with most dual-process theorists on their causal effect. The findings suggest that:

- in one process, certain unconscious mental events cause a thought while other unconscious mental events cause actions; and
- in another process, the apparent (but not real) link between cause and action is experienced consciously.

Libet's studies indicated that awareness is a unique phenomenon in itself, distinguished from the contents of awareness:

> The content of an unconscious mental process . . . may be the same as the content with awareness of the signal. But to become aware of that same content required that stimulus duration be increased by about 400 msec. (Libet 1999)

Libet's results have been replicated by others (e.g., Soon et al., 2008) using using more accurate fMRI techniques, and these later studies report that the behavioural output of a decision can be encoded in brain activity in regions of the frontal and parietal cortex as much as ten seconds before the subject becomes aware of the decision. For Libet, the awareness of the decision to veto could be thought to require preceding unconscious processes, but the content of that awareness (the actual decision to inhibit action) is a separate feature that need not have the same requirement. The role of conscious free will, on this account, would not be to initiate a voluntary act, but rather to control whether or not the act takes place. There is a weakness, however, in Libet's interpretation: if, as his results indicate, unconscious brain activity always precedes a conscious decision, then it is difficult to argue that the conscious veto is not also preceded by unconscious brain activity; thereby suggesting that the experience that "veto" is exercised consciously is also illusory.

The results of brain potential recordings of Parkinson patients suggest that the supplementary motor area, which is not available to introspection, has a priming role in the preparation and initiation of voluntary movement (Kornhuber & Deecke 1964). Studies of mirror neurons, the cells in the brain that fire when a particular action is performed and also when someone else is observed performing the same action, has led to the suggestion that this "mirroring" is the neural mechanism by which we automatically (unconsciously) understand the actions, intentions and emotions of other people[45]. Galen Strawson took the posi-

[45] University of California - Los Angeles. "First direct recording made of mirror neurons in human brain." *ScienceDaily*, 13 Apr. 2010. Web. 21 Sep. 2013.

tion that some thoughts have *qualia;* some thoughts have introspectively accessible, phenomenal features. He claimed that there are *experiences of* something (such as understanding a sentence or suddenly thinking of something) but that these experiences are not reducible to the associated sensory experiences: "Each sensory modality is an experiential modality, and thought experience (in which understanding-experience may be included) is an experiential modality to be reckoned alongside the other experiential modalities" (Strawson 1994:196). If we take the neuroscientific or cognitive approach, however, then we may be able to avoid what Gilbert Ryle (1949) tagged as a "category mistake". There is a considerable difference between discussing what we experience consciously and what causes the experience. Contrary to what we "feel" is happening, all the studies cited in this section suggest that conscious experiences are causally inert. This claim is contrary to almost all dual-process and dual system theories, especially those that are prevalent among social psychologists (see Wilson et al. 1991, 2000, Epstein 1994, Chaiken 1999, Bargh 2011) which stress the distinction between controlled (conscious) processing and automatic (unconscious) processing and hold that there are causal interactions between conscious thoughts, feelings, and perceptions, and behaviour. It should be noted that the "inhibitory" function attributed to *Type 2* could be explained as merely the monitoring and verbalization of a subconscious "conflict" resolution resulting from much the same mechanism as an alarm call would in other species.

4.3.2 **Language and Reasoning**

There are two basic assumptions behind most dual-process or dual-system theories such as that proposed by Kahneman, namely that logical reasoning is conscious, that rational judgement, in general, takes place consciously, and that such reasoning is associated with language. If this is correct, then it would appear that other non-linguistic species, or even pre-linguistic infants, cannot reason logically nor make rational judgements. It would appear that, for such theories, natural language competence is required for *Type 2* to operate successfully. I will agree that consciousness uses some form of imagery, usually verbal imagery (*pronunciation*, in the case of spoken languages; visual or even haptic in the case of sign languages), but I do not agree with these divisions between intuition and reasoning and between hypothetical and logical reasoning and will argue that all these processes are conducted at the non-conscious (*Type 1*) level. I claim that differences between the reasoning capacities of linguistic and pre-linguistic humans and non-linguistic species are questions of degree but that development of language capacities enables some reasoning to

be consciously experienced and reported**.** Several studies have demonstrated that preschoolers demonstrate deductive reasoning skills (see Dias & Harris, 1988, 1990; Hawkins, Pea, Glick, & Scribner, 1984; Richards & Sanderson, 1999). On the Piaget account of sensori-motor development, symbolic problem solving is manifest in infants from 18 to 24 months of age. Recent fMRI experiments[46], suggest that logical reasoning does not rely on the grammar of natural language and, along with other neuroimaging and neuropsychology evidence, appear to show that high-level cognitive functions, such as arithmetic, problem-solving, and theory of mind, are remarkably language-independent. Anyone who has watched a crow solving problems that require several steps (including creation of tools) is likely to accept that it too is using logical reasoning, albeit without being consciously (refectively) aware of its thought processes.

4.3.3 **Implicit and Explicit Learning**

The standard dual-process or dual-system position is that the conscious process (or **System 2**) includes rule-following. Certainly when learning a new skill, such as how to drive a car, to play a musical instrument, or to play chess, a novice will be conscious of following rules. Experts, in contrast, are usually unaware of following any rules and, as Socrates discovered when trying to find how Eurythro, a supposed expert in the field, could recognize *piety* and *justice* (see the Platonic dialogue *Euthyphro)*, experts are often unaware of how they are able to perform the skills. Given the speed at which expert chess players perform, they may depend almost entirely on intuition rather than relying, if at all, on analysis and comparison of alternatives (Dreyfus & Dreyfus 1986); experts do not follow rules consciously and sequentially. Pattern recognition is an integral part of an expert chess player's competence and that would appear to be included in *Type 1*. I would suggest that we are conscious of rule-following (attributed to *Type 2*) when learning a new skill, when examining rules, or when identifying errors in reasoning and making normative judgements, but rule-following is included in *Type 1* although not necessarily amenable to introspection.

There is a distinction between implicit and explicit learning, between implicit and explicit memory, and these are usually attributed to *Type 1* and *Type 2* respectively. There is increasing evidence that it is possible to acquire implicit knowledge without ever being able to state any related explicit rule (see Reber 1993, Sun et al. 2005). Implicit learning theorists, such as Reber, contend that implicit processing proceeds by chance without accompa-

[46]As reported by Martin M. Monti at the *Interdisciplinary Workshop on the Notion of Thought,* Ruhr-Universität Bochum, 5th - 7th June 2008

nying awareness, and that explicit processing proceeds deliberately and is always accompanied by awareness (Reber 1993). "Conscious rule-following" can, however, be described (in my opinion more accurately) as "being conscious of following rules": difficulties encountered at the unconscious level by novices result in activation of the monitoring and verbal (usually phonic) encoding systems and hence in the awareness of rule-following. The same type of difficulties do not arise in the unconscious of experts and thus their monitoring systems are rarely activated.

Procedural or motor learning (as opposed to *declarative* learning) has been defined as "…a set of [internal] processes associated with practice or experience leading to relatively permanent changes in the capability for responding" (Schmidt, 1988:346; Schmidt, 1991:51). The division of memory into two systems can help explain infantile amnesia: skills learnt in infancy may be retained but not explicitly; and adults are unable to retrieve episodic memories formed before the age of 2–4 years. This is likely because the ability to encode memories in some linguistic form is required for storing explicit memories – the hippocampus, which is necessary for storing explicit memories, does not mature prior to 3 or 4 years of age. There is anecdotal evidence from deaf individuals who did not acquire a (sign) language until relatively late (e.g., Helen Keller and the Nicaraguan deaf children who developed their own sign language[47]) that they previously did not "think". It is probably more accurate to say that, although they could think, they we*re not co*nscious of thinking before acquiring a language.

The use of short-term memory (attributed to *Type 2*) is involved in links between input stimuli (of whatever modality) and the relevant concepts and encyclopaedic knowledge stored in semantic memory (a subset of long-term memory)[48]. The existence of the relevant semantic memory items is what renders the stimuli *meaningful*. Even an incomplete semantic web of concepts and encyclopaedic knowledge can render a term like 'electron', for example, meaningful: we can "sort of know what it means" without having acquired all the scientific concepts necessary for accurate reference fixation. We do not need to keep deferring to experts whenever we use scientific or technical terms because, as long as our idiosyncratic definitions of the terms are "close enough", we will still be able to understand each other which is all that is required for day-to-day communication. When we have insufficient knowledge or understanding for a complete grasp of the meaning of the term, we may re-

[47] See Kegl (1994)
[48] See Illustration 16 "Psychological View of Memory and Cognitive Processing" on page 124 for a diagram of the relationship between short-term memory, concepts, and encyclopaedic knowledge.

sort to statements such as "you know what I mean" and most of the time that is sufficient for communication purposes. The meaning of "meaning" is discussed further in the next chapter.

The association of conscious thought with short-term (working memory) can help explain some of the attributes associated by dual-process theorists with *Type 2*: slow process of information, small capacity, and its links to general intelligence, but, as Evans (2008) states, "skeptics may see this as the only firm foundation on which the various dual-process theories stand: There is one conscious working memory system and everything else." I will argue in section 5.5 that this is the case: there is a very limited conscious mind and everything else.

4.3.4 **Evolutionarily Old versus Evolutionarily Recent**

Another contrasting attribute between *Type 1* and *Type 2* is that *Type 1* is evolutionarily relatively old (is prominent in both humans and many other species) and that *Type 2* is evolutionarily more recent (is prominent only in humans). The claim is that the two *Types* fulfil different functional roles. The claim is that certain attributes of *Type 2* (the link to language, the attributes of conscious, logical, and hypothetical reasoning, and the like) are uniquely human. The experiential *Type 1* operates rapidly and unconsciously, while *Type 2* operates slowly and consciously using the medium of language. Many of the proposed *Type 1* processes utilize, however, regions of the brain (the neo-cortex) that evolved relatively recently. There are, in addition, many ethological studies that undermine the clear distinction between the two process types. The increase in the size of the neo-cortex of primates has been attributed to the growing complexity of their social interactions given that the ability to predict the behaviour of conspecifics appears to confer a large evolutionary advantage (Dunbar 2001)[49]. The parts of the cortex responsible for social skills, which includes language functions and theory of mind, have increased in size and complexity because they improved social skills. Sensory and motor regions are located in the neo-cortex the surface of which is greater in predatory mammals than in herbivorous mammals probably because the ability to hunt successfully relies on a highly evolved sensori-motor system.

[49] What is being suggested here is that there is a "positive feedback loop": individuals with a more developed prefrontal cortex would exhibit greater associative capacities which would enable them to make their behavioural responses more unpredictable; selective pressure would then result in other members of the species developing a more complex prefrontal cortex which would enable them to predict these responses, leading to the generation of even more unpredictable behaviours, and so on.

4.3.5 **Modularity and Dual-Process Theories**

Modularity of mind theories are discussed in some detail in Chapter 3 but, briefly, Chomsky (1980) introduced the notion of domain-specific cognition by proposing that that humans are born with domain-specific systems of knowledge and Fodor (1983) presented a famous argument for modularity of mind that the mind consists of a number of functionally-specialized mechanisms which evolved in response to recurring adaptive problems. Fodor's position is not that all mental systems are modular but rather that only peripheral subsystems are modular. Consistent with Fodor's position, Evans (2012) stresses that the evidence of subject's performance on dual-processing tasks points clearly to there being more than two cognitive systems and hence there has been a shift to theorizing that there are two "minds" with different evolutionary histories and each of which are composed of multiple sub-systems. Other researchers (Moshman 2000, Osman 2004) claim that the variety of processes accomplished could not be accommodated by just two systems. Moshman actually proposes four possible types of processing: implicit heuristic processing, implicit rule-based processing, explicit heuristic processing, and explicit rule-based processing.

Fodor (1983, 2000) also proposed a form of dual-process theory but his contrasted a combination of input transducers and mandatory, informationally-encapsulated cognitive modules (such as language and vision) with central, non-modular general-purpose systems. On this model, the domain-specific input cognitive modules compute representations which become input to the central, general purpose cognitive systems (Fodor 1983:90-91); the input modules "present the world to thought" (Fodor 1983:101). The input representations and computational processes of these modules are relatively impenetrable to consciousness. Thus, this model also describes two systems: one that includes unconscious computational processing and the other that includes conscious reasoning. The representations computed by the input modules may need to be corrected (given that perceptual information is often incomplete) using background knowledge and, perhaps, output representations from other modules – Fodor calls such a process "the fixation of perceptual belief" (Fodor 1983:102). Fodor's "System 2" is more encompassing than the standard view as given in the "Dual-Process" Table above but his systems are subject to some of the same criticisms as the standard view.

Dual-process theories, such as Fodor's and several others, appear to endow *Type 2* processes with significant ability to control behaviour but, as discussed previously, many

neuro-imaging studies indicate that volitional processes are initiated unconsciously. Activity in specific regions of the human brain can be used to predict the outcome of a motor decision that the subject was not yet conscious of having made. These findings would seem to indicate that decisions are first made at a subconscious level and only after are translated into what appears to be a "conscious decision". Retrospection leads the subject to believe that an action resulted from an exercise of freewill.

Most tasks are performed automatically but, even for common tasks such as facial recognition, brushing teeth, or speaking, highly complex neural networks are employed. Many of these neural networks are highly specialized (acting as cognitive modules) and function unconsciously. An expert pianist, for instance, knows how to play the piano but would have difficulty in telling how they know what finger they use to press a particular key – in fact, when asked how they know what fingers to use, some pianists have reported experiencing a temporary inability to play. Visual perception also uses specialized neural circuits which, as Fodor (1983) has argued, are informationally-encapsulated and hence, are inaccessible to introspection and are the basis of many optical illusions.

4.3.6 **Cognitive Correlates of Consciousness**

If we examine the table " Dual-Process Theory – Differences between System 1 and System 2" on page 91, the features listed in Cluster 1 for System 2 (*Type 2*) represent cognitive correlates of consciousness for most standard dual-process views (see Evans 2008). It is possible, however, to be conscious without any of these correlates. This is especially the case when we are relaxed and, perhaps, listening to music or the sound of waves or looking at a sunset, or even engaged in yoga or transcendental meditation. In such situations, we are not conscious of reasoning; nor of explicit or controlled thought; we are not making judgements based on critical examination; we are not aware of any operations requiring effort and control; nor of reasoning logically. Yet we cannot be said to be "unconscious" in these situations. Hence, these features cannot be cognitive correlates of consciousness. In the next chapter, I suggest a different set of cognitive correlates of consciousness,

PART III: DUAL COGNITIVE ARCHITECTURAL MODELS

One of the accepted attributes (in Cluster III of the previous table) of **System 1** is that it is "parallel" whereas the related attribute of **System 2** is that it is "sequential". If such is indeed the case, then the parallel nature of **System 1** could be modelled by a connectionist cognitive architecture while the sequential nature of **System 2** could be modelled by a

symbolic cognitive architecture. Some of the relevant attributes of the two models are given in the following table (using the "Cognitive Architecture – Comparison" table in Chapter 2:

Dual-Process Theories and Cognitive Architectures — Comparison			
System 2		System 1	
System 2 Attributes	**Classical Symbolic Model Features**	**System 1 Attributes**	**Connectionist Model Features**
Sequential	Sentential; Symbol structure with combinatorial syntax and semantics; Local	Parallel	Massively-parallel; Activation patterns; Can be symbol (node) but not necessarily; Usually distributed
Linked to language	Linguistically-encoded — word-like in a language-like system	Nonverbal	Points, regions, or trajectories in a neuronal activation-vector space — "prototype"
Domain general; intentional thinking	Reference or denotation; Beliefs irrelevant	Domain specific; unintentional thinking	"Portrayal of the world" with no automatic referential connection to the external world; beliefs constitutive of meaning
Logical	Result of a practical syllogism through process of symbol manipulation	Pragmatic	Output of a neural net in response to a specific set of inputs
Rule-based	Rule-governed symbol manipulation	Associative	Vector transformation
Analytic, reflective	Inferential theory (necessarily mediated and indirect)	Holistic, perceptual	Ampliative and theory-laden activity
High effort; processes information slowly	Accessing relevant information among vast store of linguistically-encoded structures problematic	Low effort; processes information quickly	Received information results in an almost instantaneous activation of a prototype, a cognitive system is able to access the relevant consequences of a change in the environment almost immediately

Connectionist models are more biologically plausible; they are more "brain-like" than other non-connectionist architectures like the classical, symbolic model. While the connectionist model is massively parallel, it can, however, be used to model slower, sequential processing. There have been studies (Shedden & Schneider 1991) that demonstrate that, even though a connectionist system module transmits in parallel to higher levels, comparison processes within the module need to be serialized in order that acceptable accuracy levels be maintained. This serialization corresponds to the apparently serial processing by the human brain of different types of memory mapping and visual searches. John von Neumann pointed out in 1956 that the structure of the brain is such that serial processing in general would be very slow and inaccurate, but, he claimed :...large and efficient natural automata are likely to be highly *parallel*, while large and efficient artificial automata will tend to be less so, and rather to be *serial*" (italics in original) but that the brain is able to compensate for its

inability for fast and accurate logical "depth" through the use of logical "breadth" – it tends "to pick up as many logical (or informational) items as possible simultaneously, and process them simultaneously (von Neumann 1956/2000:51). Rumelhart explains:

> Given that the processes we seek to characterize are often quite complex and may involve consideration of large numbers of simultaneous constraints, our algorithms *must* involve considerable parallelism....Although the brain has *slow* components, it has *very many* of them....Rather than organize computation with many, many serial steps, as we do with systems whose steps are very fast, the brain must deploy many, many processing elements cooperatively and in parallel to carry out its activities. (Rumelhart 1989:135)

A review of the relative strengths and weaknesses of the two competing cognitive architecture models (outlined in the Comparison table in Chapter 2), suggests that accounting for all the proposed attributes of the different processes and systems might require a hybrid architecture that combines the strengths of both, at the appropriate cognitive level. There is considerable empirical evidence that supports the view of the brain as a massively-parallel processor but how this would translate into a massively-parallel mind is debatable although the fact that connectionist models can be simulated on a symbol-manipulation computer[50] is a good starting point. The problem is that thoughts that can be "broadcast" to the conscious mind[51] are sequential and are amenable to being explained in symbol-manipulation terms, but many, perhaps most of our cognitive processes do not appear to be translatable into a language-like medium, as is demonstrated by the phenomenon of expert knowledge. There is evidence for a level of representation which lies beneath that of the sentential or propositional attitudes; and for a learning dynamic that operates primarily on sub-linguistic factors. The classical symbol system cognitive architecture can be used to model very well all processes that can be converted to linguistically-encoded icons in the conscious mind, to which all processes appear to be serial. Nevertheless, the apparent serialization does not prove that the applicable unconscious processing is also sequential, and, in fact, results of brain scanning studies would suggest that unconscious processes are massively parallel.

PART IV: CONCLUSIONS

Almost everyone has experienced, at one time or another, going to sleep after being troubled by a seemingly intractable problem only to awake later with the solution. Many

[50] Digital computer simulations of a connection model treat units as virtual objects in a similar manner to the way the pieces in a computer chess game are treated.
[51] "Broadcasting" to the conscious mind is described in the next section.

computer programmers report this type of experience. It seems that our unconscious mind is very capable of finding solutions to complex problems without the involvement of our conscious mind. Many of the proposals for dual processing in higher cognition suggest an underlying neo-Cartesian dualism given their division between the automatic, instinctual, "shared with animals" attributes of the unconscious mind and the controlled, analytic, fluid intelligence, "uniquely human" attributes of the conscious mind. The claim by most dual-process and dual-system theorists that the conscious mind is causally efficacious is not supported by empirical evidence (c/f. Kornhuber & Deeke 1964; Weiskrantz 1986; Gazzinga 1988; Libet 1999, Wegner 2002, 2003; Soon et al. 2008; and many others). I contend that this claim is a result of confusing the internal "broadcasting" of some of our thought processes to the conscious mind with the initiation or control of those processes; of confusing being conscious of thoughts with thinking consciously. While I am proposing that *all* thinking is an unconscious activity, I am not proposing that consciousness is epiphenomenal. Conscious icons are linked to thoughts; the conscious mind can be taken as one stage of a cognitive "loop" rather than as a byproduct; the conscious mind is certainly not an illusion although some of the features attributed to it may very well be.

When certain thought processes are encoded in a "broadcast" form, they have been encoded as a semantic representation and combined with a verbal (usually phonic) representation to produce a linguistically-encoded icon that is "tagged" as *meaningful*. Most of the icons in our conscious mind appear as *meaningful*. It should be noticed that this "broadcasting" of meaningful linguistically-encoded icons to our conscious mind is an *internal* process and is thus of importance to an internalist theory of meaning. The "dual" system that I am proposing – a non-conscious system coupled with a causally-inert conscious system – forms a significant underpinning of the hybrid architecture model presented in the following chapter.

---

5. Proposal for a Hybrid Cognitive Architecture

5.1 Introduction

As stated in the previous chapter, one reason why the *Type 1* system proposed by most dual-process theorists operates more rapidly than the *Type 2* system is because the information processing of the former is performed in a massively parallel type of cognitive architecture whereas processes that are attributed to the latter by most dual-process theorists are necessarily sequential owing to the sequential nature of natural language and linguistically-encoded icons. This is not to say that the neural correlates of these processes are sequential but that the processes themselves seem to be sequential in so far as how they appear to the conscious mind. Given that, as I argue, the most biologically-plausible cognitive architecture is connectionist, the apparently sequential nature of thought processes is deceptive and is a result of the encoding of thought processes in a form that is suitable for communicating with oneself ("inner voice" or "inner language") as well as with conspecifics. This chapter offers a proposal for a cognitive architecture which uses, in part, Chomsky's theory of a Universal Grammar (Chomsky 2006:124) to describe the encoding of thought processes and has much in common with the theories of the role of consciousness proposed by Ray Jackendoff (1990, 2011).

Chapter 5 is divided into 4 parts:

- Part I presents a proposal for a hybrid cognitive architecture model which is broadly connectionist with a symbolic, computational component, starting with a brief overview of the proposed cognitive architecture before discussing it in more detail;
- Part II offers examples of how this hybrid model can accommodate the cognitive processes and attributes that are commonly discussed by dual-process theorists;
- Part III presents and rejects several objections that may be raised against the proposed model and against connectionist architectures in general; and
- Part IV provides conclusions and possible directions for future research.

5.2 Part I: Proposal for a Cognitive Architecture Model

5.2.1 Introduction

In attempting to design a cognitive architecture, it is necessary to emphasize the distinction between modelling cognition itself and modelling an *implementation* of cognition. Regardless of whether cognitive scientists support the classical, symbol-manipulation model or a connectionist model of cognition, they can agree that connectionism, which is the name given to the computer modelling approach based on an understanding of how information processing occurs in neural networks, may provide a useful model at the neurological, implementation level. Disagreements arise as to which provides a better model at the *cognitive* level. I will argue that connectionism provides the better model at both the cognitive *and* implementation level.

Classical, symbolic-processing systems are, I contend, too slow to correctly model reflexive and direct reasoning processes. In addition, even vociferous advocates of such systems (e.g., Fodor, Pylyshyn, Pinker, and many others) will admit that such systems cannot account for the contextual, parallel processing required for non-local reasoning. As previously discussed in section 3.3.1 "The Computational Theory of Mind", Fodor (1998b) points out that, on the Classical account, the internal syntactic structure of a thought is the sole determinant of its role in a mental process and is, hence, of little use in providing explanations of the semantic and global features of mental processes. Rational beliefs are often formed by *abductive inference*, by 'inferences to the best explanation'. If we are given only what perception presents to us as currently the fact and what beliefs are currently available to us in memory, we have the cognitive problem of finding and adopting new beliefs that are *best confirmed on balance.* Properties like relevance, strength, simplicity, and centrality apply, not to single sentences, but to belief systems as a whole, and, as Fodor states, we have no reason for supposing that such *global* properties of belief systems are syntactic. That belief systems have syntactic properties is required if the Classical, symbol-manipulation model is to be explanatorily adequate.

One reason why ***System* 1** processes as advocated by most dual-process/dual-system theorists[52] operate more rapidly than those of ***System* 2** is because the information processing of the former is performed in a massively parallel type of cognitive architecture whereas processes that are attributed to the latter by most dual-process theorists are

[52]See the table entitiled " Dual-Process Theory – Differences between System 1 and System 2" on page 91.

necessarily sequential owing to the sequential nature of conscious imagery. This is not to say that the neural correlates of these processes are sequential but that the processes themselves seem to be sequential in so far as how they appear to the conscious mind. The encoding of thoughts to produce linguistic (as opposed to other non-linguistic) icons in the conscious mind is, because it is linked to language, evolutionarily recent, but, in contrast to most dual-process theories, is, on my view, a non-conscious process. Also in contrast to other dual-process and dual-system theories, all other features commonly attributed to **System 2** are, according to the position I am taking, attributes of **System 1**; they are all attributes of the non-conscious cognitive system which is actually composed of multiple sub-systems. There is, on my view, consciousness and "everything else". The position I am advocating is that all causal processes in the mind take place non-consciously, just as results of many brain-imaging studies demonstrate. I therefore reject the position t*hat volition* and many other proposed cognitive processes advocated by dual-process theorists are correlates of consciousness. Some of these causal processes are encoded in a form that can become available to consciousness – we often become aware of difficulties and conflicts during rule-based learning, for instance – but this does not prove that causal processes, such as following rules, are conscious (**System 2**) activities. Awareness of following rules could be explained as merely the encoding of a subconscious "conflict" resolution resulting from much the same mechanism as an alarm call would in many species. Activation of an "alarm" system may be the result of a physiological reaction such as an increase in epinephrine which is part of the hormonal component of an emotional response occurring in response to stress whether environmental or psychological. The processes involved in encoding thoughts in a form required for input to the conscious mind is discussed in more detail below.

5.2.2 **Overview**

The cognitive architecture that I am proposing is predominantly connectionist. In this model, all cognitive processing is performed at the non-conscious level; the perceptual input system is modular; the output of some cognitive processes is encoded in a suitable form to become input to further thought processes; and consciousness is merely one stage in a feed-back loop – consciousness itself is causally inert. A simplified view is given in Illustration 13 on the next page. On this view, consciousness is not epiphenomenal, but can be taken to be a "spandrel", a term originally coined by Gould and Lewontin (1979), and now used by evolutionary biologists for any phenotypic characteristic that, rather than being

a direct product of adaptive selection, is merely a byproduct resulting from the evolution of some other characteristic. I contend that consciousness is a byproduct of the evolution of the alarm-call and "learning from experience" systems and is shared by many species. The "learning from experience" system will be discussed in more detail below in reference to the dual-process theorists distinction between implicit and explicit learning.

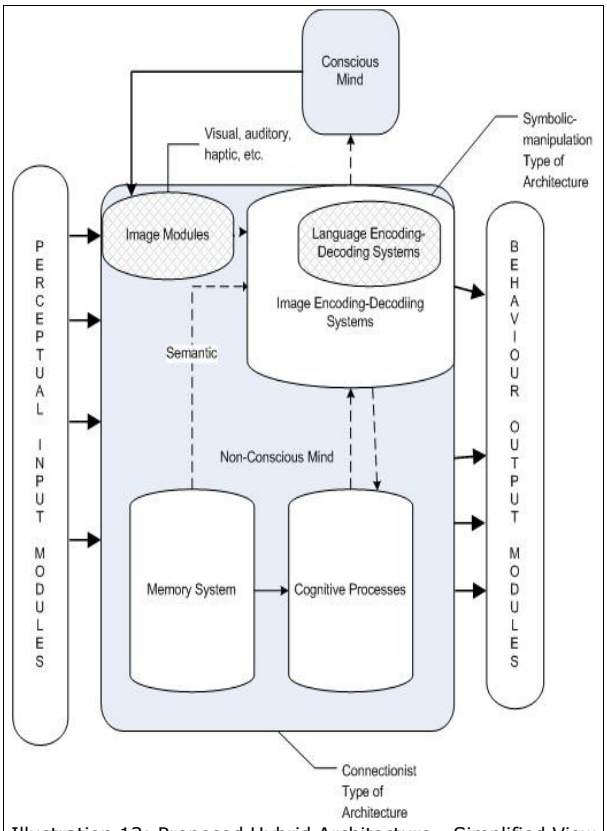

Illustration 13: Proposed Hybrid Architecture - Simplified View

5.2.3 **The Image Encoding-Decoding Systems**

The cognitive architecture model I support is connectionist. It is, however, necessary to account for our intuitions that the mind is a symbol-manipulation system, that cognition is predominantly the manipulation of propositional attitudes, and that thought and thinking take place in a language of thought consisting of a physically-instantiated system of representations with syntactic properties to which operations on these representations are causally sensitive (see, for example, Fodor 1975, 1978, Putnam 1988). For the cognitive architecture I am proposing, there is a Language of Consciousness (language of consciously-experienced icons) rather than a Language of Thought.

In the proposed cognitive architecture model, cognitive processes are not "computation over symbols". Symbols are, in this model, produced in the Encoding-Decoding Systems module which converts structured mental representations to symbolic form much the way digital computers convert internal representations to interface icons in the same way that the icons on a computer screen represent but do not resemble the complex processes of the computer (Hoffman 2009). The part of Illustration 13 labelled "Image Encoding-Decoding Systems" represents several functional systems, which may be widely distributed in the brain, and which may share functions with other processes. The input to the Encoding-Decoding Systems includes the output of perceptual processing systems such as the visual

(extensively studied by, for example, Marr, 1982; Marr & Nishihara 1978; Humphreys & Quinlan, 1987; Peterson & Rhodes, 2003), and of cognitive systems. The output from the perceptual system is "tagged" by peripheral recognition systems such as speech recognition, face recognition, and so on, and the "tags" determine which encoding-decoding systems are activated. Not all output from the Encoding-Decoding Systems is transferred to the conscious mind; on the contrary, I would contend that only a fraction of the encoded output is available to consciousness although it is often used in re-enforcing learning processes. I also contend that the syntactic structure of the encoded output is a reflection of the micro-features and structure of the input representations.

The output of the Encoding-Decoding Systems becomes input to other cognitive processes thereby re-enforcing learning – acting in the same manner as perceptual input to increase connection weights in the neural pathways. The linguistically-encoded icons are of particular importance in this regard as they constitute the "inner voice" which is of significant importance in abstract thinking and other higher mental functions such as memory and self-awareness. Recent research conducted by the Deafness Cognition and Language (DCAL) Research Centre, based at University College London, appears to demonstrate that the brains of completely deaf people never develop the "inner voice" and the theory is that, in the absence of a language, dealing with abstract concepts is highly problematic (Lyness et al., 2014). While natural language can be spoken, written, or haptic (in the case of sign language), spoken language is most crucial in developing the "inner voice" especially given the relatively small vocabulary of sign languages compared to that of spoken languages[53]. Ruben Conrad, studied the reading abilities of deaf students and reported that their reduced ability compared to hearing students was a result of their not being able to "hear" themselves reading internally (Conrad 1962, 1970, 1972). Conrad observed that short-term memory 'thrives on a speechlike input' (Conrad, 1972:231) but a later study tends support the view that Conrad's observation should be amended to "'language-like input', so that the characterization is not restricted to the speech mode" (Bellugi et al., 1975). These results suggest that the Linguistic Encoding-Decoding System has to be activated through exposure to a natural language for proper functioning (see Chomsky 1957, 1960, 1980, 2006).

[53] Sign languages typically augment hand orientations and shapes, with facial expressions and body posture, which are combined simultaneously to produce a *sign,* which corresponds to a spoke language's single word or group of words. There are currently projects to develop of computer assisted sign language recognition systems. Part of the projects involves the creation of databases, two examples of which are the Australian Sign Language database containing 7415 words, and the British Sign Language database containing 6330 signs (Melnyk et al., 2014).

Certain neurological disorders, such as blindsight, can provide evidence that, although the encoded icons are available to cognition, they may not always be available to the conscious mind. As discussed earlier[54], damage to one of the two main visual pathways can lead to subjects having absolutely no awareness of any visual stimuli but still being able to predict visual aspects such as location, or type of movement; while damage to the other pathway can lead to subjects having some awareness of visual aspects such as movement within the blind area but without having a visual percept (Armstrong 1968, Weiskrantz 1986). In regard to activation of the Linguistic Encoding-Decoding Systems, the example that can be used is that of Albert Einstein who, when describing his own thought process, said that he used images to solve his problems and only found the words later (Pais, 1982), and, even more strongly, that "I have no doubt that our thinking goes on for the most part without the use of symbols, and, furthermore, largely unconsciously" (Schilpp, pp. 8-9). Studies by Sandra Witelson et al. (1999) suggest that that the regions in Einstein's brain that are associated with speech and language are smaller, whereas the regions involved with numerical and spatial processing are larger than normal. These examples do not, of course, constitute a proof of the model I am proposing but they hint at the possibility that the mechanisms (especially that involved in linguistic encoding) for producing conscious icons are only activated under certain circumstances.

5.2.4 **Concepts and Categorization**

The mind is able to construct images and categorize percepts based on very limited and often ambiguous sensory input. Any theory of cognitive processing of perceptions must account for how the mind recognizes these incomplete patterns of perceptual input and constructs from them coherent images. The mind generalizes by extracting similarities in input patterns and uses the generalizations (concepts) in higher-level cognitive processes.

While many cognitive psychologists use *concept* to refer to a structured mental representation (e.g., Carey 1991), research psychologists are generally concerned with finding explanations for categories that are reflective of a culture and are encoded in the language prevalent in that culture (e.g., Rosch 1978). The linguistic relativity principle, also known as linguistic determinism or the Sapir–Whorf hypothesis, goes so far as to claim that differences in how languages encode cultural and cognitive categories can affect how people view the world and, hence, affect conceptual content. The difference in approaches (the

[54] See pp:41-42, Illustration 7 on page 30, and section 4.3.1

study of mental representations versus the study of the cultural component of categories) is analogous to the I-language and E-language distinction[55]. Also related to this difference in approaches are the basic principles proposed by Rosch (1973b) to explain the formation of categories: one (internalist) principle is that categorization is a system that provides the maximum information using the least cognitive effort; while another (externalist) is that the information perceived through the senses is already structured—the structure of the information is reflective of the way the perceived world is structured. The second principle assumes that each perceived material object possesses a highly correlated structure. A supporter of the internalist principle would not deny that perceptual information is structured, but claims that categorization of the structures is an internal process.

Clearly, the closer the resulting categories reflect the structure of the *perceived* world (what Kant would call "empirical reality"), the less cognitive effort will be required to obtain maximum information either through mapping categories to attribute structures or by defining attributes in order to structure categories appropriately[56]. Rosch emphasizes that the reference is to the world *as it is perceived*; the kinds of attributes that can be perceived and the way they are categorized are species-specific (Rosch 1978:383). An example of species-specific relationship to the world as-it-is-perceived is that of the ability possessed by rats (but not by humans) to distinguish between distilled normal water and heavy water (also termed 'D_2O' or 'deuterium oxide') based solely on smell and taste. The human perception of water types is not so finely-grained as that of rats; sensory input is limited by the specific perceptual system.

Concepts and Connectionism

In a connectionist model, semantical information is recorded through similarities and differences between activation patterns in the neural network; meaning is fixed via the similarity properties of neural activations. Clark and Elaismith (2002) have pointed out that connectionism "has been criticised for both holding and challenging representational views."

[55]Proposed by Noam Chomsky, "I-language" refers to the mentally-represented linguistic knowledge of a native speaker of a language. Chomsky maintained that this "internal language" is the proper object of study in linguistic theory. E-Language, on the other hand, covers all other notions of what a language is—a body of knowledge or communally-shared behavioural habits, for example—and has no utility for the study of innate linguistic knowledge. (Chomsky 1986)

[56]During the 1980's, developments in artificial intelligence on models of concept learning led to a machine-learning paradigm for unsupervised learning known as *conceptual clustering*: the inherent structure of the data (plus a conjunctive or probabilistic description language) drives cluster formation with the aim of generating a concept description for each generated class (see, for instance, Michalski 1980; Michalski & Stepp 1983).

While most connectionists[57] postulate representational states of some type, there are others, while appearing to reject the existence of representational states, actually reject the standard computational view of mental representations as memory-intensive and all-purpose forms of internal representation, and hold instead to a view of internal representations as sparse and action-oriented forms which exploit stimuli from both the body and the external world to produce a response from which is built the representation itself (see Clark & Eliasmith 2002). On this view, concepts correspond to patterns over large numbers of units.

Connectionism has sometimes been viewed as a return to radical empiricism that would only be of use in discussions of neurological implementation as, for instance, in analyses of how grammar could be realized in the brain (Pinker & Prince, 1988). Certainly some of the early connectionist models seemed to rely on a type of representation based on some form of associationism which had proved to be totally inadequate for the Classical Theory of concepts to provide an explanation of many aspects of cognition such as linguistic knowledge (Fodor & Pylyshyn, 1988; Pinker & Prince, 1988). The view of concepts for the Classical Theory of Concepts is that they are mentally-represented definitions that encode a set of *necessary* and jointly *sufficient* characteristics. Membership in the extension of a concept, on this view, is strictly determinable; concepts are either applicable or not — the relevant category is "unfuzzy". Concepts of this type, namely "unfuzzy", definable concepts, could easily be instantiated in terms of inference rules and sentential representations in a classical symbol-manipulation computer.

There is, nevertheless, empirical evidence that very little, if any knowledge of categories is organized in terms of necessary and sufficient conditions. Knowledge seems to be more organized in terms of clusters of features such that only a sufficient number of features need be satisfied, and with some features considered to be more significant than others – some features are *characteristic* or *core-defining*. The idea that knowledge is organized in terms of features (rather than necessary and sufficient conditions) is basic to the Prototype Theory[58] (and, to a lesser extent, the Exemplar Theory[59]) of concepts and lends itself to implementation in a connectionist network. In a connectionist model, some

[57]As opposed to *eliminativism* for which the appropriate level of psychological theorizing is neurological and the semantic notion of representation is, therefore, not required.

[58] See Rosch (1973a) for a description of Prototype Theory.

[59] See Medin & Schaffer (1978) and Nosofsky (1986). Exemplar Theory was used in psychology as a model of perception and categorization, and was later applied to speech perception (Pierrehumbert 2001). It and has also been used by linguists such as Johnson (1997) and Goldinger (1996, 1997).

features would have higher connection *weights* than others. On the connectionist view, instances cluster around the "prototype" which is the central region in a hyper-dimensional semantic space[60]. Many philosophers and cognitive psychologists have argued that concepts and categories may be delimited through some kind of family resemblance or similarity to a prototype (as illustrated in Illustration 14). A graded notion of category membership, such as that proposed by the Prototype Theory, is well suited to implementation in connectionist models given that neural nets are capable of "learning" the difference between subtle statistical patterns that would be very hard to implement in rule-based classical models.

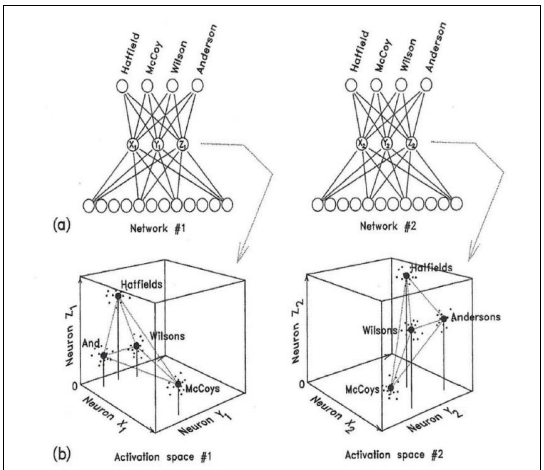

Illustration 14: Churchland 2007 - Fig 8-2

When nets of different structures are trained on the same task, they develop activation patterns which are strongly similar, thereby suggesting that it might be possible to produce empirically well-defined measures of similarity of concepts and thoughts between different individuals (Churchland 1989). It is possible to measure the degree of similarity between two conceptual frameworks, as Illustration 14 is purported to show[61] (Churchland 2007). Fodor and Lepore (1999) have pointed out, there are problems with developing a standard theory of meaning based on similarity: such a theory would need to be able to assign truth conditions to sentences based on the meaning of their constituents, but similarity alone, they claim, does not seem to be sufficient to fix denotation as would be required by any standard theory[62]. It should be emphasized, however, that many of the presuppositions of standard theories are rejected by most connectionists who advance similarity-based accounts of meaning. Such connectionists are attempting to develop an alternative theory of meaning which may reject or modify the presuppositions of a standard theory of meaning but which is still consistent with what is known about human linguistic

[60] This differs from the Exemplar Theory which allows only for actual instances to be stored in memory.

[61] In this illustration (Churchland 2007:Fig 8.2), the top part of the diagram, labelled '(a)', shows two distinct networks which have been trained to discriminate photographs of faces belonging to one of four families; and the bottom diagram, labelled '(b)', shows the two resultant activation spaces of the respective middle layers of the two networks. Each network has acquired a structured family of "prototypical" family regions, within which facial inputs from each of the four families typically produce an activation pattern.

[62] See the section "Connectionism and Concept Individuation" on Page 116.

abilities. Unlike the unstructured, mutually-independent concepts of informational atomism (IA), all concepts, for connectionists, have complex internal structure with no automatic referential connection to the external world. Furthermore, connectionist concepts are not *acquired* but are *learned*; connections are formed and re-enforced through cognitive processes. Changing the processing or knowledge structure in a connectionist system involves modifying the patterns of interconnectivity. This can involve three kinds of modification: development of new connections; loss of existing connections; or modification of the strengths of connections that already exist. A network's knowledge is *stored in the strengths of its connections*.

Connectionism and Innate Concepts

Computational cognitive science tends toward the view that the mind is computational; that it is composed of distinct modules each of which specializes in the processing of distinct types of information, has specialized functions, and is informationally and functionally encapsulated (Chomsky, 1980; Fodor, 1983). In contrast to the cognitivist position of evolutionary psychologists and many others who support at least some form of the modular mind hypothesis, connectionists tend to be non-cognitivist and anti-innatist, holding instead that only a general capacity for learning is genetically inherited and that all cognitive capacities are the result of learning and experience. The mind functions as a pattern recognition system. That is to say, there are perceptual processes that match limited sensory input to existing "patterns" in memory – a flash of yellow and black in a jungle setting is likely to be matched rapidly to a 'tiger' concept as long as such a concept had previously been acquired or learnt. In order for the pattern recognition system to perform initially, however, there would have to be some innate patterns – such as the sensory concepts proposed by British Empiricists[63]. Almost all theories of concepts support innateness in this regard; they differ in how many and of what type these concepts might be. The idea that there may be innate concepts is not only applicable to humans. There is some evidence of innate concepts in non-linguistic species, a famous example of which is the Tinbergen/Lorentz hypothesis that some bird species have an innate recognition of the shapes of birds of prey. Lorenz and Tinbergen used, in their 1937 experiments, a "hawk/goose dummy", as rendered in Illustration 15, which, when flown to the right (as a "hawk") elicited escape behaviour ("fixating, alarm calling and marching off to cover") in

[63]The British Empiricists, such as John Locke (1690/1975), held that human concepts are grounded in a set of innate primitive ideas in terms of which all complex concepts are defined. Concept acquisition is explained by the combination of a subset of these innate primitives with associative mechanisms by means of which complex concepts are constructed.

young, experimentally-naïve turkeys, whereas no escape behaviour was elicited, when the model was flown to the left (as a "goose") (Lorenz, 1939). There are non-human species (some primates, dolphins, and elephants) that are able to recognize themselves in a mirror demonstrating that they have at least some rudimentary concept of self.

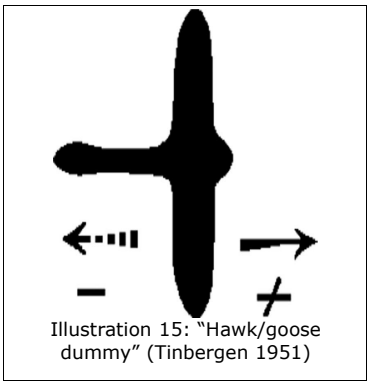

Illustration 15: "Hawk/goose dummy" (Tinbergen 1951)

From a connectionist perspective, neural patterns are generally of three main types:

- Fused ("hard-wired") neural networks: innate concepts (e.g., instincts) and encapsulated modules might be of this type;
- Heavily weighted neural patterns which would include: innate capacities that need triggering by relevant input such as natural language and bird-song; "acquired" concepts, and those biases and beliefs that are not available to conscious awareness; and
- New patterns created by sufficient sensory input, including: "learned concepts" and beliefs and biases amenable to encoding as images in consciousness.

Connectionism and Concept Individuation

According to connectionist models, semantical information is recorded through similarities and differences between activation patterns in a network such that *meaning* is fixed through the similarity properties of neural activations. As previously stated, Churchland (1989) claims that, when networks of different structures are trained on the same task, they develop activation patterns which are strongly similar, thereby suggesting that it might be possible to produce empirically well-defined measures of similarity of concepts and thoughts between different individuals. Nevertheless, semantic identity is not based on causal connections to the external world; it is based on an *internalist* account of sameness and similarity (Churchland 2007:134). The brain of each person has a unique configuration of synaptic weights and connections which is a result of individual biology and, especially, of lived experience. Each person's empirical reality (the reality of one's experiences) may nevertheless reflect enduring features (e.g., natural kinds) in the external world which may result in similarities in the structure of families of activation spaces in different individuals (Churchland 2007:34). Churchland (1995) has referred to mappings between referents and to individuation of certain types of computational states in terms of their role in a

connectionist network. Rather than having a semantic and syntactic structure similar to that of a natural language, as some other theories of concepts posit, internal representations are sub-patterns that include the micro-features that are specific to the context. For instance:

> "The coffee in the cup" would...have a subpattern that stands for "coffee." But that subpattern will be heavily dependent on context and will involve microfeatures that are specific to the in-the-cup context (Clark 1989:148).

Variations between and within groups of language users could be implemented in this manner. Zhang, Segalowitz, and Gatbonton have reported systematic differences in the use of linguistic expressions by Chinese (Mandarin) and English speakers for certain topological spatial relationships: for example, where an English speaker would say "the bird is *in* the tree", a Mandarin speaker would say the equivalent of the English "the bird is *on* the tree" (Zhang, Segalowitz, & Gatbonton 2010). The sub-pattern that stands for "tree" in the mind of an English-speaker would have micro-features specific to the "in-the-tree" context, and in the mind of the Mandarin-speaker micro-features specific to the "on-the-tree" context. From a connectionist perspective, there is not an equivalence *per se* between the the Mandarin preposition translated as "on" in English and the English "on" – in fact, the prepositions themselves may not be represented at all. Systematic differences such as these may be, at least in part, responsible for the development of the linguistic relativity principle (or Sapir–Whorf hypothesis) which holds that differences in the way languages encode cultural and cognitive categories affect thought processes. From a connectionist perspective, however, the reverse is the case: cultural and cognitive categories affect the way a concept is acquired, which results in differences in micro-features of the internal representation, and these differences are reflected in the natural language surface structure.

Computational states individuated by their role in a connectionist network could thus be able to differentiate MORNING STAR from EVENING STAR (Frege's famous example of co-extensive concepts) by having two subpatterns that stand for STAR: one would include the micro-feature specific to the "in-the morning" context and the other would include the micro-feature specific to the "in-the-evening" context. In fact, Frege's Puzzle, which has been discussed at length in the philosophical literature, does not exist for connectionist models. By treating internal representations as sub-patterns that include the micro-features that are specific to the context, *a* connectionist model could also explain how two different referents may result in the same activation pattern if the input stimuli are similar enough; it is therefore very likely that a connectionist network would treat jadeite and nephrite as the same,

namely as jade. An expert could "learn" to discriminate between jadeite and nephrite and, consequently, would develop two distinct activation patterns, a JADEITE prototype and a NEPHRITE prototype. We can restrict the use of a term by stipulation but require *learning* to restrict application of the concept; to refine conceptual content.

5.3 Part II: Cognitive Attributes

In this section I attempt to demonstrate that the proposed cognitive architecture has the attributes which are associated with standard views of the differences between the two systems of a dual-process theory.

5.3.1 Conscious and Unconscious Processes

As discussed in the previous chapter, experiments using electroencephalography performed by H. H. Kornhuber and I. Deeke (1964), Libet (1999), Wegner (2002, 2003), Soon et al. (2008), and many others, support the claim that the volitional process is initiated unconsciously; that awareness is a unique phenomenon in itself which is distinguished from the contents of awareness; and that there is no difference between the experience of having personally caused an action and the experience of cause and effect in general. These experiments lend support to the view that there are two related cognitive processes:

- in one process, certain unconscious mental events cause a thought while other unconscious mental events cause actions; and
- in another process, the apparent (but not real) link between cause and action is experienced consciously.

There is a considerable difference between discussing what we experience consciously and what causes the experience, and, as many neuroscientific studies suggest, conscious experiences are causally inert. The results of these studies negate the distinction between controlled (conscious) processing and automatic (unconscious) processing. The "inhibitory" function usually attributed to controlled (conscious) processing is, I claim, merely the monitoring and verbalization of a subconscious "conflict" resolution resulting from much the same mechanism as an alarm call would in other species.

5.3.2 Rational and Intuitive Reasoning

For most dual-process theorists, reasoning is a combination of fast, heuristic processes (of **System 1**) which may provide contextualized content as input to a conscious, analytic

process (of **System 2**); thus, **System 2** provides normatively correct reasoning, whereas **System 1** processes often result in cognitive biases. This division is highly questionable, especially in the case of expert reasoning and judgements. As the results of de Groot's 1965 study (discussed previously, in more detail, in Chapter 4) demonstrate, expert performance is a question of (non-conscious) pattern-matching rather than conscious rule-following. Gary Klein (1999) has termed this expert performance "recognition-primed decision making." After studying the decision making of groups of firemen and paramedics, he concluded that experts typically:

- Recognize a situation as of the same type as one previously encountered; and then
- Rapidly retrieve a schema which provides a solution.

For Klein, the most cogent process in intelligent action is *automatic retrieval*. Non-conscious pattern-matching is very simply modelled by a connectionist network and is, in fact, crucial to the way connectionism handles prototypes.

The Wason Selection Task has been widely used to study conditional reasoning (see a detailed review in Evans & Over, 2004, Chapter 5), Many theories have been postulated to explain the results of the studies: for instance, Leda Cosmides and John Tooby (1987) state that the results of the Wason Selection Task justify their claim of a so-called "Cheater Detection Module", suggesting that humans have no single "reasoning faculty" and that the brains of *hominids* have evolved a cluster of cognitive adaptations to deal with social interactions. Wason (1966) originally proposed that the errors were the result of a *confirmation bias:* the subjects were trying to prove the conditional true. More recent studies (see Evans & Over 2004) suggest that errors are actually caused by a *matching bias*: subjects tend to select the options matching the entities explicitly given in the rule regardless of the logical appropriateness of doing so. The underlying causes of matching bias may reflect a general phenomenon which applies to many other types of logical rules in addition to conditionals (Roberts 2002). The Wason Selection Task has been modelled by different connectionist systems with some success. In particular, BioSLIE (BIOlogically-plausible Structure-sensitive Learning Inference Engine), which is a low-level, spiking neuron model of a high-level cognitive behaviour, has been applied to the Wason Selection Task (Eliasmith 2005). Chris Eliasmith reports that a detailed computational model, such as BioSLIE, demonstrates how a domain general mechanism can account for the observed

cognitive behaviour of the subjects, and, further, that each subject can solve the problem idiosyncratically rather than in accordance with a pre-specified, discrete set of "schemas":

> At the behavioral level, the model not only meets Cosmides' challenge of specifying an inductive, domain general inference mechanism, it also makes it possible to predict behavioral variations on the task given a learning history. For instance, it should be possible to predict the effects of varying the kind of feedback that a subject receives in similar and dissimilar contexts. (Eliasmith 2005)

A "matching bias" is highly consistent with such a model.

5.3.3 **Preference and Inference**

The division by dual-process theorists between two systems of processes, one of which often results in cognitive biases (**System 1**) and the other which provides normatively correct reasoning (**System 2**), is similar to that made by many contemporary psychologists who distinguish between the processes of *preference* and *inference*. The traditional psychological view, is not, however, the either/or view represented by dual-process theorists, but that both preference and inference are part of cognitive processing with preference (affective reaction) following after inference (analytic computation). As reported by R. B. Zajonc (1980), the traditional view is that *affect* is post-cognitive, that cognition precedes evaluation. Zajonc argues strongly against this view, reversing the order so that preference precedes inference whenever the information being processed has affective qualities. His points out that:

> It is much less important for us to know whether someone has just said "You are a friend" or "You are a fiend" than to know whether it was spoken in contempt or with affection. (Zajonc 1980)

The intention of a someone speaking ironically or sarcastically is conveyed by context and by the intonation of the speaker rather that by the symbols and syntax he used. Several studies (such as those by Dawes and Kramer, 1966; and Scherer, Kolvumaki, and Rosenthal, 1972) demonstrate that we can reliably encode the emotions expressed in an utterance even when the content of those utterances are almost completely obliterated by electronic "noise" or random splicing, or even when the utterance is in a language unknown to us. We evolved to make rapid judgements with very limited perceptual input for, as Kant observed, rational processing would be too slow and unreliable in dangerous situations that are very likely to arise in natural environments (Kant 1785/1956:395). We are more likely to survive

if our experience of a flash of black and yellow in a jungle results in a fear and flight response even if we are never sure what caused the sighting. Even in normal, everyday situations, most of our perceptions have at least some emotional component: "affect is *always* present as a companion to thought, whereas the converse is not true for cognition. In fact, it is entirely possible that the very first stage of the organism's reaction to stimuli and the very first elements in retrieval are affective" (Zajonc 1980). For a connectionist, this claim is not surprising as the weighting of neural connections are likely to be stronger for emotional components. Even though Zajonc's position is that preference precedes inference, he does note that all cognitions are accompanied by the form of experience generally termed "feeling" and that feeling derives from a system that is parallel and separate from that responsible for inference. His view is thus consistent with a connectionist model in which processing is massively parallel and holistic; it is merely a question of which processing, preference or inference, dominates in a particular situation, of which neural pathway has the higher weighting. Psychotropic drugs[64] are often used to "excite" one pathway at the expense of another, acting as a sort of "counter-irritant" and behavioural therapy works in a similar way.

5.3.4 **Natural Language and Reasoning**

There are two basic assumptions behind most dual-process or dual-system theories such as that proposed by Kahneman: first, that logical reasoning is conscious, that rational judgement, in general, takes place consciously, and, secondly, that such reasoning is associated with language. Several studies have demonstrated that pre-linguistic preschoolers demonstrate deductive reasoning skills (see Dias & Harris, 1988, 1990; Hawkins, Pea, Glick, & Scribner, 1984; Richards & Sanderson, 1999). On the Piaget account of sensori-motor development, symbolic problem solving is manifest in infants from 18 to 24 months of age.

A proponent of the view that the brain is a symbol-manipulation device (the Symbol System Hypothesis), tends to claim that sentences (and linguistic symbols in general) provide a model of how a fact can be represented, and, further, that the brain must contain representational states. It is natural to theorize that these states literally *are* sentences (or sentence-like), that representations are sentence-like strings of symbols. But what does it mean to say that the brain literally contains sentences? According to the Symbol System

[64]Psychotropic drugs are chemical substances that influence brain function, thereby altering perception, cognition, mood, and behaviour.

Hypothesis (SSH) symbols are realized in the brain by some sort of electrical or chemical activity. How they are actually realized in the brain is an empirical matter. When we think, do we always do so in a language? SSH supporters do not claim that cognition occurs in a natural language, such as English, even though thinking in a natural language may seem to be what is happening at a conscious level. Below consciousness (and below the level of mental imagery) the language of the brain cannot be a spoken language. After all, infants who have not yet acquired a native language can nevertheless form mental representations and there is evidence from the behaviour of many non-linguistic species that they, too, form mental representations. It is claimed that the language of thought is some kind of innate brain code, sometimes called 'Mentalese' (discussed previously, in more detail, in paragraph 2.6.2). On the SSH view, cognitive processes invoke explicit rules operating on syntactically structured representations.

A contrary view is held by connectionists for whom these processes should be understood instead in terms of the activation of nodes or patterns of nodes in neural nets that are not governed by explicit rules nor have any syntactic structure. Perceptual pattern recognition, for instance, cannot be modelled by a symbol-manipulation, Language of Thought (LOT) system – Fodor (1983) proposed a combination of specialized transducers and encapsulated modules to perform such cognitive functions. There is much empirical evidence showing that the brain, when physically damaged, exhibits graceful degradation; connectionist systems exhibit a similar response to physical damage but such is not the case for symbol-manipulation, LOT systems. Much of the input we receive is "noisy" or incomplete but, unlike symbol-manipulation, LOT systems, degraded input can be managed by the pattern recognition power of connectionist systems. Further, as Fodor (2000:47) himself has recognized, a major failure of computational cognitive psychology is its inability to provide a convincing account of abduction, an account that connectionism can provide. Thus, connectionism is more biologically plausible than symbol-manipulation systems.

Support for the connectionist position are provided by recent fMRI experiments[65] which strongly suggest that logical reasoning does not rely on the grammar of natural language and, along with other neuroimaging and neuropsychology evidence, appear to show that high-level cognitive functions, such as arithmetic, problem-solving, and theory of mind, are remarkably language-independent. That our awareness of rule-following is encoded in the grammar of a natural language cannot be used as a justification for claiming

[65]As reported by Martin M. Monti at the *Interdisciplinary Workshop on the Notion of Thought,* Ruhr-Universität Bochum, 5th - 7th June 2008

that logical reasoning itself relies on the grammar of a natural language. The appearance that it is so reliant is the result of reverse engineering: the form of the reasoning as it appears to consciousness is taken to be the form of the reasoning processes. But this reverse-engineering is no more justified than using it to claim that the structure of the output from a digital computer is the structure of its internal processes.

5.3.5 **Consciousness and Volition**

Almost all dual-processing theorists claim v*olition* to be a cognitive correlate of consciousness or, at least, that the associated attributes are of *Type 2*. Consequently, it is reasonable to question how it is possible to have free-will if all thought processes are unconscious as I am proposing. In response to such a question, I would ask why free-will has to be a *conscious* process and why satisfying the metaphysical requirement of being responsible for one's actions would necessarily have to be controlled *consciously*. Whether we accept that free will is merely the ability by which we select a course of action in order to fulfil some desire or that we require the selection of a course of action based on our desires and values be deliberative, we are still not committed to the selection process itself being controlled by the conscious mind. As discussed in the preceding chapter, there is considerable empirical evidence supporting the view that the conscious exercise of free will is an illusion.

Many neuroscientific experiments, most notably those performed by Libet, Wegner, and others cited in the preceding chapter, support objections to the traditional view of free will. In particular, these experiments demonstrate that our brains prepare to act before we are conscious of the decision to act. If such is the case, then we cannot be making a *conscious* decision to act. New studies by Jesse Bengson, *et al.,* appear to show that "brain noise" might actually create a possibility of “free will” in so far as it "inserts a random effect that allows us to be freed from simple cause and effect":

> This finding provides evidence for a mechanistic account of decision-making by demonstrating that ongoing neural activity biases voluntary decisions about where to attend within a given moment. (Bengson et al 2014)

That “free will” is just the result of random effects would not satisfy those believers in a free-will for which consciousness is causally-efficacious. Bengson's results merely provide a means for explaining why we are not always able to recreate the reasoning behind an apparent act of free-will. In any case, the effect of “brain noise” is unconscious and random

neural firings can more easily be explained using the connectionist model than the symbol manipulation model of cognition.

5.3.6 **Consciousness and Natural Language**

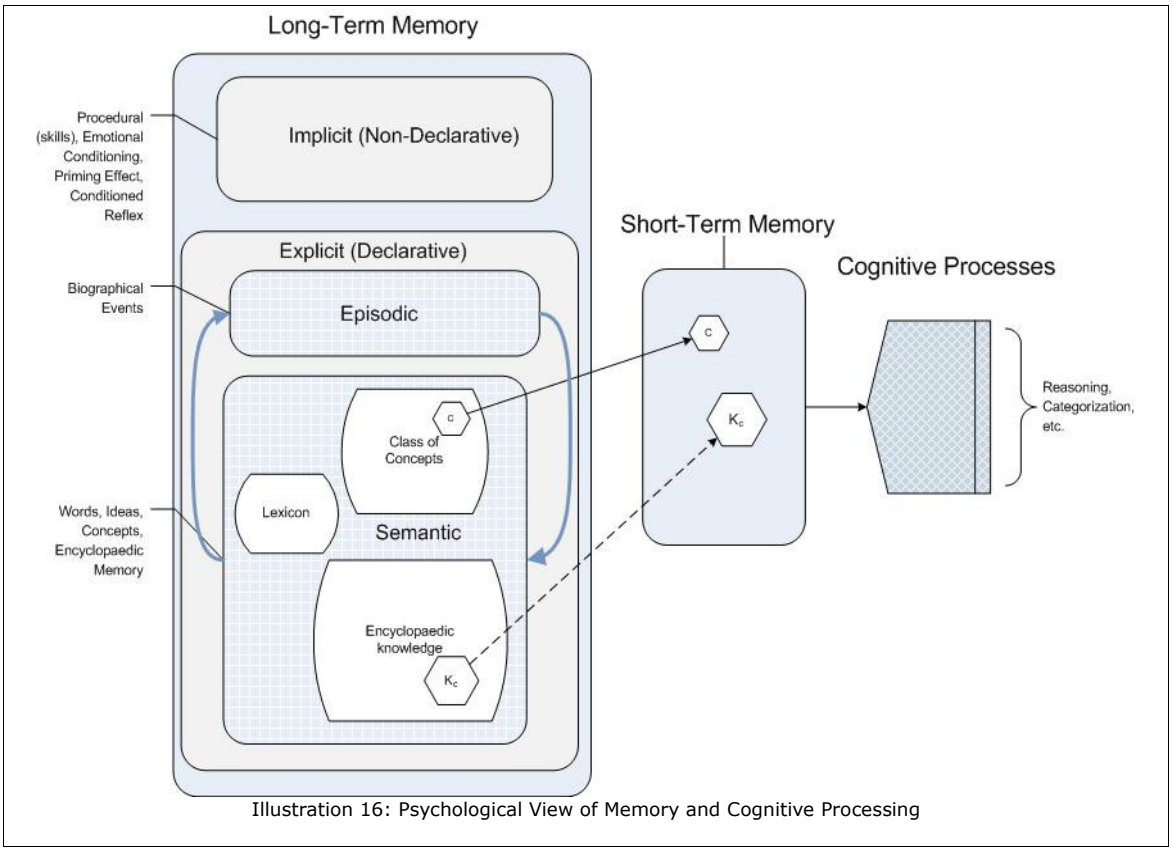

Illustration 16: Psychological View of Memory and Cognitive Processing

On the psychological view of the relationship between memory and cognitive processing (Illustration 16 above), relevant cognitive processes retrieve the same body of knowledge, the same mental representation of an entity, from long-term memory whether we categorize the entity, draw an inductive inference about the extension it belongs to, draw an analogy between it and some other entity or entities, or even understand a sentence in which a term referring to the entity appears. In other words, a concept is moved from a class of concepts stored in semantic memory to short-term memory, perhaps along with other non-conceptual knowledge of the item in question, for use in a cognitive process (e.g., deduction, induction, categorization). To summarize, according to this view concepts are characterized as bodies of knowledge stored in long-term memory.

Using Chomsky's theory of a Universal Grammar (Chomsky 2006:124) as a basis, I support the position that language-related (usually speech) input is "decoded" by being

input to the Universal Grammar (UG) which takes the linguistic representation, transforms it to the surface structure, and then to the deep structure which, along with a mapping to a semantic structure (using semantic memory which consists of the class of concepts, the lexicon, and encyclopaedic knowledge, and is found in long-term memory as shown in Illustration 16 above), and becomes input to the cognitive systems. The *encoder* does the reverse: the internal deep structure is transformed to the surface structure which, in turn, is converted, by the phonological component, to the phonic representation, and, if the relevant semantic information exists, the deep structure is also mapped to a semantic representation. If the UG is able to assign a semantic representation to the input, then the thought is experienced as *meaningful*. A highly simplified diagram of this process is given in the following diagram (Illustration 17). The link to external behaviour is not shown.

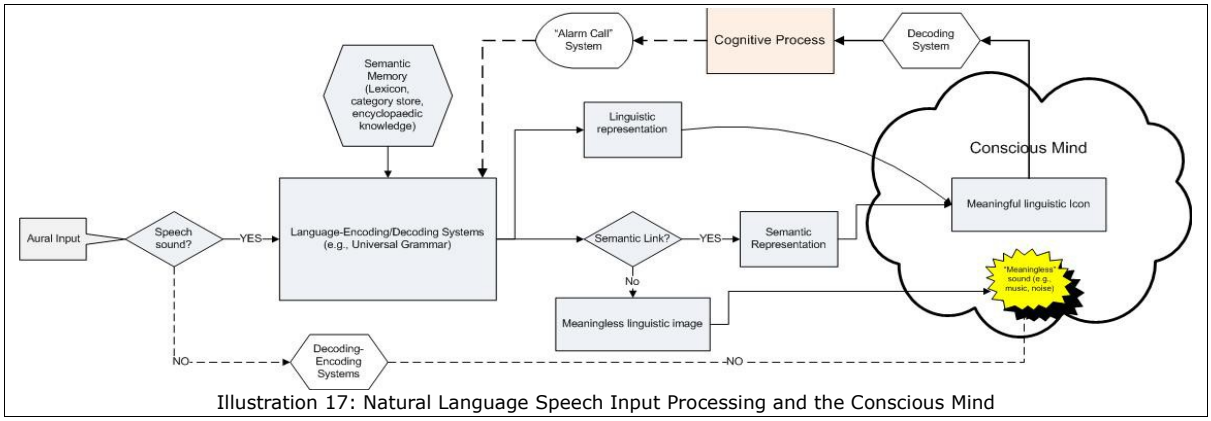

Illustration 17: Natural Language Speech Input Processing and the Conscious Mind

In the situation given in the preceding diagram, the *form* of the experience is a linguistic "speech" icon. As also proposed by Jackendoff (1990, 2011), I contend that the linguistic icon is linked to a thought; it provides a *form* of the experience; and is a cognitive correlate of consciousness. Semantic representations are encoded as icons with a "meaningful" *feature*. It should be noted, that neither the semantic representations nor speech representations themselves are cognitive correlates of consciousness. The form of the experience as it appears to the conscious mind depends on the input modality; it may be tagged with a type of "external" feature for aural input as well as being linked to one or more phenomenal features, such as "meaningful", all of which contribute to the conscious experience. Internal speech follows a similar process except for being tagged with an "internal" feature, although this feature may be absent or replaced with a false "external" feature for people suffering from delusional disorders like schizophrenia. The "internal" feature is usually absent during the dream state – the exception is the rare case of "lucid" dreaming. Of course, some people might deny that the conscious mind is active at all during

non-lucid dreaming, but such a denial would be based on a conflation of consciousness and awareness. I would suggest that our conscious mind *is* active during dreaming and the differences between the dreaming and awake states are all at the unconscious level. Dreams are often defined as "vivid, sensorimotor hallucinations with a narrative structure" which are experienced consciously. Consciousness during dreams is unlike wakeful consciousness given the usual absence of the ability to introspect except during lucid dreaming. During sleep, many unconscious processes cease temporarily, or mental states cease temporarily to be in the same causal relations as they are in the awake state. Nevertheless, some of these mental states may still be encoded in a form that is a cognitive correlate of consciousness. That is to say that 'consciousness' and 'awareness' are not synonymous. Being aware of a conscious icon is a different process from having an icon in consciousness. *Becoming aware* requires two (unconscious) synergistic processes (one that produces the encoded icon of the thought of which one is "aware" and one that produces the linguistic or a non-linguistic icon produced by the monitoring process). In the awake state, it possible to be *aware* of being conscious, but, in the dream state, it (usually) is not.

In the dream state, the linguistic encoding and other encoding processes are highly active without any external input. Clearly, aural input is not the only initiator of the linguistic encoding process. For advocates of the "Language of Thought Hypothesis" (Fodor 1975), the type of cognitive processes that would be amenable to the linguistic encoding of internal speech are those conducted (non-consciously) in a "language of thought". The Cluster 1 "features" of System 2, listed in the table on page 103, may provide some indication of which thought processes are amenable to such encoding given that the types of representations considered to belong to **System 2** appear to have have a combinatorial syntax and semantics and thus to belong to a representational or symbolic system (c/f. Fodor and Pylyshyn 1988:12–13). It should be emphasized that speech encoding does not always occur even when the input is aural. We are not always conscious of sensory input yet the input may be registered in semantic memory and later linked to a thought process. There have been several experiments (see, e.g., Allport 1977; Maki et al. 1997; Besner & Stolz 1999) that indicate that printed words do not require attention or task relevance for their meanings to be accessible; semantic access occurs for printed words even when spatial attention, task relevance, or awareness for these words are lacking. Semantic representations are still activated even when the words are presented peripherally and no attention is drawn to them. These results are similar to those for dichotic listening

experiments which demonstrate that it is possible to extract meaningful content from irrelevant or unattended inputs depending on the strength of distracting stimuli (see, for example, Corteen & Dunn 1974; Rees *et al*. 1999).

On the construal of consciousness that I am proposing, *qualia* ("the introspectively accessible, phenomenal aspects of our mental lives") are features of mental states which may, or may not be encoded. Without this encoding, we are unaware of pain, for instance, even though the nervous system acts in response to the stimuli that caused the mental states. The case of an athlete who is not conscious of an injury while performing but becomes conscious of the injury when the performance is over can be explained by other mental processes taking priority over the encoding process during the performance. When discussing concepts like *qualia*, we have to be careful not to commit what U. T. Place (1956) called the "phenomenological fallacy" which is confusing properties of objects experienced with properties of experiences. The encoded icons of mental states with phenomenal features (*qualia*) are not the mental states themselves but are the haptic icons of those mental states. Just as internal speech aids in memory consolidation through a "feedback loop" by which experiences are encoded in speech icons[66] and fed back into the unconscious thought process, *qualia*, such as the feeling of pain, unfortunately, may also be "re-enforced" through such a feedback loop.

Also on this construal of consciousness, pre-linguistic children and non-linguistic species are not precluded from having a conscious mind. Conscious imagery is not limited to linguistic icons. Nor, as Fodor himself has suggested, is the language of thought limited to humans: some species may have a proto-LOT even if encoding in a natural language form is limited to humans (according to current evidence, at least). The claim that almost all the features attributed to *Type 2* by most dual-process/dual-system theorists are actually features of the non-conscious (*Type 1*) does not detract from the claim that there is a considerable evolutionary advantage to being able to encode thought processes linguistically whether or not the results also appear as icons in the conscious mind: not only are we individually aware of many of our thought processes but we are able to communicate them to our conspecifics; and the encoding of the output of a thought process in a form that can be used internally as input to cognitive processes improves learning and abstract thought.

[66]These proposals also apply to sign-language in which case the icons are haptic or, perhaps, visual.

5.3.7 **Implicit and Explicit Learning**

As discussed in Section 4.3.3 "Implicit and Explicit Learning", the distinction between implicit and explicit learning is predominantly a distinction between non-conscious and conscious learning, of non-conscious knowledge and concept acquisition and conscious rule-following. As previously noted, it is possible to acquire implicit knowledge without ever being able to state any related explicit rule (see Reber 1993, Sun et al. 2005) but, while implicit processing proceeds by chance without accompanying awareness, explicit processing proceeds deliberately and is always accompanied by awareness (Reber 1993). For connectionist models of learning, there are two types of feedback: external and internal. In either case, if there is an internal measure of error, the learning is called "monitored". There are, usually, internal consistency/coherence measures which can be used in the improvement of internal representations and which can be internally monitored; "error detected by a monitor in one part of the nervous system is a plausible teaching signal for another part of the nervous system" (Churchland & Sejnowski 1992/1999:97). I postulate that "conscious rule-following" is actually "being conscious of following rules". It should be noted that connectionist models are able to produce rule-following *behaviour*. Difficulties encountered at the non-conscious level, especially by novices, result in activation of the encoding (or "monitoring") systems, producing (usually linguistically-encoded) icons in the conscious mind and, hence, in an impression of rule-following. Because the same type of difficulties are less likely to arise in the non-consciousness of experts, their monitoring systems are rarely activated. The following diagram illustrates the position of the conscious mind in a learning process.

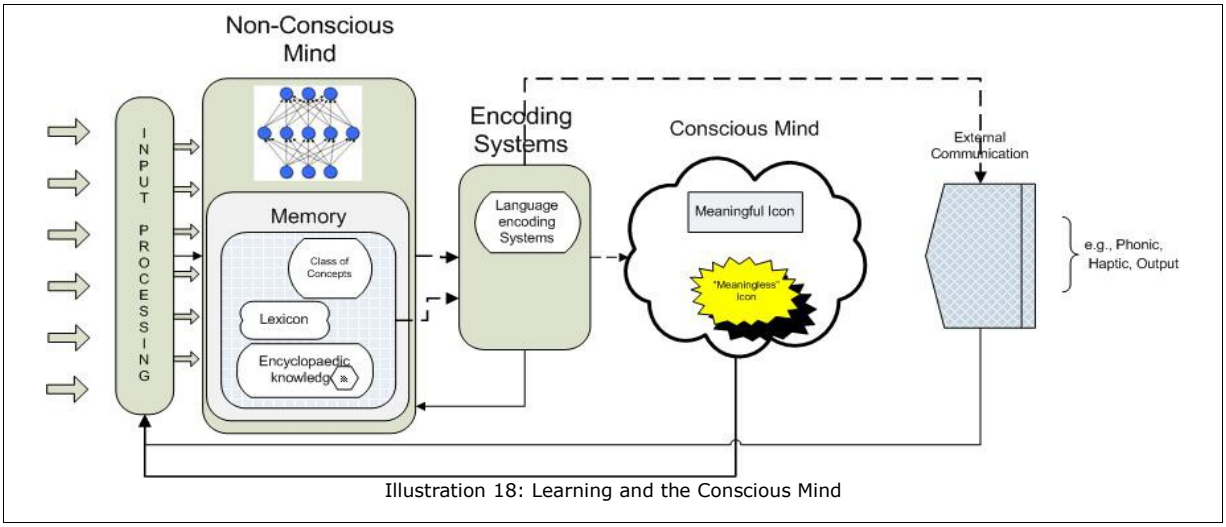

Illustration 18: Learning and the Conscious Mind

"Feedback" is part of a normal learning process and avoidance behaviour is a normal response to an unpleasant experience. There is strong evidence that even crustaceans such as shore crabs feel pain and learn to avoid causes of the pain (Magee & Elwood 2013). The experience of pain is part of avoidance learning: animals, including humans, learn from pain; they do not just continue to respond to a stimulus. Stress of any kind can be part of a learning experience. An emotional response occurring in response to stress, whether from an environmental or psychological cause, results in a sequence of hormonal changes and physiological responses. I contend that activation of an emotional response to a subconscious reasoning "conflict" may result in the linguistic encoding of parts of the reasoning process that are amenable to symbolic and syntactical expression. Whether or not the learning processes are encoded as icons in the conscious mind accounts, at least in part, for the apparent division between implicit and explicit learning.

5.3.8 **Evolutionarily Old versus Evolutionarily Recent**

Until recently, scientists believed that the more developed prefrontal cortex in humans explained humans' unique abilities for planning and abstract reasoning, but further studies (see Dunbar 2001) have questioned this belief. Recent studies have used magnetic resonance imaging to measure the relative size of the prefrontal cortex in all species of great apes (chimpanzees, bonobos, gorillas, and orangutans) as well as humans. The results show that the relative size of the prefrontal cortex was almost the same in humans as in the great apes. The superior abilities of humans to anticipate and to plan has, in more recent studies, been attributed to other specialized regions of the cortex and also to the fact that the larger volume of white matter (composed of myelin-covered axons) in the human prefrontal cortex provides greater connectivity between the prefrontal cortex and the rest of the brain (Smaers et al. 2010). The type, number, and complexity of the neural connections in different species are a matter of degree and hence there is no justification for distinguishing between evolutionarily old and evolutionary recent areas of the brain, nor, in particular, for dividing types of cognition based on such areas. The claim that there are two cognitive system types, where *Type 1* cognition operates rapidly and unconsciously and *Type 2* operates slowly and consciously using the medium of language, cannot be supported by proposing that two different areas of the brain have distinct histories of evolution. That there are specialized regions of the cortex, for example, can be justified but these are still neural networks distinguished by the number and type of connections. There is no area of the brain that is uniquely human.

5.3.9 **Modularity and Dual-Process Theories**

The inability of the symbol-manipulation system of human cognition to model all types of cognitive functions – of which perceptual pattern recognition is a prime example – has led to proposals for models that include multiple different cognitive processing systems, ranging from dual-process and dual-system models to the massively modular, and various combinations of these. In principle, there is nothing about dual-process theories that precludes modularity; the modules could be of just two types: for instance transducers and encapsulated, domain-specific modules for input processing, along with central, domain-general, unencapsulated systems (see Fodor 1983). The proliferation of views about different types of processing systems is, however, indicative of basic problems with viewing the mind as a symbol-manipulation computer. Classical, symbolic-processing systems are too slow to correctly model reflexive and direct reasoning processes and such systems cannot account for the contextual, parallel processing required for non-local reasoning.

In addition, attempts by Artificial Intelligence researchers to model symbol manipulation have often resulted in a version of the *Frame Problem* because of the difficulty of drawing on the relevant background knowledge – accessing relevant information among a vast store of linguistically-encoded structures is highly problematic. The frame problem cannot be avoided by a massively modular mind, despite what supporters of the Massively Modular Mind Hypothesis (MMM) might claim, because there would have to be some mechanism whereby input to these encapsulated, domain-specific modules is limited to relevant information only (Fodor 2001), and no such mechanism, consistent with a symbol-manipulation model, has been found. Even though some Artificial Intelligence researchers have developed a variety of adequate, but not necessarily biologically-plausible, means for avoiding the frame problem, the epistemological frame problem remains for philosophers of mind: Dennett (1978:125) asks how "a cognitive creature … with many beliefs about the world" updates those beliefs when it performs an act so that they remain "roughly faithful to the world"; and Cosmides and Tooby, strong supporters of massive modularity of mind, note that the more a cognitive creature inferentially integrates information, the more it risks error propagation (Cosmides & Tooby, 2000). Connectionist systems, on the other hand, are not subject to the frame problem because the received information results in an almost instantaneous activation of a prototype and, in consequence, a cognitive system is able to access the relevant consequences of a change in the environment almost immediately (Churchland 1989:178).

5.3.10 Cognitive Correlates of Consciousness

Many dual-process theories identify, albeit implicitly, several cognitive attributes that appear to be correlates of consciousness[67]. These include: conscious reasoning; judgements based on critical examination; reasoning that requires effort and control; logical reasoning, inhibition used when **System 1** fails to form a logical/acceptable conclusion; intentional thinking; analytic and reflective reasoning; and so on. An examination of each of these leads to the conclusion that each of these may become encoded in the conscious mind but that each of these could be performed without conscious awareness at all. Experts are rarely able to express the reasoning processes they use; expert judgements are most often made non-consciously with only the results appearing to consciousness. Most people have experienced reaching a solution to a problem without being at all aware of how the solution was reached. When we catch a ball, or cross a road in heavy traffic, we are making highly complex calculations but would fail miserably if we had to be conscious of the calculations. It is unlikely that, without advanced mathematical knowledge, we could even state what calculations have to made when judging the point at which we should catch the ball or what distance from oncoming cars and what speed we should walk to make it safely to the other side of the road. These calculations are performed non-consciously; although the result of failure in our calculations will be encoded in consciousness!

I suggest that the only cognitive correlates of consciousness are the encoded icons which appear to the conscious mind. When some thoughts, phonetic, haptic, and aural images are encoded in a form for input to consciousness, then these are cognitive correlates of consciousness for, without them, there would be no consciousness at all. The same cannot be said for the correlates of consciousness associated with the attributes identified in *Cluster 1* of **System 2**, even if problems encountered during a reasoning process may well trigger the encoding of the thoughts in a form for input to the conscious mind.

5.4 PART III: POSSIBLE CRITICISMS

Chapter 3 of this dissertation contains a comparison of the strengths and weaknesses of the Classical (aka "computational", "symbol-manipulation") and connectionist cognitive architectures[68]. In this section, I will concentrate on criticisms related to connectionism as a *theory of mind* rather than as a cognitive architecture implementation model.

[67]See the table "Dual-Process Theory – Differences between System 1 and System 2" in Section 4.2 of the preceding Chapter.

[68] See the table on page 81.

5.4.1 **Connectionism and Computation over Symbols**

Fodor (1975) presented a major argument for the Language of Thought Hypothesis (LOTH) which was to the effect that a computational/representational medium is required for our best models of higher cognition to be true. He considered three types of cognitive phenomena: perception as the fixation of perceptual beliefs; concept learning as hypothesis formation and confirmation; and decision-making as a form of representing and evaluating the consequences of possible actions carried out in a situation using a pre-existing set of preferences. These "best" cognitive models, he argued, all accept mental processes as being computational processes defined over representations. He concluded that, if these models are correct in their treating mental processes as computational, then there has to be a language of thought (LOT) over which they are defined. Hence, for Fodor, there is "no computation without representation." According to Pylyshyn (1993), human thought is the manipulation of sentences of an internal mental language or code; and this 'Mentalese' (or LOT), may, but is unlikely to, resemble the binary code of today's computer systems. Some of what makes minds rational is their ability to perform computations on thoughts, where thoughts, like sentences, are assumed to be syntactically structured and where 'computations' means formal operations in the manner of Turing.

In supporting this position, Fodor and Pylyshyn wrote an influential critical examination of the difference between computational (à la Turing) and connectionist cognitive architectures (Fodor & Pylyshyn 1987). They claim that the architecture of the mind/brain cannot be connectionist at the cognitive level and that connectionism may provide at most an account of the neural structures in which the Classical, computational cognitive architecture is implemented. They have to accept the latter possibility given that the brain itself is composed of neural structures. Fodor and Pylyshyn (henceforth F&P) outline what they consider to be serious flaws in a connectionist theory of mind but most of their arguments rest on what they see as a failure of connectionism to support a compositional semantics whereby the meaning of an expression is a function of the meaning of its constituent parts. This is a curious claim since there is at least one example of a neural network that supports compositional semantics, namely the brain itself, and F&P themselves accept that there are connectionist *implementations* of the Classical architecture that support compositional semantics. Nevertheless, F&P claim that connectionist models, as a theory of mind, cannot, in principle, have compositional semantics. They use a "toy" localist network to support their claim:

> To simplify the exposition, we assume a 'localist' approach, in which each semantically interpreted node corresponds to a single Connectionist unit; but nothing relevant to this discussion is changed if these nodes actually consist of patterns over a cluster of units. (Fodor & Pylyshyn 1988:15n)

Again, it is curious that F&P would assume a localist approach given that a central tenet of connectionism is that an atomic token, or node, is not capable of carrying sufficient information to make it useful in a discussion of human cognition. Connectionists place a strong emphasis on the distinction between local and distributed representations and there are examples of the connectionist approach to compositionality being used successfully to operate directly on distributed representations (see examples in Pollack 1988 and 1990, and Smolensky 1990). It is significant that Fodor himself later argued that Turing's account of computation is, in at least two respects, *local* given that it does not look past the form of sentences to their meanings, and that it assumes that the role of thoughts in a mental process is determined entirely by their internal (syntactic) structure. Unfortunately, there are some rational processes which are not local in either of these respects. For Fodor, it may be that:

> Wherever either semantic or global features of mental processes begin to make their presence felt, You reach the limits of what Turing's kind of computational rationality is able to explain. As things stand, what's beyond these limits is not a problem but a mystery. (Fodor 1998a)

5.4.2 **The Mind as Computer**

The symbol-manipulation hypothesis is basically that we have mental states in virtue of symbol-processing operations performed by the brain, and, hence, that all creatures with mental states are symbol-processing systems. The fundamental assumption of Artificial Intelligence (AI) is the symbol system hypothesis which claims that the computer is an appropriate kind of machine to think; and, further, a physical symbol-system has the necessary and sufficient means for general and intelligent action. This leads to treating the brain of all thinking creatures as symbol-manipulating computers.

Compared to a digital computer, however, the brain is very slow. For example, the brain takes a million times longer than a personal computer to perform a basic electrical operation. Hence the brain cannot be running traditionally conceived AI programs. Even if they are considerably faster than a human brain, traditional computers are sequential processors – they perform their instructions serially. This fact constitutes strong evidence

that the brain is some kind of parallel device: anatomical studies show that the pattern of parallel, simultaneously operating layers of neurons is repeated throughout the brain:

> If current estimates of the number of [connections] in the brain are anywhere near correct, it would take even the fastest of today's computers something between several years and several centuries to simulate the processing that can take place within the human brain in one second. (Rumelhart, McClelland, et al. 1986)

A cursory examination of the differences in how a classical, von Neumann computer and a human brain performs memory recall operations demonstrates significant differences:

- Each specific memory location in a computer has a unique address, but the human brain retrieves a memory using a description of the information required;
- A piece of information in a computer can only be retrieved if its unique numerical address is stated in the program, but the human brain can use any one of an open-ended collection of descriptions to retrieve the same information – memories are content-addressable – they are stored holistically in networks of memories.

Through content-addressability, humans have an open-ended access to memories. There has been no success, so far, in reconciling the computer's "content-blindness" with the open-endedness of human recall. In a computer, each string of symbol tokens exists at a specific physical location in the hardware, whereas the human memory system makes a very large use of distributed storage. This difference in how memories are stored makes computer hardware very subject to memory loss through hardware damage; the human brain, on the other hand, can withstand injury: its distributed storage of memories allows for a "graceful" degradation. As damage increases, the performance of a brain degrades gradually but computers "crash" very easily.

As yet, no AI system has been able to perform such common human cognitive functions as abduction. The cognitive processes that prove so difficult to implement in a classical computer system are those that rely on parallel processing but also those whose mechanisms we still do not understand. Trying to understand human cognition using the computer model is, in my opinion, fruitless. If we must continue using technology as an analogy, then the Internet might be more productive. It is, like the brain, massively-parallel, and we could view the individual computers in the network as analogous to neural nuclei. Nonetheless, I contend that our time would be better spent attempting to understand the

mind using the results of such empirical research areas as neuroscience and neurophysiology.

5.4.3 **The Computational Brain**

That the brain is a computational device is non-tendentious. The disagreement between supporters of the symbol-system hypothesis (aka *computationalists*) and connectionism lies in how and over what computation is performed. Both theories accept the existence of representational mental states but supporters of the former position hold to a symbol-level of representation and, even, to the idea that there is a correspondence between these representations and actual brain circuits. As discussed previously, some early supporters of the symbol-system hypothesis have expressed misgivings about the ability of the hypothesis to explain many crucial aspects of human cognition, but they (e.g., Fodor, Putnam) are not willing to reject the hypothesis completely in what they see as an absence of an viable alternative. Recently, cognitive scientists, such as C. R. Gallistel and Adam Philip King (2010), have argued strongly in support of the hypothesis and believe that it can present a fruitful direction for neuroscientific research. In examining their arguments, it must be emphasized that both they and connectionists can talk about the "computational brain", but they do not agree on what form the computing takes. For supporters of the symbol-system hypothesis, the computation is à la Turing (computation over symbols), and for connectionists, the computation is over neural networks (computational neurobiology). At first glance, they appear to be talking at different levels of explanation (cognitive versus implementation), but Gallistel and King (2010) appear to argue for *computation over symbols* as more than a cognitive level of explanation; they also see it as an implementation research strategy:

> there must be an addressable read/write memory mechanism in brains that encodes information received by the brain into symbols (writes), locates the information when needed (addresses), and transports it to computational machinery that makes productive use of the information (reads). (Gallistel & King 2010:Preface)

Fodor and Pylyshyn (1988) noted that a symbol-manipulating device could be implemented in a connectionist system, but stated that the mind/brain architecture is not connectionist *at the cognitive level*. Gallistel and King (2010) attempt to show how the mind as a symbol-manipulation computer could be realized in the brain. In doing so, they criticize connectionist models for relying too heavily on neuroscience for evidence of how neural

mechanisms might mediate computation rather than on the architecture that a powerful computing machine requires as demonstrated by "well-established results in theoretical and practical computer science" (Gallistel & King 2010:ix).

In the previous section, 5.4.2 "The Mind as Computer", I present arguments against viewing the mind as a computer à la Turing or von Neumann, but Gallistel and King (2010) not only accept such a view of the *mind* but also attempt to show just how it could be instantiated in the *brain*. The connectionist and neural network approach is that a memory consists of neurons distributed throughout the brain[69], but Gallistel and King claim that the existence of a neurobiological *read-write memory* mechanism is indispensable for explanations of psychological phenomena (Gallistel & King 2010:xii; Gallistel 2014:5). They do acknowledge, however, that neuroscience has not yet proved that such a memory exists (Gallistel 2014:5). The existence of such a memory is essential to their arguments concerning how computations are carried out in the brain; this is because of their assumption that the brain is a Turing-style, digital computing machine for which a read/write, sequentially structured symbolic memory is an essential component:

> to get machines that can do computations of reasonable complexity, a specific, minimal functional architecture is demanded, an architecture that includes a read/write memory (Gallistel & King 2010:128)

Gallistel (2014) proposes that the way information is stored in DNA may provide an indication of how a read-write memory may be neuro-biologically realised. He describes the indirect addressing process whereby a transcription factor initiates the reading of a codon sequence to specify a protein which leads to the synthesis of still other proteins which may yield further transcription factors. The example he provides is:

> The eye gene codes only for one protein. That protein does not itself appear anywhere in the structure of the eye. It is a transcription factor. It gives access to the addresses of other transcription factors and, eventually, through them, to the addresses of the proteins from which the special tissues of the eye are built and to the transcription factors whose concentration gradients govern how those tissues are arranged to make an organ. (Gallistel 2014:29)

He suggests that the brain would need a similar architecture for accessing stored information. The suggestion indicates a view that memories are stored hierarchically rather

[69] For the connectionist and neural network approach, memories are content-addressable and meaning is holistic. When trying to retrieve a memory, we may use many different means: we may go through the alphabet hoping that a letter of the alphabet will trigger the memory of a name; we may use the memory of a particular context in which we learnt the name; we may use a mnemonic (such as a "memory palace"); and so on. We are able to do this because each memory structure is linked in a neural network of memory structures.

than in networks and, hence, it is unclear to me how we could take different paths to retrieve the same information or the same starting information to retrieve different memory structures: how, for example, I could use the letter of an alphabet to trigger the memory of someone's name, or how I could use the memory of a smell to retrieve the memory of a location.

Gallistel and King contend that neither connectionism nor neuroscience can account for the existence of such a read/write memory but, in fact, they themselves note that it is possible to realize it using accepted neural mechanisms such as a "reverberating" neural activity (which functions in connectionist models as *working memory*) and in the connection strengths between neurons in a network (which functions in connectionist models as *long term memory*) (Gallistel & King 2010:285). Thus, their requirement for a read/write memory cannot be used as a criticism of connectionist models, even though they continue their arguments as if it is.

In order to support their position and show that connectionism, and neural network approaches in general, are unsatisfactory, they use the example of dead reckoning (path integration) and the model developed by Samsonovich and McNaughton (1997):

> This model relies on the only widely conjectured mechanism for performing the essential memory function, reverberatory loops. We review this model in detail because it illustrates so dramatically the points we have made earlier about the price that is paid when one dispenses with a read/write memory. To our mind, what this model proves is that the price is too high. (Gallistel & King 2010:xv)

They criticize the "enormous" complexity of the model (the "amount of computation that must be done rapidly gets out of hand"), the "*ad hoc* fixes" it requires, and its inability to handle "noisy" input (Gallistel & King 2010:254-256). These criticisms are specific to the model under consideration and do not apply to connectionist models in general: *ad hoc* fixes are not specific to connectionism; and some of the strengths of connectionist models are their ability to handle complex problems with limited numbers of neurons and pathways, their ability to handle "noisy", degraded, or limited input, and their resistance to "computational explosion". As concerns dead reckoning, later connectionist models (see Mole 2014 for a discussion) are not subject to the same criticisms. It is, hence, questionable to present a criticism of neural network architecture that is based on the failings of a particular model. In addition, it is, in my opinion, unwise to ignore the empirical evidence that neuroscience has and continues to produce when constructing a theory of cognitive architecture. I contend that greater advances will result by using the results of

neuroscience to aid computer science in developing more "human-like" architectures for computing machines, rather than the other way around.

5.4.4 **Neuroscience and Cognition**

As stated previously, Gallistel and King (2010:ix) criticize connectionist models for relying too heavily on neuroscience for evidence of how neural mechanisms might mediate computation. Their criticisms, I claim, result from their applying computation as understood by supporters of the symbol-manipulation hypotheses: they are seeking evidence from neuroscience of how neural mechanisms might mediate computation *over symbols*, a type of computation that connectionists do not accept. Supporters of the symbolic-manipulation paradigm posit symbolic models that, they contend, have a similar structure to underlying brain structures; connectionists, on the other hand, use "low-level" modelling with the aim of producing models that resemble neurological structures. It is clear, therefore, that connectionists would rely consistently on the results of neuroscience to refine their models. In the domain of psycholinguistics, for example, the search for neural architecture underlying such areas as written word processing and spoken word forms would be influenced by which paradigm is supported. For example, Broca's Area is purported to be the centre of language production, although there is disagreement about its precise location (Poeppel, 1996). For connectionists, mental processes are distributed and massively parallel, and this view is supported by results of neural imaging which purport to show that every lobe of the brain is involved in some measure in language production (Pulvermüller, 1999).

There has been skepticism as to the direction being taken by some of the research combining the study of language and the the brain. David Poeppel and David Embick, in particular, have identified what they considered to be serious problems with these research programs (see, especially, Poeppel 1996, Poeppel & Embick 2005). They identify two main problems, the first of which is the:

> **Granularity Mismatch Problem** (GMP): Linguistic and neuroscientific studies of language operate with objects of different granularity. In particular, linguistic computation involves a number of fine-grained distinctions and explicit computational operations. Neuroscientific approaches to language operate in terms of broader conceptual distinctions (Poeppel & Embick 2005:2-3);

Their complaint is that the elemental concepts of neurobiology and cognitive neuroscience are "coarse-grained" relative to the corresponding linguistic primitives, it is not possible to

formulate hypothesis that link neuroscience and linguistics (Poeppel & Embick 2005:2-3). It is interesting to note that Poeppel and Embick explicitly question whether there is any point to a science of language and brain which does not advance the understanding of either. They contend that this situation will not be resolved "until certain fundamental problems are identified, acknowledged, and addressed" (Poeppel & Embick 2005:2).

I suggest that this granularity mismatch is likely to result from a (perhaps implicit) acceptance of a form of the symbol-system hypothesis which has lead many philosophers, psychologists, and others to think of the brain as modular. The level of cognitive analysis adopted by the sub-symbolic paradigm, favoured by connectionists, is lower than the level traditionally adopted by the symbolic paradigm for which processes are sensitive to the structure of the symbols that are conceptual-level representations. For connectionists, entities are represented by a large number of sub-symbols, operations on which "often consist of a large number of finer-grained operations" (Smolensky, 1988:§1.3).

The second problem they identify has to do with whether the fundamental units of linguistic theory can be reduced to the biological units identified by neuroscience:

> **Ontological Incommensurability Problem** (OIP): The units of linguistic computation and the units of neurological computation are incommensurable (Poeppel & Embick 2005:4).

This problem is a result, they contend, of the independence of the development of the two ontologies, and claim that the problem will never be resolved as long as the current conceptual architectures of linguistics and neuroscience remain as they are. I contend that a resolution would be for linguistic theory to discard the symbol-system hypothesis.

5.4.5 **Non-conscious and Conscious Mind Distinction**

Most arguments against the thesis that all thinking is non-conscious tend to be of the form: "but when I make logical inferences, I am conscious of so doing." As discussed in the previous chapter, the results of neuroscientific studies (by Kornhuber et al., Libet, Wegner, and others) indicate that the impression that we think consciously is an illusion. I contend that all cognitive processes are performed non-consciously and, while we may become consciously aware of some aspects of a thought process, the consciousness itself is causally inert. What I am proposing is that the encoded images that appear in consciousness may become input to further cognitive processes and hence give the impression that the conscious mind is causally efficacious. The benefit of the encoding process is that aspects of

a cognitive process are produced in a form that makes them suitable as input for further processing, whether or not we become consciously aware of them. This is of significance during learning.

5.5 Part IV: Conclusions

Currently, the thesis that the human brain is a symbol-manipulating computer can be neither confirmed nor refuted by empirical means. Most support for such a position relies on philosophical arguments one of which is notably incorrect: many aspects of brain function, such as vision, clearly do not involve symbol manipulation. As Dreyfus and Dreyfus (1986) noted, human intelligence and expertise appear to depend on non-conscious instincts and not on conscious symbol-manipulation. Thus, there are, at least some, mental states which do not arise as the result of the manipulation of symbols. Many problems of cognition, such as the frame problem, arise for the Classical model as long as we adopt its assumption of the explanatory value of computation over representations. Yet, many philosophers of mind still support the view that human cognitive processes are, for the most part, inferences over a set of propositions using some type of computation. Some philosophers such as Fodor (1998a) have more recently accepted that the Classical, computational model has serious limitations, yet it still remains for many philosophers and other cognitive scientists "the only game in town." The Symbol-System Hypothesis, on which the Classical model is based, is an empirical theory – its credibility can only be established by careful study. Understanding human cognition is not a topic for armchair philosophical speculation but requires empirical research especially by neuroscientists. No existing digital computers are able to compete with the human brain in such areas as abductive reasoning or computing reasonable solutions based on very limited information. The failure of artificial intelligence research to produce machines that match the computing power of the human brain suggests that, rather than modelling human cognition on the Turing machine or on a von Neumann symbol-manipulation system, a more fruitful approach would be to develop computing systems that more closely simulate what we have learnt about the brain from neuroscience and neurobiology.

Fodor's "mystery" concerning the limits of what the Computational Theory of Mind is able to explain is, on my view, a result of insisting on the idea that the brain is, at least in most respects, some kind of Turing-style computer; that is, of insisting that computation over symbols is the best model of human cognition. Connectionism supports the more biologically-plausible position that variously weighted elements are the principal means of

computation as well as constituting the main memory store (Churchland 2007:33). Connectionism provides a more explanatory model of cognition and the intuition that cognitive processes constitute computation over symbols arises from the way these processes appear to the conscious mind. As well as dispensing with Fodor's "mystery", connectionism also avoids the linguistic puzzles discussed in section 2.3.9 "Externalist Semantics and Linguistic Puzzles".

In offering an explanation of consciousness, I propose a mechanism of encoding and decoding such that the conscious mind is like a computer screen in being an "interface" to the hidden, non-conscious, cognitive processes. On this analogy, whatever thoughts are encoded in a form that can appear in the conscious mind are like computer interface icons. The "icons" that appear in the conscious mind, like the icons on a computer screen, are causally-inert – recent evidence from neuroscience supports the view of the conscious mind as causally inefficacious.

Empirical research will be required to substantiate the claims made in this dissertation. One research project might be to determine whether the "appearance" of icons in the conscious mind results from the activation of a form of "alarm" system which is normally the result of a physiological reaction. Ethology studies like that of the Brosnan & de Wall (2001) study on capuchin monkeys (discussed on page 67) and the study of learning in the shore crab (Magee & Elwood 2013), mentioned in section 5.3.7 "Implicit and Explicit Learning" might be of use in this respect. In addition, there is a need for more research in linguistics and neuro-linguistics as to how connectionist models might handle such topics as *wh-movement* – the ability to explain how connectionism would handle such topics is crucial if linguists are to accept the cognitive model presented in this dissertation. A related topic is that of variable binding which is an essential component for modelling human cognition, especially for modelling the language capacity (Pinker & Prince 1988). Connectionists have already developed several solutions to the variable binding problem which offer promising directions for future research (see Browne & Sun 2001)

Recent progress in neurophysiology and the understanding of neural networks has led to connectionist models that have overcome many of the early problems noted by critics of connectionism. The parallel distributed processing of connectionist models reflects the massively parallel nature of the brain, whereas symbol-manipulation models have no resemblance whatsoever to underlying brain structure. Finally, I contend that classical symbol-manipulation, computational models and cognitive science, in general, rest on Cartesian

assumptions that need to be rejected. There is, on my view, a very limited consciousness and "everything else", and the "everything else" is best modelled, at the cognitive level, by a connectionist architecture.

---

6. Bibliography

Allen, C., & M. Bekoff (1997) *Species of Mind, The philosophy and biology of cognitive ethology*. Cambridge, MA: MIT Press

Allport, D. A. (1977) On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.) *Attention and performance VI* London:Academic Books.

Anderson, J. R. (1995) *Learning and Memory: An Integrated Approach*. New York: John Wiley & Sons, Inc.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind.* London: Routledge and Kegan Paul.

Armstrong, S., Gleitman L., and Gleitman, H. (1983) 'What Some Concepts Might Not Be', *Cognition*, 13:263-308.

Astington, J. (2006) The Developmental Interdependence of Theory of Mind and Language. In *Roots of Human Sociality*. N.J. Enfiled and Stephen C. Levinson (eds) pp. 179-206. Berg: NY.

Ballard, D. (1991) Animate Vision *Artificial Intelligence* 48, 57-86

Ballard, D., Hayhoe, M., Pook, P. and Rao, R. (1997) "Dieictic Codes For the Embodiment of Cognition" *Behavioral and Brain Sciences* 20:723-767

Bargh, J. A. (2011). Unconscious Thought Theory and its discontents: A critique of the critiques. *Social Cognition*, 29, 629-647

Barkow, J., Cosmides, L., and Tooby, J., eds. (1992) *The Adapted Mind: Evolutionary psychology and the generation of culture,* New York: Oxford University Press

Barsalou, L. W. (1987) The Instability of Graded Structure: Implications for the Nature of Concepts. In U. Neisser, ed., *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. Cambridge: Cambridge University Press.

Barsalou, L. W., Pecher, D., Zeelenberg, R., Simmons, W. K., and Hamann, S. (2003) 'Multi-Modal Simulation in Conceptual Processing'. W. Ahn, R. Goldstone, B. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Festschrift in honor of Douglas L. Medin*. Washington, DC: American Psychological Association

Bechtel, W. and Graham, G. (1998) (eds.) *A Companion to Cognitive Science*, Blackwell Publishing

Bellugi, U., Klima, E.S., and Siple, P. (1975). Remembering in signs. *Cognition*, 3(2), 93-125.

Bengson, J. J., Kelley, T. A., Zhang, X., Wang, J-L., Mangun, G. R. (2014) Spontaneous Neural Fluctuations Predict Decisions to Attend, *Journal of Cognitive Neuroscience,*16 April 2014

Bennett, M., Dennett, D., Hacker, P., and Searle, J. (2007) *Neuroscience & Philosophy: Brain, Minda, & Language*, New York: Columbia University Press).

Berlin, B., and Paul Kay (1969) *Basic color terms: their universality and evolution*. Berkeley: University of California Press.

Besner, D., and Stolz, J. A. (1999) Unconsciously controlled processing: The Stroop effect reconsidered. *Psychonomic Bulletin & Review*, 6, 449-455.

Block, N. ed. (1981), *Readings in Philosophy of Psychology, Vol. 2,* Cambridge, Mass.: Harvard University Press.

_____, (1986) 'Advertisement for a semantics for psychology', in French et al (1986).

Boden, M. (2006) *mind as machine*, in two volumes, (Oxford: Clarendon Press)

Boghossian, P. (2000) "Knowledge of Logic," in P. Boghossian and C. Peacocke (eds.): *New Essays on the A Priori* (Oxford:Oxford University Press, 2000)
Bogdan, R. (1993), 'The Architectural Nonchalance of Commonsense Psychology, *Mind & Language* 8(2): 189-205.
Boyd, R, (1988) "How to Be a Moral Realist", in G. Sayre-McCord (ed.) Essays on Moral Realism, Cornell University Press
_____, (1991) "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds", *Philosophical Studies* 61: 127–148.
_____, (1999a) "Homeostasis, Species, and Higher Taxa", in R. Wilson (ed.), *Species: New Interdisciplinary Essays*, Cambridge, Massachusetts: MIT Press: 141–186.
_____, (1999b) "Kinds, complexity and multiple realization: comments on Millikan's 'Historical Kinds and the Special Sciences'", *Philosophical Studies*, 95: 67– 98
Brooks, R. (1991) Intelligence without representation *Artificial Intelligence* 47, 139-159
Brosnan, S. F., and de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature* 13, 2002, p. 129–152.
Browne, A., and Sun, R. (2001) Connectionist Inference Models. *Neural Networks* 14 (2001) 1331-1355.
Bryson, J. J. (2006) *The adaptive advantages of knowledge transmission*.
Brooks, R. A., "Intelligence without Representation", in J. Haugeland (1997).
Butterworth B, Reeve R, and Lloyd D. (2008) Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences* 105(35): 13179-13184
Campbell, J. (2002) *Reference and Consciousness* (Oxford: Oxford University Press).
Carey, S. (1985) *Conceptual Change in Childhood* (Cambridge, Mass., MIT Press).
_____, (1991) "Knowledge Acquisition: Enrichment or Conceptual Change?" in S. Carey and R. Gelman,eds., *Epigenesis of Mind: Essays on Biology and Cognition.* 1991, Lawrence Erlbaum Associates, Inc., Publishers, reprinted in Margolis & Lawrence (1999, Chapter 20)
Carey, S. and Spelke, E. (1994) Domain-specific knowledge and conceptual change. In Hirschfeld & Gelman (Eds.) (1994), 169-200
_____,(1996) Science and Core Knowledge. *Philosophy of Science*, 63, 515-533.
Caramazza, A., and Shelton, J. R. (1998) Domain-specific knowledge systems in the brain the animate-inanimate distinction, *Journal of Cognitive Neuroscience*, 1998 Jan;10(1):1-34.
Carnap, R. (1952) "Meaning postulates" in *Philosophical Studies, III* (now in *Meaning and Necessity*, 1956, 2nd edition)
_____, (1947/1956) *Meaning and Necessity* (Chicago, Ill.: University of Chicago Press).
Carruthers, P. (2001) 'Review of Fodor, The Mind Doesn't Work That Way,' *Times Literary Supplemen*t, 5 October, p. 30.
_____, (2006) *The Architecture of the Mind*, (Oxford: Clarendon Press).
Carruthers, P., Laurence, S., and Stich, S. (2005) (eds.) *The Innate Mind: Structure and Contents*, (Oxford:Oxford University Press).
_____, (2006) (eds.) *The Innate Mind: Culture and Cognition*, (Oxford:Oxford University Press).
_____, (2007) (eds.) *The Innate Mind: Foundations and the Future*, (Oxford:Oxford University Press).
Cartwright, H. M. (1970) 'Quantities', *Philosophical Review* 79, 25–42.
Cartwright, H. M. (1975) 'Some Remarks About Mass Nouns and Plurality', *Synthese* 31, 395–410.

Cattaneo, L., and Rizzolatti, G. (2009). The mirror neuron system. *Arch. Neurol.* 66, 557–560.
Chaiken, S., ed. (1999) *Dual-Process Theories in Social Psychology*, The Guilford Press
Chalmers, D. (1990) Why Fodor and Pylyshyn Were Wrong: the Simplest Refutation. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society,* Cambridge, Mass.
Cheney, D. L., and Seyfarth, R. M. (1990) *How Monkeys See the World: Inside the Mind of Another Species*. University of Chicago Press.
Chomsky, N. (1957/2002) *Syntactic Structures*, (Berlin, New York: Mouton de Gruyter).
_____, (1965) *Aspects of the Theory of Syntax*. MIT Press
_____, (1980) *Rules and representations,* New York: Columbia University Press
_____, (1986) *Knowledge of Language: Its Nature, Origin, and Use*, (Conneticut: Praeger).
_____, (2000/2005) *New Horizons in the Study of Language and Mind*, (Cambridge: Cambridge University Press).
_____, (2006) *Language and Mind*, Third Edition, (Cambridge: Cambridge University Press).
Churchland, P. M. (1981) Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, Vol. 78, No. 2 (Feb., 1981), pp. 67-90
_____, (1987) 'Epistemology in the age of neuroscience', *Journal of Philosophy*, 84: 544-53.
_____, (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science* (Cambridge, Mass.: MIT Press).
_____, (1990) "On the Nature of Theories: A Neurocomputational Perspective", reprinted in Haugeland (1997), Chap. 10
_____, (2007) *Neurophilosophy at Work.* Cambridge University Press
Churchland, P. S., and Sejnowski, T. J. (1992/1999) *The Computational Brain*, (Cambridge, Mass.: MIT Press)
Clark, A. (1989) *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, Mass: MIT Press
_____, (1993) *Associative Engines,* Cambridge, Mass: MIT Press
Clark, A., and Eliasmith, C. (2002) Philosophical Issues in Brain Theory and Connectionism. *The Handbook of Brain Theory and Neural Networks*, Second edition, (M.A. Arbib, Ed.), Cambridge, MA: The MIT Press
Clarke, M. (2004) *Reconstructing Reason and Representation*, (Cambridge, Mass.:MIT Press).
Clayton, N, Dally, J. M., and Emery, N. (2007) 'Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist.' *Philos Trans R Soc Lond B Biol Sci*. 2007 April 29; 362(1480): 507–522
Conrad, R. (1962) An association between memory errors and errors due to acoustic masking of speech. *Nature*, 193, 1314-1315.
_____, (1970) Short-term memory processes in the deaf. *Brit. J. Psychol*., 61, 179-195.
_____, (1972) Speech and reading. In J. Kavanagh and I. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, M.I.T. Press.
Cook, M., and Mineka, S. (1990) "Selective Association in the Observational Conditioning of Fear in Rhesus Monkeys" printed in the *Journal of Experimental Psychology: Animal Behaviour Processes*, Vol. 16, No. 4, 372-389.
Copeland, J. (1993) *Artificial Intelligence: A Philosophical Introduction,* Blackwell Publishing
Corballis, M. C. (1989) Laterality and human evolution. *Psychological Review*, 96(3):492-505.
_____, (2004a) FOXP2 and the mirror system. *Trends in Cognitive Sciences*, 8(3):95-96

_____, (2004b) The Origins of Modernity: Was Autonomous Speech the Critical Factor? *Psychological Review*, 111(2): 543–552
Corteen, R. S., and Dunn, D. (1974). "Shock-associated words in a non attended message: A test for momentary awareness.". *Journal of Experimental Psychology* 102: 1143–1144.
Cosmides, L., and Tooby, J. (1987) "From evolution to behavior: Evolutionary psychology as the missing link.", in J. Dupre (Ed.), *The latest on the best: Essays on evolution and optimality*, Cambridge, MA: The MIT Press
_____, (1994) "Origins of Domain Specificity: The Evolution of Functional Organization", in *Mapping the Mind*, (L. Hirschfeld and S. Gelman, eds.) p. 85-116
_____, (1997) *Evolutionary Psychology: A Primer*, Center for Evolutionary Psychology, University of California)
Dupre (Ed.), The latest on the best: Essays on evolution and optimality, Cambridge, MA: The MIT Press, 1987.
Cowie, F. (1999) *What's Within* (Oxford: Oxford University Press).
Damasio, A. R., and Tranel, D. (1993) "Nouns and verbs are retrieved with differently distributed neural systems." *Pro. Natl., Acad. Sci. USA,* Vol. 90, pp.4957-4960, June 1993, Neurobiology
Davidson, D. (1974) "On the Very Idea of a Conceptual Scheme", *Proceedings and Addresses of the American Philosophical Association*, 47: 5–20
_____, (1975) "Thought and Talk", in S. Guttenplan (ed.) *Mind and Language*, Oxford: Oxford University Press
_____, (2001) *Subjective, Intersubjective, Objective* (Oxford: Oxford University Press).
Daugherty, K., and Seidenberg, M.S. (1992). Rules or connections? The past tense revisited. *Proceedings of the 14th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
Davis, G., E. Gray, A. Simpson, D. Tall, and M. Thomas. (2000) "What is the object of the encapsulation of a process". *Journal of Mathematical Behavior*, 18(2) pp. 223-241.
Dawkins, R. (2004) *The Ancestor's Tale*, Boston: Houghton Mifflin
Deacon, T. W. (1999) Language evolution and neuromechanisms. In W. Bechtel & G. Graham (eds.) *A Companion to Cognitive Science*. (pp. 212-225). Oxford, UK: Blackwell Publishing.
De Almeida, R. G. (1998) *The representation of lexical concepts: A psycholinguistic inquiry.* Unpublished Ph.D. dissertation, Rutgers University, New Brunswick, NJ.
_____, (1999) "What Do Category-Specific Semantic Deficits Tell Us about the Representation of Lexical Concepts?", *Brain and Language* ***68***, 241-248
de Groot, A. D. (1965) *Thought and choice in chess*. The Hague: Mouton & Company.
Dehaene S. (1997) *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
Dennett, D. (1969) *Content and Consciousness*, London: Routledge and Kegan Paul
_____, (1978) *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: Bradford Books/MIT Press
_____, (1982). 'Beyond belief', in A. Woodfield (ed.) *Thought And Object* (Oxford: Clarendon).
_____, (1987) *The Intentional Stance*, (Cambridge: Mass.: MIT Press)
_____, (1990) The Interpretation of Texts, People and Other Artifacts" *Philosophy and Phenomenological Research (Supplement)* 50:177-194
_____, (1995). Darwin's Dangerous Idea. New York: Touchstone/Simon & Schuster.
Dessalles, J-L. (2006) Du protolangage au langage : modèle d'une transition. Marge linguistiques 11:142-152.

Deutsch, J. A. (1960) *The Structural Basis of Behavior* (Chicago, Ill.: University of Chicago Press).
Devitt, M., and Sterelny, K. (1987) *Language and Reality* (Cambridge, Mass.: MIT Press).
Dingfelder, S. F. (2005) Autism's smoking gun? *Monitor on Psychology. American Psychology Association*, 36:9
Donnellan, K. S. (1977) Reference and definite descriptions. In S. P. Schwartz (Ed.), *Naming, Necessity, and Natural Kinds,* Ithaca, NY: Cornell University Press, pp. 42-65
_____, (1990) "Belief and the Identity of Reference." In A.C. Anderson and J Owens (eds.), *Propositional Attitudes*, (pp. 201-214). CSLI, Stanford.
Dowty, D. (1979) *Word Meaning and Montague Grammar*, Dordrecht: Reidel
Dretske, F. (1981) *Knowledge And The Flow of Information* (Cambridge Mass.: MIT Press).
_____, (1995) "Meaningful Perception", in D. N. Osherson et S. M. Kosslyn (eds), *An invitation to Cognitive Science, vol. 2: Visual Cognition*, MIT Press
Dreyfus, H. L. (1978) *What Computers Can't Do* (London: Harper Collins).
Dreyfus, H. L., and Dreyfus, S. E. (1986) *Mind over Machine: the power of human intuition and expertise in the era of the computer* Oxford; Basil Blackwell
_____, (2004) From Socrates to Expert Systems: The Limits and Dangers of Calculative Rationality. Retrieved from: *http://socrates.berkeley.edu/~hdreyfus/html/paper_socrates.html*
Dudman, V.H. 1976, 'Bedeutung for Predicates', in M. Schirn, (ed.) *Studien zu Frege III: Logik und Semantik*, Fromann-Holzboog, Stuttgart, see also H. Sluga, (ed.) The Philosophy of Frege, Vol 4, Garland, New York, 1993.
Dummett, Michael. *Frege: Philosophy of Language*, chap.7, pp.211-219.
Dunbar, R. (1996) *Grooming, gossip, and the evolution of language*. Cambridge: Harvard University Press.
_____, (2001) Brains on two legs: group size and the evolution of intelligence. In *Tree of Origin: What Primate Behavior Can Tell Us About Human Social Evolution* (173-191). London: Harvard University Press.
Eckhardt, B. (1993) *What is Cognitive Science?* Cambridge, Mass.: MIT Press
Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996) *Rethinking Innateness,* Cambridge, Mass.: MIT Press
Emmorey, K. (2002) *Language, Cognition, and the Brain*, New Jersey: Lawrence Erlbaum Associates
_____, (2006) The signer as an embodied mirror neuron system: neural mechanisms underlying sign language and action. In M. A. Arbib (ed.) *Action to Language via the Mirror Neuron System*. (pp. 110-135). Cambridge, UK: Cambridge University Press
Epstein S. (1994) Integration of the cognitive and psychodynamic unconscious. *Am. Psychol*. 49:709–24
Erickson, M. A., and Kruschke, J. K. (2002) Rule-based extrapolation in perceptual categorization, *Psychonomic Bulletin & Review*. 2002, 9(1), 160-168
Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press
Evans, J. St-B. T., and Over, D. E. (1996) *Rationality and Reasoning*. Hove: Psychology Press
Fitch, T. W. (2005) The evolution of language: a comparative review. *Biology and Philosophy* 20:193-230.
Fodor, J. A., Garrett, M., and Brill, S. L. (1975) Pe, ka, pu: the perception of speech sounds in prelinguistic infants. *M. I. T. Quarterly Progress Report*, January, 1975. (Submitted to *Science*)
Fodor, J. (1968) 'The appeal to tacit knowledge in psychological explanation'*, Journal of Philosophy,* 65: 627-40
_____, (1975) *The Language of Thought*, (Cambridge, Mass.: Harvard University Press)

_____, (1978) “Propositional Attitudes”, *The Monist,* 61:501-523
_____, (1981) *Representations* (Cambridge, Mass.: MIT Press)
_____, (1983) *The Modularity of the Mind*, (Cambridge, Mass.: MIT Press).
_____, (1985) “Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum”, *Mind*, New Series, Vol. 94, No. 373, (Jan., 1985), pp. 76-100
_____, (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
_____, (1990) *A Theory of Content* (Cambridge Mass.: MIT Press).
_____, (1994) *The Elm and the Expert* (Cambridge Mass.: MIT Press).
_____, (1995) “Concepts: A Potboiler”, *Philosophical Issues*, Vol. 6, pp. 1-24
_____, (1998a) *Concepts* (Oxford: Oxford University Press).
_____ (1998b) The Trouble with Psychological Darwinism, From *The London Review of Books,* Vol 20, No 2
_____, *(*2000) *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology* (Cambridge, Mass.: MIT Press).
_____, (2001) 'Doing Without What's Within', *Mind,* 110/437: 99-148.
_____, (2004) 'Language, thought and compositionality', *Mind and Language,* 16/1: 1-15.
_____, (2008) 'Against Darwinism', *Mind and Language,* 23/1: 1-24.
_____, (2008) *LOT2: The Language of Thought Revisited*, (Oxford: Oxford University Press).
Fodor, J. D., Fodor, J. A., and Garrett, M. F. (1975) “The psychological unreality of semantic representations”, *Linguistic Inquiry*, 6, 515–531.
Fodor, J., and Lepore, E. (1992) *Holism* (Oxford: Blackwell).
_____, *(*1996) The Red Herring and the Pet Fish: Why Concepts Still Can't Be Prototypes, *Cognition, 58, 253-270.(*
_____, (2002a) 'The Emptiness of the Lexicon', *Linguistic Inquiry,* 29/2: 269-88*.*
_____, (2002b) *The Compositionality Papers* (Oxford: Oxford University Press).
_____, (2007) 'Brandom Beleaguered', *Philosophy and Phenomenological Research,* 74/3: 677-91.
Fodor, J. D., Fodor, J. A. & Garrett, M. F. (1975). The psychological unreality of semantic representations. *Linguistic Inquiry*, 4: 515–531.
Fodor, J., and McLaughlin, B. (1990) 'Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work', *Cognition,* 35: 183-204·
Fodor, J., and Pylyshyn, Z. (1988) 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition,* 28: 3-71.
Fogassi, L. & Ferrari, F.(2007) Mirror neurons and the evolution of embodied language. *Current Directions in Psychological Science*, 16(3):136-141.
Franks, F. (Ed.) (1972) Water, A Comprehensive Treatise. *Vol. 1: The Physics and Physical Chemistry of Water,* New York: Plenum Press
Frege, G, (1984) 'The Foundations of Arithmetic' in *Philosophical Writings of Gottlob Freqe,* Blackwell, 1952
_____, (1891a) 'Function and Concept' in *Philosophical Writings of Gottlob Freqe,* Blackwell, 1952
_____, (1891b) “Letter to Husserl, 24.5.1891” in *Philosophical Writings of Gottlob Freqe,* Blackwell, 1952
_____, (1892a) “On Concept and Object” in *Philosophical Writings of Gottlob Freqe,* Blackwell, 1952
_____, (1892b) “On *Sinn* and *Bedeutung”* in *Philosophical Writings of Gottlob Freqe*, Blackwell, 1952
_____, (1918) “Thought*”* in *Philosophical Writings of Gottlob Freqe*, Blackwell, 1952

Fumerton, R. (1998) "Externalism and Epistemological Direct Realism," *The Monist*, Vol. 81, No. 3, 1998, 393-406.
Gallistel, C. R. (2014) Learning and Representation, To appear in *Learning and memory: A comprehensive reference*. R. Menzel (Vol. Editor) & J. Byrne (Editor). New York: Elsevier
Gallistel, C. R., King, A. P. (2010) *Memory and the computational brain : why cognitive science will transform neuroscience*, John Wiley & Sons Ltd
Gazzaniga, M. S, (1988) 'Brain Modularity: Towards a Philosophy of Conscious Experience'. In Marcel, A. J., Bisiach, E. (eds) 1988. *Consciousness in Contemporary Science*, Oxford: Clarendon Press, pp. 218-38.
Gazzaniga, M.S., Bogen, J. E., Sperry, R. W. (1965) 'Observations on Visual Perception After Disconnection of the Cerebral Hemispheres in Man'. *Brain*, 8, pp. 221-36.
Gazzaniga. M. S., Le Doux. J. E. (1978) *The Integrated Mind*. New York: Plenum.
Gelman S A; Markman E M (1986) Categories and induction in young children. *Cognition* 1986;23(3):183-209.
Gelman S A; Markman E M (1987) Young children's inductions from natural kinds: the role of categories and appearances. *Child development* 1987;58(6):1532-41.
Gelman, S., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science*, 10, 489–493.
Gelman, S. A., and Medin, D. L. 1993. What's so essential about essentialism? Different perspective on the interaction of perception, language, and conceptual knowledge. *Cognitive Development* 8:157-167.
Gentilucci, M. & Corballis, M. C. (2006) From manual gesture to speech: A gradual transition. *Neuroscience and Biobehavioral Reviews*, 30:949-960.
Gertler, B. and Shapiro, L. (2007) (eds.) *Arguing About the Mind* (New York: Routledge)
Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems* (Boston, Mass.: Houghton Miffiin).
Gigerenzer, G., and Hug, K. (1992) Domain-specific reasoning: social contracts, cheating and perspective chang. *Cognition* **43**:127-71
Gigerenzer, G., and Selten, R., eds.(2001) *Bounded Rationality: The Adaptive Toolbox,* The MIT Press
Gleick, J. (1988) *Chaos: Making a New Science. Penguin Books*
Gleitman, L., Cassidy, K., Nappa, R., Papafragon, E., and Trueswell, J. (2005) 'Hard words', *Journal of Language Learning and Development,* 1/1: 23-64.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology* 22: 1166-1183.
_____, (1997). Perception and production in an episodic lexicon. In Keith Johnson and John W. Mullennix, editors. *Talker Variability in Speech Processing*, pages 33-66. Academic Press, San Diego 1997.
Goodman, N. (1968/1976). *Languages of art: An approach to a theory of symbols* (2nd ed.). Indianapolis, IN: Hackett.
Gopnik, A. (1988) Conceptual and semantic development as theory change. *Mind and Language* 3:197-216
_____, (1996a) The Scientist as Child. *Philosophy of Science*, 63, 485-514.
_____, (1996b) A Reply to Commentators. *Philosophy of Science*, 63, 552-561.
Gopnik, A., and Meltzoff, A. N. (1997) *Words, Thoughts and Theories,* Bradford Books: MIT Press.
Gould, S. J. & Lewontin, R. C. (1979) The spandrels of San Marco and the Panglossian paradigm. A critique of the adaptionist programme. *Proceedings of the Royal Society of London*, 205:281-288

Grandin, T. (2000) My Mind is a Web Browser: How People with Autism Think. *Cerebrum*, 2000 Winter Vol. 2, Number 1, pp. 14-22
Grice, H. P. (1969). 'Vacuous Names', in *Words and Objections*, ed. Donald Davidson and Jaakko Hintikka, Dordrecht: Reidel, 118–145
Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know and do not know? A*nimal Behavior*, 61, 139-151.
Harris, Z. (1988) *Language and Information* (New York: Columbia University Press).
Haugeland, J. (1997) (ed.) *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, (Cambridge, Mass.: MIT Press)..
_____, (1979) "Understanding Natural Language." *The Journal of Philosophy*, Vol. 76, No. 11, Seventy-Sixth Annual Meeting of the American Philosophical Association, Eastern Division (Nov., 1979) pp. 619-632
Hempel, C. G. (1951) "The Concept of Cognitive Significance: A Reconsideration", *Proceedings of the American Academy of Arts and Sciences*, Vol 80, No. 1, Jul., 1951
Herrnstein. R. (1979). Acquisition. Generalization, and Discrimination Reversal of a Natural Concept. *Journal of Experimental Psychology: Animal Behavior Processes*. 5. 118-129.
_____, (1984). Objects. Categories. And Discriminative Stimuli. In H. Roitblat. T. TelTace (Eds.), *Animal Cognition* (pp. 233-261). Hillsdale, NJ: Lawrence Erlbaum Associates.
Hill, J. H. and Mannheim, B. (1992) "Language and World view", *Annual Review of Anthropology* 21: 381–406
Hirschfeld, L. A., Gelman, S. A., eds. (1994) *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge University Press
Hoffman, D. D. (1998) *Visual Intelligence: How We Create What We See*, New York: W. W. Norton & Company
_____, (2008) Conscious Realism and the Mind-Body Problem. *Mind & Matter* Vol. 6(1), pp. 87–121
_____, (2009) The Interface Theory of Perception. *Object Categorization: Computer and Human Vision Perspectives*, edited by Sven Dickinson, Michael Tarr, Ales Leonardis and Bernt Schiele. Cambridge University Press, 2009, pages 148-265
_____, (2010) Human Vision as a Reality Engine. In a *Psychology Reader: Foundation for the Advancement of Behavioral and Brain Sciences*
Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York: Basic Books.
Hohenhaus, P. U. (1998) "Non-Lexicalizability - As a Characteristic Feature of Nonce-Formations in English and German." *Lexicology* 4.2, 237-280.
_____, (2000) "An Overlooked Type of Word-Formation: Dummy-Compounds in German and English." In: Christopher Hall & David Rock (eds.), *German Studies towards the Millennium. CUTG Proceedings vol. 2.* Oxford, Bern, Berlin, Bruxelles, Frankfurt a.M., New York, Wien: Peter Lang, 241-260
Horwich, P. (1992). Chomsky versus Quine on the analytic–synthetic distinction. *Proceedings of the Aristotelian Society*, 92: 95–108.
Hume, D. (1739/1985) *A Treatise of Human Nature* (London: Penguin).
Humphreys, G., & Quinlan, P. (1987). Normal and Pathological Processes in Visual Object Constancy . In G. Humphreys, & M. Riddoch (Eds.), *Visual Object Processing: A Cognitive Neuropsychological Approach* (pp. 43-105). London: Lawrence Erlbaum Associates .
Hurford, J. R. (2004) Language beyond our grasp: what mirror neurons can, and cannot, do for language evolution. In D. Kimbrough Oller and Ulrike Griebel (eds.) *Evolution of Communication Systems: A Comparative Approach*, pp. 297-313. Cambridge, MA: MIT Press.

Isac, D., and Reiss, C. (2008) *I-Language: An Introduction to Linguistics as Cognitive Science,* (Oxford: Oxford University Press).
Inhelder, B., and Piaget, J. (1958) De la logique de l'enfant à la logique de l'adolescence,. PUF; English translation. The Growth of Logical Thinking from Childhood to Adolescence. London: Routledge & Kegan Paul, 1958 [1955].
Jackendoff, R. (1983) *Semantics and Cognition* (Cambridge, Mass.: MIT Press).
_____, (1997) "Semantics and Cognition." *The Handbook of Contemporary Semantic Theory*. Lappin, Shalom (ed). Blackwell Publishing
_____, (1989) "What Is a Concept That a Person May Grasp It? *Mind & Language*, 4. 1989, Blackwell Publishers Ltd, reprinted in Margolis & Lawrence (1999, Chapter 13)
_____, (1993) *Languages of the Mind* (Cambridge, Mass.: MIT Press).
_____, (1997) *The Architecture of the Language Faculty, (*Cambridge, Mass: MIT Press)
_____, (2002) *Foundations of Language*, (New York: Oxford)
_____, (2012) *A User's Guide to Thought and Meaning*, Oxford University Press
Jaynes, J. (1976) The evolution of language in the late pleistocene. *Annals New York of the Academy of Sciences*, 280:312-325
Jeffrey, G. A. (1997) *An Introduction to Hydrogen Bonding,* Oxford: Oxford University Press
Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics* 50: 101-113.
Johnson-Frey, S. H (2003) Mirror neurons, Broca's area and language: Reflecting on the evidence, *Behavioral and Brain Sciences*, 26: 226-227
Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.
Kant, I. (1785/1956) *Groundwork of the Metaphysics of Morals*, Trans. H. J. Paton, New York: Harper Torchbooks
_____, (1963) *Critique of Pure Reason*, Translated by Norman Kemp Smith, St Martin's Press, New York
Kaplan, D. (1979) "On the logic of demonstratives", *The Journal of Philosophical Logic*, 8:81-98
Kay, P. and C. McDaniel (1978) The linguistic significance of the meanings of color terms. *Language* 54, 610-646.
Kegl, J. (1994) The Nicaraguan Sign Language Project: An Overview. *Signpost*. vol.7, no. 1, Spring, pp. 24–31.
Keil, F. (1989) *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
Kim, J. (1992) 'Multiple Realization and the Metaphysics of Reduction', *Philosophy and Phenomenological Research,* 52: *1*-26.
_____, (2005) *Physicalism or Something Near Enough*, Princeton: Princeton University Press
Klein, G. (1999) *Sources of Power*, Cambridge, Mass. : MIT Press
Koffka, F. (1935) *Principles of Gestalt Psychology* (New York: Harcourt).
Komatsu, L. K. (1992) Recent Views of Conceptual Structure. *Psychological Bulletin*, Vol. 112, No. 3, 500-526
Kornblith, H. (1993) *Inductive Inference and Its Natural Ground*. Cambridge, MA: MIT Press.
_____, (2002) *Knowledge and its Place in Nature*, Oxford University Press
Kornhuber, H.H. & Deecke, L. (1964) Hirnpotentialindertangen helm Menschcn vor und nach Willktirbewegungen, dargestellt mit Magnetbandspeicherung und Rtckwttlsanalyse (Changes in human brain potential before and after voluntary movement studied by recording on magnetic tape and reverse analysis). PfldgersArch—Ear. J. Physiol. 281:52.1964.
Krifka, M. (2004) Bare NPs: Kind-referring, Indefinites, Both, or Neither? *Empirical Issues in Formal Syntax and Semantics*, O. Bonami & P. Cabredo Hofherr, eds. 2004. pp. 111–132

Kripke, S. (1971) 'Identity and Necessity', in M. Munitz (ed.), *Identity and Individuation,* New York: New York University Press
_____, (1972/1980) *Naming and Necessity*, Cambridge MA: Harvard University Press
_____, (1977) 'Speaker's Reference and Semantic Reference.' *Midwest Studies in Philosophy*, II
_____, (1979) 'A puzzle about belief', in A. Margalit (ed.) *Meaning And Use,* Dordrecht: Reidel, 239-283 in A. P. Martinich, *The Philosophy of Language*, Oxford University Press, 1976
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B. (1992) "Linguistic experience alters phonetic perception in infants by 6 months of age.", *Science*. 1992 Jan 31;255(5044):606-8.
Kuhn, T. (1962) *The Structure of Scientific Revolutions.* The University of Chicago Press.
Kushner, D. J., Baker, A., and Dunstall, T. G. (1999). "Pharmacological uses and perspectives of heavy water and deuterated compounds". *Can. J. Physiol. Pharmacol*. 77 (2): 79–88
Ledoux, J. (2002) *The Synaptic Self* (NY: Viking).
Leibniz, G. (1988) *Discourse on Metaphysics and Related Writings*, edited and translated by R.N.D. Martin and Stuart Brown. New York: Manchester University Press
_____, (1996) *New Essays on Human Understanding.* P Remnant & J. Bennett (eds.), Cambridge University Press.
Libet, B. (1999), "Do We Have Freewill?", *Journal of Consciousness Studies*, 6, No. 8–9, 1999, pp. 47–57
Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K. (1983). 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act', *Brain*. 106 (3):623–642.
Lieberman, P. (2006) *Toward an Evolutionary Biology of Language*. Cambridge, Mass.: The Belknap Press of Harvard University Press
Lewis, D. (1970) 'General Semantics', *Synthese* 22:18-67.
_____, (1972) 'Psychophysical and Theoretical Identification, *Australasian Journal of Philosophy,* 50*:* 249-58.
Limber, J. (1982). What can chimps tell us about the origins of language. In S. Kuczaj (Ed.) *Language Development Volume 2* (pp. 429-446). Hillsdale, NJ: L. E. Erlbaum.
Locke, J. (1690/1975) *An Essay Concerning Human Understanding*, Oxford, UK: Oxford University Press
Locke, J. L. (2001) Rank and relationships in the evolution of spoken language. *Journal of the Royal Anthropological Institute*, 7:37-50.
Locke, J. L. & Bogin B. (2006) Language and life history: A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, 29:259-325.
Loewer, *B.,* and Rey, G. (1991) (eds.) *Meaning In Mind: Fodor And His Critics* (Oxford: Blackwell).
Lowe, E. J. (2011), Locke on Real Essence and Water as a Natural Kind: A Qualified Defence. *Aristotelian Society Supplementary Volume*, 85: 1–19. doi: 10.1111/j.1467-8349.2011.00193.x
Lyness, R. C., Alvarez, I., Sereno, M. I., MacSweeney, M. (2014). Microstructural differences in the thalamus and thalamic radiations in the congenitally deaf. *Neuroimage* doi:10.1016/j.*neuroimage*.2014.05.077.
MacAndrew, A. (2003) FOXP2 and the evolution of language. Retrieved November 3, 2007, from http://www.evolutionpages.com/FOXP2_language.htm

Macdonald, c., and Macdonald, G. (1995) *Connectionism* (Cambridge, Mass.: Blackwell).
McDonough, L., Choi, S., and Mandler, J. M. (2003) "Understanding spatial relations: Flexible infants, lexical adults", *Cognitive Psychology* 46 (2003) 229–259
McDowell, J. (1994) *Mind and World* (Cambridge, Mass.: Harvard University Press).
Magee, B., and Elwood, R. W. (2013) Shock avoidance by discrimination learning in the shore crab (Carcinus maenas) is consistent with a key criterion for pain. *The Journal of Experimental Biology* 216, 353-358
Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the United States of America*, 97(8), 4398-4403.
Maki, W. S., Frogen, K., and Paulson, K. (1997) Associative priming by targers and distractors during rapid serial visual presentation: Does word meaning survive the attentional blink? *Journal of Experimental Psychology:Human Perception and Performance*, 23, 1014-1034
Margolis, E. (1999) 'How to acquire a concept', in E. Margolis, and S. Laurence (1999).
Margolis, E., and Laurence, S. (1999) *Concepts: Core Readings,* Cambridge, Mass.: MIT Press
Markman, E. M. (1989) *Categorization and Naming in Children: Problems of Induction*, Cambridge, Mass.: MIT Press
Marr, D. (1982) *Vision,* Cambridge, Mass.: MIT Press
Marr, D., & Nishihara, H. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London*, 200, 269-294.
Masataka, N. (2007) Music, evolution and language. *Developmental Science* 10(1):35–39.
Mates, B. (1950) 'Synonymity', in *Meaning and Interpretation, University of California Publications in Philosophy*, 25: 201-6, reprinted in L. Linksy, *Semantics and the Philosophy of Language* (Urbana: University of Illinois Press, 1952): pp. 111-38.
Medin, D., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.) *Similarity and analogical reasoning* (pp. 179–195). New York: Cambridge University Press.
Medin, D. L., and Schaffer, M. M. (1978). Context theory of classification. *Psychological Review*, 85, 207–238.
Melnyk, M., Shadrova, V., and Karwatsky, V. (2014) Towards Computer Assisted International Sign Language Recognition System: A Systematic SurveyInternational. *Journal of Computer Applications* (0975 – 8887) Volume 89 – No 17, March 2014 44
Michalski, R. S. (1980) "Knowledge acquisition through conceptual clustering: A theoretical framework and an algorithm for partitioning data into conjunctive concepts". I*nternational Journal of Policy Analysis and Information Systems* 4: 219–244.
Michalski, R. S.; Stepp, R. E. (1983) "Learning from observation: Conceptual clustering". In Michalski, R. S.; Carbonell, J. G.; Mitchell, T. M. (Eds.). *Machine Learning: An Artificial Intelligence Approach*. Palo Alto, CA: Tioga. pp. 331–363.
Mill, J. S. (1961) *A System of Logic,* Longmans
Miller, G., and Johnson-Laird, P. (1987) *Language and Perception* (Cambridge, Mass.: Belknap).
Millikan, R. G. (1984) *Language, Thought, and Other Biological Categories*. MIT Press.
_____, (1999) "Historical Kinds and the Special Sciences", Philosophical Studies 95: 45–65
_____, (2000) *On Clear and Confused Ideas.* Cambridge: Cambridge University Press.
Minsky, M. (1967). *Computation: Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, NJ.

_____, (1974), A Framework for Representing Knowledge, *MIT-AI Laboratory Memo* 306, June, 1974.
Mitkov, R. (2003/2004) (ed.) *The Oxford Handbook of Computational Linguistics*, (Oxford: Oxford University Press).
Mole, C. (2014) Dead Reckoning in the Desert Ant: A Defence of Connectionist Models. *Review of Philosophy and Psychology*, 2014; 5:277-290
Moshman, D. (2000). "Diversity in reasoning and rationality: metacognitive and developmental considerations". *Behavioural and Brain Sciences* 23: 689–690.
Murphy, D. (2006) 'On Fodor's Analogy', *Mind and Language,* 21/5, 553-64.
Needham, P. (2002) 'The discovery that water is H_2O', *International Studies in the Philosophy of Science*, 16: 3, 205 — 226
Neville, H. J. & Bavelier, D. (2002) Human brain plasticity: Evidence from sensory deprivation and altered language experience. In M.A. Hofman, G.J. Boer, A.J.G.D. Holtmaat, E.J.W. van Someren, J. Verhaagen and D.F. Swaab (Eds.) *Plasticity in the adult brain: From genes to neurotherapy*. (pp.177-188). Amsterdam: Elsevier Science.
Nosofsky, R. M. (1986). Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53–97
Onishi, K., and R. Baillargeon (2005) Do 15-Month-Old Infants Understand False Beliefs? *Science* 8 April 2005: Vol. 308 no. 5719 pp. 255-258
Osgood, C., and Tzeng, O. (1990) *Language, Meaning and Culture* (Westport, Conn.: Praeger).
Osherson, D., and Smith, E. (1981) 'On the adequacy of prototype theory as a theory of concepts', *Cognition,* 9: 35-58.
Osman, M. (2004). "An evaluation of dual-process theories of reasoning". *Psychonomic Bulletin & Review* 11 (6): 988–1010.
Oztop, E., Kawato, M., & Arbib, M. (2006) Mirror neurons and imitation: a computationally guided review. *Neural Networks*, 19(3):254 – 271
Pais, A. (1982). *Subtle is the Lord: The Science and the Life of Albert Einstein.* Oxford: Oxford University Press.
Parsons, K. P. (1973) Three Concepts of Clusters. *Philosophy and Phenomenological Research*, Vol. 33, No. 4 (Jun., 1973), pp. 514-523
Partee, B. (1995) "Lexical semantics and compositionality", in *Invitation to Cognitive Science,* 2nd edn., vol. I, ed. L. Gleitman and M. Liberman. Cambridge, Mass.:MIT Press
Pears, D. (1990) *Hume's System* (Oxford: Oxford University Press).
Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Joan L. Bybee and Paul Hopper, editors, *Frequency and the Emergence of Linguistic Structure*, pp. 137-157. John Benjamins, Amsterdam.
Perry, J. (1988) "Cognitive Significance and New Theories of Reference", *Noûs*, Vol. 22, No. 1, pp. 1-18
Peterson, M. A., & Rhodes, G. (Eds.) (2003) *Perception of Faces, Objects and Scenes: Analytic and Holistic Processes*. New York: Oxford University Press.
Pigliucci, M., and Kaplan, J. (2006) *Making Sense of Evolution* (Chicago, Ill.: University of Chicago Press).
Pinker, S. (1994) *The Language Instinct*, Penguin Books
_____, (1995) Language Acquisition. *Language: An Invitation to Cognitive Science*, Cambridge, Mass.:The Mit Press, 135-182
_____, (1998) Excerpts: *How the Mind Works*, retrieved from *http://www.cse.iitk.ac.in/users/amit/books/pinker-1998-how-mind-works.html*

_____, (1999) *How the Mind Works*, W. W. Norton and Company
_____, (2005) 'So how does the mind work?', *Mind and Language,* 20/1: 1-24.
Pinker, S., and Mehler, J. (1989) (eds.) C*onnections and Symbols,* (Cambridge, Mass.: MIT Press).
Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-194.
Place, U. T. (1956) "Is Consciousness a Brain Process?" *British Journal of Psychology*, 47: 44–50. doi: 10.1111/j.2044-8295.1956.tb00560.x
Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language*, 55, 317-351
Poeppel, D. and D. Embick (2005) 'Defining the relation between linguistics and neuroscience' (2005) In A. Cutler ed. *Twenty-first century psycholinguistics: Four cornerstones*, Lawrence Erlbaum
Pollack, J. B. (1990) Recursive Distributed Representations. *Artificial Intelligence*. 46: 77-105
Pond, R. (1987) Fun in metals, *Johns Hopkins Magazine*, April 1987, 60-68
Premack, D. (1989) Language, Nonhuman. In D. Kimura (ed.) *Speech and Language*. Pages 7-8. *Readings from the Encyclopedia of Neuroscience*, Boston: Birkhäuser
Prentice, D. A., & Miller, D. T. (2007) Psychological Essentialism of Human Categories, *Current Directions in Psychological Science,* August 2007 vol. 16 no. 4 202-206
Prince, A., and Smolensky, P. (1997) "Optimality: From Neural Networks to Universal Grammar", *Science*, Vol 275, 14 March 1997
Prinz, J. (2002) *Furnishing the Mind* Cambridge, Mass.: MIT Press)
Pulvermuller, F. (1999) Words in the Brain's Language, *Behavioral and Brain Sciences*, 22, 253-336
Pustejovsky, J. (1995) *The Generative Lexicon.* Cambridge, Mass.:MIT Press.
Putnam, H. (1954) "Synonymity and the Analysis of Belief Sentences", *Analysis*, April, 1954, pp. 114-122.
_____, (1970) "Is semantics possible?'", *Philosophical Papers, Vol. 2: Mind, Language and Reality*, Cambridge University Press.
_____, (1973a) "Explanation and Reference'", *Philosophical Papers, Vol. 2: Mind, Language and Reality*, Cambridge University Press.
_____, (1973b) "Meaning and Reference", *The Journal of Philosophy*, Vol. 70, No. 19, Seventieth Annual Meeting of the American Philosophical Association Eastern Division (Nov. 8, 1973), pp. 699-711
_____, (1975) "The Meaning of 'Meaning'", *Philosophical Papers, Vol. 2: Mind, Language and Reality*, Cambridge University Press.
_____, (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.
_____, (1982). 'Why there isn't a ready-made world' in *Synthese*, Vol. 51, No. 2, May.
_____, (1983a) "Two Dogmas Revisited", *Philosophical Papers, Vol. 3: Realism and Reason*, Cambridge University Press
_____, (1983b) ""Vagueness and Alternative Logic," in H. Putnam, *Realism and Reason*, Cambridge: Cambridge University Press.
_____, (1990) "Is Water Necessarily H_2O?", *Realism with a Human Face*, J. Conant (ed.), Cambridge University Press
_____, (2000) *The Threefold Cord: Mind, Body and World* (New York: Columbia University Press)
_____, (2004) *Ethics Without Ontology*, Cambridge: Harvard University Press
Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, 22, 253-336.

Pylyshyn, Z. (1984) *Computation and Cognition: Towards a Foundation for Cognitive Science*, (Cambridge, Mass.: MIT Press).
_____ (1987) (ed.) *The Robot's Dilemma,* Norwood, NJ: Ablex
_____ (1999) Is Vision Continuous With Cognition? The Case For Cognitive Impenetrability Of Visual Perception
_____, (2003) *Seeing and Visualizing,* Cambridge, Mass.: MIT Press
Pylyshyn, Z., and Demopoulos, W. (1986) (eds.) *Meaning and Cognitive Structure: Issues in the computational theory of mind*, (New Jersey: Ablex Publishing Corporation).
Pylyshyn, Z., and Bannon, L. J. (1989) (eds.) P*erspectives on the Computer Revolution,* (New Jersey: Ablex Publishing Corporation).
Quine, W. V. O. (1951) "Two Dogmas of Empiricism," *The Philosophical Review* 60: 20–43. Reprinted in:
_____, (1953) *From a Logical Point of View*. Harvard University Press
_____ (1960). *Word and Object*. Cambridge, Mass:MIT Press
_____, (1969) Natural Kinds. in *Ontological Relativity and Other Essays*: Columbia Univ. Press.
Ramsey, W., Stich, S. and Garon, J. (1990), "Connectionism, Eliminativism, and the Future of Folk Psychology", *Philosophical Perspectives* 4: 499-533.
Reber, A. S. (1993) *Implicit Learning and Tacit Knowledge*. Oxford, UK: Oxford Univ. Press
Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, 3, 382–407.
Rees, G; Russell, C., Frith, C., and Driver, J. (1999). "Inattentional blindness versus inattentional amnesia for fixated but ignored. words". *Science* 286 (5449): 2504–2507.
Reid, T. (1764) *An Inquiry into the Human Mind*, D.R. Brooks, ed., Edinburgh: Edinburgh University Press (1997)
_____, (1983) *Thomas Reid's Inquiry and Essays* (Indianapolis, Ind.: Hackett).
Rey, G. (1981), "Introduction: What Are Mental Images?" in Block 1981: 117-127.
_____, (1991), "Sensations in a Language of Thought," in E. Villaneuva, ed., *Philosophical Issues 1: Consciousness*, Atascadero: Ridgeview Publishing Company: 73-112.
Rhees, R. (1963) 'Can there be a private language?', in C. Caton (ed.) *Philosophy and Ordinary Language* (Urbana, Ill.: University of Chicago Press).
Rizzolatti, G., & Arbib, M. (1998). "Language is within our grasp. *Trends in Neurosciences*. 21(5):188–194
Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27:169–192
Roberts, M. J. (2002) The elusive matching bias effect in the disjunctive selection task. *Exp Psychol.*;49(2):89-97.
Roelofs. A. (1992) A spreading-activation theory of lemma retrieval in speaking. *Cognition*.42, 107· 142
Rosch, E. (1973a) 'On the internal structure of perceptual and semantic categories', in T. Moore (ed.) *Cognitive Development and the Acquisition of Language* (New York: Academic).
_____, (1973b) Natural categories. *Cognitive Psychology*, 4, 328-350.
_____, (1975) Cognitive Representation of Semantic Categories. *Journal of Experimental Psychology* v. 104:192-233
_____, (1978) "Principles of Categorization." in E. Rosch and B. Lloyd. Eds.. *Cognition and Categorization,* 1978, Lawrence Erlbaum Associates, Inc., Publishers, reprinted in Margolis & Lawrence (1999, Chapter 8)
Rosch, E., Mervis, C. B., Gray, W., Johnson, D., and Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.

Rouder, J. N., and Ratcliff, R. (2006) "Comparing Exemplar and Rule-Based Theories of Categorization". *Current Directions in Psychological Science* 15: 9–13
Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. I: Foundations*. Cambridge, Mass.: MIT Press
Russell, B. (1905) "On Denoting." in A. P. Martinich, *The Philosophy of Language*, Oxford University Press, 1976
Ryle, G. (1949) *The Concept of Mind* (London: Hutchinson).
Salmon, N. (1979) "How Not To Derive Essentialism from the Theory of Reference," *The Journal of Philosophy*, LXXVI, 12 (December 1979):703-725
_____, (1982) *Reference and Essence*. Oxford: Basil Blackwell
Samuels R. (2006) The magical number two, plus or minus: some comments on dual-processing theories of cognition. In *Two Minds: Dual Process Theories of Reasoning and Rationality*. 2006.
Schilpp, P. (Ed.) (1979) *Albert Einstein: Autobiographical Notes*. La Salle, Ill: Open Court.
Schmidt, R.A. (1988). *Motor Control and Learning: A Behavioral Emphasis*. 2nd ed. Champaign, IL: Human Kinetics.
_____, (1991). Motor learning principles for physical therapy. In: *Foundation for Physical Therapy. Contemporary Management of Motor Control Problems: Proceedings of the II-STEP Conference*. Alexandria, VA: Foundation for Physical Therapy.
Searle, J. R., (1980) "Minds, Brains, and Programs", *Behavioral and Brain Sciences*, 3(3): 417-457
_____, (1983) *Intentionality*, Cambridge: Cambridge University Press
Seidenberg, M.S. (1994) Language and Connectionism: the developing interface.*Cognition* 50 (1994) 385-401
Sellars, W. (1955) 'Putnam on Synonymity and Belief', *Analysis* 15 (1955): 117-120
_____, (1956) 'Empiricism and the philosophy of mind', in *Minnesota Studies in the Philosophy of Science,* 1 (Minneapolis, Minn.: University of Minnesota Press).
Silby, B. (2000) *Revealing the Language of Thought*, An e-book by BRENT SILBY This paper was produced at the Department of Philosophy, University of Canterbury, New Zealand Copyright © Brent Silby 2000
Sipser, M. (2006), *Introduction to the Theory of Computation: Second Edition*, Thompson Course Technology div. of Thompson Learning, Inc. Boston, MA.
Shedden, J. M., & Schneider, W. (1991) 'A connectionist simulation of attention and vector comparison: The need for serial processing in parallel hardware.' ln *Proceedings of the Thirteenth Annual conference of the cognitive science society*. chicaqo (pp. 546-551). Hillsdale, NJ: Lawrence Erlbaum.
Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., and Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101(3):260-277.
Sokolik, M.E. (1990) 'Learning without rules: PDP and a Resolution of the Adult Language Learning Paradox' T*esol Quarterly* vol 24, no. 4 pp 685-696.
Sokolik, M.E., and Smith, M. E. (1992) Assignment of gender to French nouns in primary and secondary language : a connectionist model. *Second Language Research* February 1992 vol. 8 no. 1 39-58
Slawinski, E. G. and Fitzgerald, L. K (1998) Perceptual development of the categorization of the /r-w/ contrast in normal children. *Journal of Phonetics* 26, 27-44.
Smaers, J. B., Schleicher, A., Zilles. K., Vinicius, L. (2010) Frontal White Matter Volume Is Associated with Brain Enlargement and Higher Structural Connectivity in Anthropoid Primates. *PLoS ONE* 5(2): e9123. doi:10.1371/journal.pone.0009123

Smith, E. (1995) Concepts and Categorization. In *Thinking: An invitation to Cognitive Science*, Vol. 3, second edition, E. Smith and D. Oshershon (Eds.), Cambridge MA:MIT Press, pp. 3-33
Smith, E., and Medin, D. (1981) *Categories and Concepts* (Cambridge, Mass.: Harvard University Press).
Smith. E., Osherson. D., Rips, L, and Keane, M. (1988) "Combining Prototypes: A Selective Model" *Cognitive Science 12* © 1988 by Ablex Publishing Corporation
Smolensky, P. (1988) 'On the proper treatment of connectionism', *Behavioral and Brain Sciences, 11*: 1-23.
_____, (1990) Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems. *Artificial Intelligence*, 46: 159-216
_____, (1995) 'Connectionism, Constituency and the Language of Thought', in C. Macdonald and G. MacDonald, (1995).
Soames, S. (1998) 'The Modal Argument: Wide Scope and Rigidified Descriptions', *Noûs* 32:1, pp. 1-22
Soon C.S., Brass M., Heinze H.J., and Haynes J.D. (2008) Unconscious determinants of free decisions in the human brain. *Nat Neurosci.* 2008 May;11(5):543-5
Spelke, E. (1994) 'Initial Knowledge: Six Suggestions'. In *Cognition*, 50, pp. 431-445
Sperber, D. (1997) 'Intuitive and reflective beliefs', in *Mind and Language* 12 (1) (1997). pp. 67-83
_____, (2002) 'A Defense of Massive Modularity', in Depoux (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* (Cambridge, Mass.: MIT Press).
Sperling, G. (1960) 'The Information Available in Brief Visual Presentations', *Psychological Monographs,* 74: 1-29.
Stanford, P. K, and Kitcher, P. (2000) Refining the causal theory of reference for natural kind terms. *Philosophical Studies* 97 (1):97-127
Stanovich, K. E. (1999) Who is Rational? *Studies of Individual Differences in Reasoning*. Mahway, NJ:
Lawrence Elrbaum Associates
Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding/acceptance principle. *Cognitive Psychology*, 38, 349–385.
_____, (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–726.
_____, (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13, 225–247.
_____, (2008a). On the failure of intelligence to predict myside bias and one-sided bias. *Thinking & Reasoning*, 14, 129–167.
_____, (2008b). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94, 672–695.
Sterelny, K. (2003) *Thought in a Hostile World: The Evolution of Human Cognition.* Blackwell Publishing
Stiles, J., Reilly, J., Paul, P. & Moses. P. (2005): Cognitive development following early brain injury: evidence for neural adaptation, *Trends in Cognitive Sciences*, 9(3)
Stich, S. (1983) *From Folk Psychology to Cognitive Science*, Cambridge, Mass.: The MIT Press.
Suits, B., with Godine, D. R. (1978) *The Grasshopper: Games, Life and Utopia* (University of Toronto Press

Sun, R., Slusarz, P.,Terry. C. (2005) The interaction of the explicit and the implicit in skill learning: a dual-process approach. *Psychol. Rev*. 112(1):159–92
Tinbergen, N. (1951) *The study of instinct*. Oxford University Press
Tomasello, M., Call, J., and Hare, B. (2003) Chimpanzees understand psychological states: The question is which ones and to what extent. *Trends in Cognitive Scienc*e 7:153–156
Tooby, J., and Cosmides, L. (1992) The psychological foundations of culture. In Barkow, Cosmides, & Tooby (1992), pp. 19–136
Treisman, A., and Schmidt, H., 'Illusory Conjunction in the Perception of Objects', *Cognitive Psychology,* 14/1: 107-42
Tversky, A. (1977) 'Features of Similarity', *Psychological Review*, 84(4), 327-352
Tye, M. (1991), *The Imagery Debate*, Cambridge, Mass.: The MIT Press.
_____, (1995) *Ten Problems of Consciousness: a Representational Theory of the Phenomenal Mind*, (Cambridge, Mass.: MIT Press).
_____, (2000), *Consciousness, Color, and Content*, Cambridge, Mass.: The MIT Press.
Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., et al, (2001). I know what you are doing: A neurophysiological study. *Neuron*, 31(1):155-165
Velman, M., and Schneider, S. (2007) (eds.) *The Blackwell Companion to Consciousness*, (Blackwell Publishing)
van Gelder, T., 1990, Compositionality: A connectionist variation on a classical theme, *Cognitive Science*, 14:355-384.
van Hateren, J.H., Srinivasan, M.V., and Wait, P.B. (1990) Pattern recognition in bees: orientation discrimination. Journal of Comparative Physiology A, 167:649-654
von Neumann, J. (1956/2000) *The Computer and the Brain*. New Haven, CT:Yale University Press.
Warren, R.M. (1970). "Restoration of missing speech sounds". *Science* 167 (3917): 392–393.
Wegner, D.M. and Wheatley, T.P. (1999) Apparent mental causation: sources of the experience of will. *Am. Psychol.* 54, 480–492
Wegner, D. M. (2002) *The illusion of conscious will*. Cambridge, MA: MIT Press.
_____, (2003) The mind's best trick: how we experience conscious will. In *TRENDS in Cognitive Sciences* Vol.7 No.2 February 2003
Weiskrantz, L. (1986) *Blindsight: A Case Study and Implications*. Oxford University Press
Werker, J. F. & Tees, R. C. (2005) Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain, 2005 Wiley Periodicals, Inc., *Dev Psychobiol*, 46: 233–251, 2005.
Wiggins, D. (1967) *Identity and Spatio-temporal Continuity*, Oxford: Blackwell
_____, (1980) *Sameness and Substance.* Oxford: Blackwell
Wilson, T. D., Lindsey, S., Schooler, T. Y. (2000) A model of dual attitudes. *Psychol. Rev.* 107(1):101–26
Wilson, T. D., Schooler, J. W. (1991) Thinking too much: introspection can reduce the quality of preferences and decisions. *J. Personal. Soc. Psychol*. 60(2):181–92
Witelson, S. F., Kigar, D. L., and Harvey, T. (1999) The exceptional brain of Albert Einstein. *Lancet* 1999; 353: 2149–53
Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford: Basil Blackwell
_____, (2007) *Zettel: 40th Anniversary Edition*, University of California Press
Wright, C., and Hale, B., (1996) (cds.) *A Companion to the Philosophy of Language* (Oxford: Blackwell).
Xu, F., and Spelke, E. S. (1999) Large number discrimination in 6-month-old infants. *Cognition* 74 (2000) B1±B11

Yamauchi, T., Love, B. C., & Markman, A. B. (2002) Learning Nonlinearly Separable Categories by Inference and Classification, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2002, Vol. 28, No. 3, 585–593
Zajonc, R. B. (1980) Feeling and Thinking: Preferences Need no Inferences. *American Psychologist*, 35:151-175
Zelazo, P. D., Moscovitch, M., & Thompson, E. (2007) (eds.) *The Cambridge Handbook of Consciousness* (Cambridge University Press)
Zhang, Y., Segalowitz, N., and Gatbonton, E. (2010) "Topological spatial representation across languages and within language: IN and ON in Mandarin Chinese and English." *The Mental Lexicon*, Vol. 6, No 3, 414-445(32)
Zuberbühler, K. (2005) The phylogenetic roots of language: Evidence from primate communication and cognition. *Current Directions in Psychological Science*, 14(3):126-130.
Zwicky, A. M., & Sadcock, J. M. (1975) Ambiguity tests and how to fail them, *Syntax and Semantics,* Vol. 4, (J. P. Kimball, ed.), Academic Press, New York, 1-36