****

THE NATURALISATION OF INTENTIONALITY AND RATIONALITY USING SYSTEMS

A Functional Explanation of Mind, Agency and Intentionality in Human Linguistic Communities.

****

William Alfred Cameron

M. Litt.
Glasgow University, Dept. of Philosophy
September 1998

Abstract

This thesis recognises two axioms of materialism. Firstly, that the human or other agent is within and is comprised of the same stuff as, a universe made up of material things, each of which is, in principle, explicable in materialist terms. Secondly, that the theorist is himself an agent, explicable within the theory of agency that he propounds. The author contends that any attempt to explain complex conscious human agency 'from the top down' faces either a potential regress of explaining the aetiology of human agency in terms of some agency of design or the view, canvassed by Colin McGinn, (1989), that the human mind is 'cognitively closed' to the concepts that would explain how human consciousness can arise from the material substance of the brain.

The author has avoided this dilemma by postulating an austere characterisation of agency, from which the rich and manifold nature of human agency and intentionality has developed by the accidents of evolution. He holds that this austere agency may be explained by natural accidents of chemical combination that have led, by the accidents of evolution, to the phenomena of reproductive life, functionally characterised in this explanation by agency and autopoiesis. This austere characterisation of agency is an example of a functional system. Agency is a capacity of an entity within the physical system of an agent in the world. This capacity is enabled by functions that the author has named; 'perception', 'representation', 'cognitive process' and 'action'. Through the processes of perception, physical states of the world physically cause changes in representational states of agency; within cognitive processes, representational states combine to cause actions of agency that change states of the world, that includes the agent, in ways that maintain or tend towards those goal states of the agent in the world, that are postulated in a theory of agency.

It is argued that this concept of agency is functionally isomorphic with the technological concept of regulation. Two theorems from regulation are particularly relevant. Firstly, Ashby's theorem that for successful regulation the variety of possible states in the regulator must at least match the variety of states regulated against. Secondly, the very idea of regulation stems from the epistemic contingency for the agent, of events regulated against. Life, as we know it on earth, is reproduced by reproductive behaviour that follows and reproduces the programmes encoded in DNA. The autopoietic maintenance of the structure of the living organism against the contingencies of an unpredictable world is enabled by the mechanisms of agency.

The structure of the thesis and the ontological commitments of the author are set out in a first introductory chapter. In the second chapter the author summarises the history and currant range of application of the system concept and describes the philosophical implications of his notion of a physical system. The notions of physical cause, accident, function, supervenience, representation and alternative realisation that are assumed within the thesis are also described in this chapter. The third chapter is devoted to the development of the concept of agency as a capacity, characterized by goals and intentionality and enabled by the functions listed above. Examples of agency in the world are described in the fourth chapter. These range from the simple reflex agency of a governor, unicellular organism or part of a plant to the complex integrated agency of production control systems, advanced vertebrates, including us, and social groupings such as a colony of social insects or some aspects of a human corporation. Also, within this chapter, the author considers the impact of language on human social agency, the implications of social agency for the attribution of personhood and through semantic ascent, the social practices of attribution of meaning, truth and mind, and the propositional attitudes. He concludes that, since agency necessarily involves an agent in its world and human language is about the world as it is for the human agent; language, agency and the world are explanatorily inseparable.

In the fifth and sixth chapters the author applies his theory of human agency to the computational theory of mind and the apparent tension between determinism, free will and personal responsibility. The author concludes: Firstly, that the brain as an organ of representation, is not a computer since computation is an act of agency, although parts of the brain may have a combinatorial function within such acts. Secondly, if freedom is defined as an absence of physical constraint then a free agent is physically responsible for its acts. Within the social practice of attribution of personhood to the continuous ongoing agent within the community, each person is held responsible for his actions, including those that change the future agency of himself and others, for better or for worse, according to the valuation of the community. In a final chapter the author summarises some of the philosophical implications of his thesis. The notions of variety in regulation and of autopoiesis as a necessary criterion of life are used in the thesis and are explained in each of two appendices.

ACKNOWLEDGEMENTS

Acknowledgement is due to Jim Edwards without whose encouragement this project would never have been started and whose unfailing generosity of time and advice has enabled its completion.

Also, to my wife whose patience and common sense have kept my feet near to the ground.

CONTENTS

1 INTRODUCTION

1.1 SUMMARY OF THESIS 1
1.2 OUTLINE OF THESIS 2
1.3 NOT A PHYSICALIST THESIS 3
1.4 CAUSE AND ACCIDENT 3
1.5 THE PLACE OF BIOLOGY 4
1.6 SYSTEM DESCRIPTION 4
1.7 INTENTIONALITY 5
1.8 AGENCY 6
1.9 REPRESENTATION 8
1.10 LEVELS OF AGENCY 9
1.11 TELEOLOGY AND GOALS 10

2 HISTORY & METAPHYSICS OF SYSTEM

2.1 INTRODUCTION 13
2.2 PHYSICAL SYSTEMS 14
2.3 FORMAL SYSTEMS 20
2.4 SYSTEM AS A MODE OF DESCRIPTION 21
2.5 THE COUNTERFACTUALITY OF SYSTEM DESCRIPTION 23
2.6 CAUSE 26
2.7 FUNCTION 35
2.8 SYSTEM AND SUPERVENIENCE 37
2.9 ALTERNATIVE REALISATION 42
2.10 MILLIKAN FUNCTIONS 44
2.11 THE FUNCTION OF REPRESENTATION 45
2.12 MALFUNCTION AND MISREPRESENTATION 51
2.13 CONCLUSIONS 54

3 SYSTEM, INTENTIONALITY & AGENCY

3.1 INTENTIONALITY 56
3.2 AGENCY 59
3.3 FUNCTIONS OF AGENCY 62
3.4 GOALS OF AGENCY 64
3.5 REPRESENTATION WITHIN AGENCY. 70
3.6 THE CONTENT OF REPRESENTATIONS WITHIN AGENCY 72
3.7 MISREPRESENTATION WITHIN AGENCY 79
3.8 AGENCY, REGULATION AND CYBERNETICS 81
3.9 CONTINGENCY 83
3.10 VARIETY 86
3.11 REALISATIONS OF AGENCY 88
3.12 CONCLUSIONS 90

4 LEVELS OF AGENCY

4.1 INTRODUCTION 93
4.2 THE AGENCY OF SIMPLE ORGANISMS 96
4.3 THE AGENCY OF ARTEFACTS 98
4.4 MORE COMPLEX ARTEFACTS 107
4.5 MORE COMPLEX ORGANISMS 109
4.6 LEARNING AND ADAPTATION 119
4.7 SOCIAL AGENCY 123
4.8 THE AGENCY OF SOCIAL GROUPS 126

4.9 An Example of Social Agency 130
4.10 Personhood 135
4.11 The Impact of Language 135
4.12 A Theory of Mind 142
4.13 Hypothesis for the Status of Beliefs 144
4.14 Support for the Hypothesis from the Supervenience of Intrinsic and Conventional Representations 146
4.15 Conclusions 149

5 AGENCY & THE COMPUTATIONAL THEORY OF COGNITION

5.1 Introduction 152
5.2 Computation 153
5.3 Representation by Convention 163
5.4 Functions of Conventional Representation 167
5.5 Conventional Representations that Misrepresent 167
5.6 Cognition as Computation? 172
5.7 Conclusions 183

6 DETERMINISM & AGENCY

6.1 The Problem of Materialism 185
6.2 The Determinist Debate 187
6.3 Cause and Determination 190
6.4 Regulation 194
6.5 Actions of Agency 199
6.6 System Descriptions of the Temporal Agent 203
6.7 The Person in the Community 207
6.8 Attributions of Freedom 213
6.9 Attributions of Choice 215
6.10 Attributions of Responsibility 218
6.11 Spontaneity and Responsibility 221
6.12 Conclusions 225

7 IMPLICATIONS 231

APPENDIX A

AUTOPIESIS, ALLOPOIESIS, AETIOLOGY AND AGENCY:

CRITERIA FOR LIFE

A1.1 The Concept of Autopoiesis 236
A1.2 Criteria for Autopoiesis 237
A1.3 The Place of the Observer 239
A1.4 Autonomy and Function 239
A1.5 The Capacity of Autopoiesis 241
A1.6 The Complementarity of Autopoiesis and Agency 242
A1.7 Teleology and Aetiology 244

APPENDIX B

REQUISITE VARIETY 246

GLOSSARY 254

BIBLIOGRAPHY 260

1

INTRODUCTION

1.1 Summary of Thesis

In this thesis I shall develop an account of agency that is based on a system description of the behaviour of an entity in an environment as action for the goals of the entity consequent on perception of the environment. This differs from standard accounts of agency principally in two respects.

Firstly, previous accounts have explained actions in the world by reference to desires and knowledge that are states of the agent and some capacity in the agent to gain beliefs from perception and choose action appropriate to beliefs and desires. I claim that these accounts leave desire as an explanatory dangler. By explaining agency as functionally isomorphic to regulation, action is explained as a functional component in a causal regulatory process between an entity and its environment. Regulation depends on goals and I explain the goals postulated in a theory of natural agency as the result of the accidents of evolution by which each organism is so structured that it is apt for system description in terms of agency. If regulation and agency are functionally isomorphic then our artefacts of regulation are also apt for system description in terms of agency.

Secondly, my concept of agency extends to simple organisms and artefacts of regulation. An austere definition of agency in system terms leads to a functional specification that is realised in a vast range of entities familiar in our world.

This is an explanation of mind and language from the bottom up. The argument is aetiological and can be summarised in eight steps.

(i) There is a material universe of stuff constrained by physical laws

(ii) The aetiology of natural agency is accidental and evolutionary.

(iii) Acts of agency are concerned with what is contingent for the agent

(iv) Agency is realised in the material stuff of this universe

(v) By its acts each agent divides the material stuff into a world of *things* for it.

(vi) Evolution of the cooperative social agency of human kind has led to languages that denote *things* and the practices of reference to *things* that are *for* humans.

(vii) Semantic ascent within language has enabled human practices of attribution of propositional attitudes and an articulated theory of mind.

(viii) Human theorising has led to a reification of mind and the invention of the mind-body problem.

1.2 Outline of Thesis

This thesis is in three parts. Firstly, a methodology of explanation based on the notion of system is developed. Secondly, this methodology is applied to an analysis of agency, as the capacity for intentional action in the world. Thirdly, the value of this analysis is demonstrated by its application to two problem areas; the computational model of cognition and the tension between the notions of determinism and responsibility. Chapter 2 deals with the system methodology of explanation. The application of this methodology to agency and intentionality is considered in Chapters 3 and 4. The two problem areas are considered in Chapter 5 and 6. The system concepts of autopoiesis and requisite variety are used in the thesis and are explained in two appendices.

1.3 Not a Physicalist Thesis

A principal aim of the thesis is to establish a methodology which will enable self-consistent explanation of ourselves as organisms, individuals and members of our various social and linguistic communities without postulating any stuff separate from the varieties of physical stuff studied by our physical sciences or any causal agency which is not within the purview of the physical sciences. This makes me a monist and some kind of a naturalist, but not a physicalist. Wilkes, (1978, 1), explicates the 'physical' as " picking out all and only the items, processes, concepts, laws, hypotheses, theories or theoretical postulates used essentially by physical scientists". Like Wilkes, I take as axiomatic that "... the physical sciences which study the brain and human activity can have or devise a vocabulary adequate for the identification of any physical event, state or process." However, in my own way and for my own reasons which will emerge in the course of this thesis, I recognise functional theories based on the notion of a system and these theories are not reducible to physical theories, hence I reject physicalism. In a narrative description singular physical events have singular causal explanations but the teleological implications of a functional explanation within a system description may preclude reduction, in the Nagel, (1961), usage, of function-theoretical terms to the theoretical terms of physical science. However, this system description is consistent with a background assumption that none of the entities or events that realise a system contradict the laws of physical science. As will emerge, I am a realist about the representational states of agency and an instrumentalist about the goals of agency. Propositional attitudes such as belief are to be interpreted instrumentally, although states of attribution of each of the attitudes are real states of the linguistic community within which they are attributed.

1.4 Cause and Accident

My notion of 'cause' is physical and explanatory. I assume a Humean account of causality. Certain phenomena show a constant conjunction and if two events are causally linked in this way then; if one event had not been the second never had existed. Natural events, those lacking agent cause, are generally linked by various phenomena and causal explanation is easier in principle than in practice. However, over the billions of years between the big bang and the advent of life on this planet, events led to events and were susceptible to physical causal explanation had we been around to do it. I take no view as to whether the world, as described by the 'true physics' is deterministic or not. The actions of agency are responses to events that are epistemically contingent for the agent. Although many events that we note in our world today bear the mark of agent cause, the very fact of agency is the result of accidental events, constrained by physical law.

1.5 The Place of Biology

My concept of natural agency is within the explanatory scheme of biology. Many writers have contended that the concepts of biology are not reducible, although no entities discussed by biologists are inconsistent with the laws of physics and this is the stance that I will assume. Since my notion of agency includes both biological agents and artefacts of regulation, it is unsurprising that that it extends continuously throughout the phylogenetic tree to primitive and complex organisms and also, within a complex agent, whether natural or artefact, to its regulatory supporting structures.

1.6 System Description

System descriptions are essentially counterfactual. A system description picks out a set of possible worlds. Its use is to explain how the range of outcomes is constrained by the theories, components and boundaries which define the entity described. The notion of system is an important tool in this thesis and a chapter is devoted to the history, metaphysics and application of this notion.

Systems may be described in terms of abstract functions, which, if realisable, may be capable of alternative realisations. Alternatively, a system description may be of a specific structure of actual physical components. Such a structure may be apt for characterisation in functional terms and if so, these functions may be realisable in alternative physical structures.

In Chapter 2, I describe the history of the system concept and how this has led to the application of system thinking to explanation and intervention within complex human organisations and artefacts. Within this chapter, I explain my own specific usage of 'system' as a mode of description of an entity which is characterised by the counterfactuality of event and outcome. This usage requires precise definitions of the notions of cause and function and these are given. In a causal explanation, effect follows cause within a flow of events constrained by physical laws. In a functional explanation, certain phenomena are seen as apt for explanation in terms of the capacities of entities. Such an entity may be described as comprised of components, which have functions that enable the capacities of the entity. Although the capacities of any given entity may be explained uniquely in terms of specific functions that are realised in specific structures; for any given function, alternative physical realisations may exist.

1.7 Intentionality

Brentano, (1874), described the objects of mental phenomena, the components of our desires and fears as 'intentional objects'. His contention was that: since these objects relate to propositions that are immaterial entities, mental properties must be irreducible and materialism must be false. By this thesis, intentionally is the hallmark of the mental as manifest in articulate thought and much of current philosophy of mind and language is concerned with the implications of this assumption. I will argue that this assumption draws a false distinction between man and other living creatures. Man recognises and describes objects that are intentional for him and distinguishes between living things and artefacts of regulation. Animals with nervous systems recognise objects that are intentional for them. All organisms and regulatory artefacts behave in ways which man can describe as regulatory, goal seeking and intentional: by their behaviour such entities classify objects that are intentional for them.

1.8 Agency

In Chapter 3, I discuss the interrelated concepts of agency, intentionality, perception, representation, goal and action. Although my usage of these concepts is close to that current in the philosophy of mind and cognition, some of this usage is technical. In particular, I refine the concept of human agency to an austere agency that is a capacity of an entity to react to the events of an unpredictable environment so as to further its goals.

I follow Davidson in defining agency as characterised by action that is intentional under a description and from this I derive a functional characterisation of agency in terms of perception, representation, goal and action. By this austere definition, any artefact that displays regulatory behaviour can be described as having agency. In the terms of Dennett, (1987, 15), it is an intentional system.[1]

Within a theory of agency; perception, representation, cognitive process and action are functions of the entity that enable its agency. By perception, states of the world change the representational states of the agent; through cognitive

[1] What Dennett actually writes is: "What it is to be a true believer is to be an intentional system, a system whose behaviour is reliably and voluminously predictable via the intentional strategy." Dennett is however, an instrumentalist regarding the attitudes. Whether Dennett is a realist about systems is unclear. The distinctions between Dennett's position and my own will emerge.

processes, representational states combine to cause learning and action. Action is change in the agent that changes the world for the goals of the agent. Perception is the process by which representational states are maintained as representations of the world for the agent. Representational states are about the world for the agent; as the world is presented to the agent so these states re-present the world for the agent. Goals are postulates within a theory of agency that explain how the actions of agency are *for* the agent; they are possible worlds picked out by the system that describes the agency of this agent.

This austere concept of agency can be compared with an equally austere concept of regulation by which entities that may be organisms or artefacts act towards their goals in an unpredictable environment. Cybernetics is the discipline concerned with the theory of regulating mechanisms. If agency is equivalent to regulation then the explanatory tools of cybernetics are explanatory for agency. If regulation is counter entropic then agency is counter entropic. Each act of agency is determined since it follows from the state of the world and the representational and states and goals of the agent but each state of the world is unpredictable for the agent; if it were otherwise it would be programmed. Thus, each act of agency is a spontaneous response to an event that is epistemically contingent for that agent. Successful regulation requires requisite variety. If an organism has evolved to survive in virtue of its agency then its representational variety will match the variety of states in the world that are relevant to its survival.

My naturalisation of agency and intentionality is based on the thesis that the very notion of intentionality is implicit in the regulatory behaviour of living things and artefacts of regulation. Several authors have contributed to the thesis that the maintenance and behaviour of living things is essentially cybernetic[2]. A

[2] Cannon, (1932), Ashby (1956), Sommerhoff, (1950), Wiener, (1961), and Nagel, (1979), to name but a few.

separate application of systems to biology has been the work of Maturana on cognitive systems leading to his collaboration with Varela in describing the autopoietic account of life as self-maintenance. Cybernetics is the science of control and in organisms or artefacts, control can be homeostatic or heterotelic, Sayre, (1976, 52-54). Homeostatic control is concerned with internal maintenance of states of an entity by variation within the entity. Heterotelic control is concerned with the control of states of an entity by action in an environment.

By our current physical theories, our universe is characterised by entropy. Although the closed organisation by which living things maintain their identity is apt for description in terms of autopoiesis, the open organisation by which living things sustain their autopoiesis by maintaining essential relations with their environment must be described in cybernetic terms and this maintenance is counter entropic.

1.9 Representation

The concept of representation is central to the theory of agency, which iis the central topic of chapter 3. The concept of representation is introduced in Chapter 2, where the intrinsic and non-conventional representations of agency are distinguished from the extrinsic conventional representations of language and symbol. The content of an intrinsic representation is attributed within a theory of agency. Such theories will be within some domain of description chosen by the theorist and a description of the content of a representation within agency will be comprised of concepts, within this domain and possessed by the theorist, that are of whatever is represented. If such a domain is realistic, in that the description is of a world real for the theorist, then bridging laws available to the theorist will link alternative physical descriptions, also available to the theorist. Alternatively, functional domains will include functional objects, real for the theorist. Either of these domains of description may be used to describe and explain the same instance of agency.

Different theories will invoke different representations that may be differently realised but if an entity is apt for description in these different theories of agency then cognitive variety of representation will be available to realise each of the representations invoked. However, the content of all such representation is described in concepts of the theorist, and these are not, in general, concepts of the agent.

Concepts of the theorist are abstracta realised in the practices of language and in language the terms that denote concepts are representations .by convention The distinction between representations and misrepresentations within agency is explored in Chapter 3. Chapter 4 contains examples of natural and artefactual agency that illustrate misrepresentation within agency. Representation by convention and associated misrepresentation is considered in Chapter 5. The distinction between conventional representations and representations within agency is pertinent to the computational theory of cognition which is considered and opposed in Chapter 5.

1.10 Levels of Agency

Using Ashby's, (1956), notion of requisite variety, I suggest that the development of complexity in life can be seen as an evolutionary response to the possibilities of complexity inherent in the natural world. This can then be described as an increased complexity of those objects that may be defined, within our system description of agency, as intentional for the creature. These objects comprise the world of the creature that recognises them. By 'world' I do not mean planet Earth, although for all of us our world is within it. We are complex creatures and our concepts of the intentional objects, whether chairs, elephants, computers, quarks or galaxies, which comprise our world reflect our complexity. Although my hypothesis is that all of these worlds are emergent descriptions of the same physical stuff and the project of physics is to describe this stuff, I cannot demonstrate this. Neither can I demonstrate the reality of the objects of our world.

The rich variety and complexity of human agency and our artefacts of regulation is a long way from the austere concepts of regulation, representation and goal. In Chapter 4 I give some examples of landmarks on the spectrum between the basic regulation of a simple control mechanism or prokaryotic cell, and the complex integrated agency manifest in our diverse responses to events in our world.

1.11 Teleology and Goals

The thesis is frankly teleological in that I explain agency by the attribution of goals and agency as a capacity within the essence of an organism or artefact of regulation. I follow Aristotle in attributing a telos to such entities, and this telos will include the processes of operation, maturation and change prescribed by the genome of an organism and the design of an artefact. I also follow Aristotle in holding that, once the organism is born or the artefact is switched on, it is of the essence of the entity that, ceteris paribus, it will behave thus and so. If we define a local necessity as following the telos and a local accidentality as causation by factors not within the telos, (instantiations of other things not being equal) then Aristotle's terminology applies. This necessity is law like but these are not the physical laws of our universe, such laws have evolved or for an artefact, have been designed, and they could have been otherwise. Accidents of evolution or the design of artefacts have occurred within the constraints of physical laws. Whether these physical laws could have been otherwise is beyond the scope of this thesis

Teleological explanation can be characterised by the invocation of ends and goals and each of these is at risk of reification. In my analysis I shall define function in explanatory terms following Cummins, (1975). Although I use Millikan's, (1984), notion of Normal[3] function as a natural biological function that explains current functionality because of its reproductive history, I do not follow Millikan in her reification of such functions, Millikan, (1993, 19). The behaviour of organisms can be explained in terms of functions and each such explanation is of some capacity of an organism. It is in these terms that the functional system of autopoiesis explains the survival and autonomy of an organism, the functional system of circulation explains the oxygen metabolism of the mammalian body and the functional system of agency explains the interaction between an organism and its environment. I define the goals of agency as possible worlds picked out by the cognitive processes of agency, within the system of an agent in the world. Survival and reproduction are theoretical terms within a theory of the agency of organisms. In these terms, current realisations of natural agency have evolved through the survival and reproduction of previous generations against the accidents of previous environments. The complex goals described within a theory of natural agency are enabled by the prime goals of the natural agent; which are to survive and reproduce. Each agent organism within each generation can be described as pursuing its goals against epistemically contingent events within the world that includes itself and other agents.

In so far as realisation of perception, representation and action are appropriate to the satisfaction of its goals against the accidents of its environment, an agent whether natural or artefact, will be successful in its environment. In so far as this realisation is appropriate to its goals of survival and reproduction against the accidents of its environment the organism, which is a natural agent, will survive and reproduce. In so far as its realisation is appropriate to the goals prescribed by its designer the agency of the artefact of regulation will be successful. The goals of agency are instrumental within a theory of agency.

[3] I shall follow Millikan in capitalising 'normal' when it is used in this technical sense.

They do not represent realities in the world. In Chapter 6 I use these conclusions to resolve the apparent tension between determinism and personal responsibility.

2

HISTORY & METAPHYSICS OF SYSTEM

2.1 Introduction

In this chapter I will outline the history of the concept of system and describe how the application of this concept has become an analytical methodology with application to most of the physical and social sciences. Application of this concept to my own project of refining and extending the concept of agency requires a clear definition of system as a mode of description and the ontological consequences of this definition are considered.

Within language and aided by science, we each maintain an internal physical model of our world with which we derive meaning from current and remembered perception. One aspect of this modelling is our separation of the world into component structures of phenomena: things that we can recognise, describe and comprehend.

Some of these component structures we accept as other organic creatures, more or less like ourselves. Some are naturally occurring things like hills, rocks, planets, stars or rivers. Others are artefacts of humankind like cars, houses, computers and chairs, or repositories of information like books, computers, and other people whom we may question.

Many of these component structures can be integrated into other structures which we can name and recognise, or can be analysed into further component structures which we can name and recognise. Analysis and synthesis are not unique. We analyse a structure into its component structures with reference to some context of interest and there will generally be several alternative contexts.

A car can be analysed in terms of engine, body, wheels and transmission. Alternatively it can be seen as comprising; steel, rubber and plastic or as an assembly of components from various suppliers. Each analysis is in a different context. Each of these describes the same real thing which is the car. A term in general use for a whole which, in a given context we recognise as an integration of other wholes which relate or are combined together in a comprehensible way is 'system'.

The use of 'system' to denote a whole which is the sum of it's parts has a long tradition in Western language. In ancient Greek, systema (from sys-hystemi, "to [make to] stand together") originally meant something joined together - a connected and composite whole. The first recorded occurrences of the word appear to be in the texts of Aristotle and Plato. In "The Generation of Animals" Aristotle uses 'system' for the animal or organism as an entity comprising body and soul. In the Epinomis, Plato referred to "every diagram and 'system' of number". This distinction between physical and formal system is fundamental to the usage of the term and has endured through the millennia.

An associated notion is that of 'holism'. Holism was defined by Smuts, (1926), as "the tendency in nature to produce wholes from the ordered groupings of units". Another alternative is 'holon'. Koestler, (1967, 48), defined 'holons' as " those nodes on the hierarchic tree which behave partly as wholes or wholly as parts, according to the way you look at them."

2.2 Physical Systems

Modern system thinking has developed through several strands. Mayr has described the development of feedback control as existing since the Greeks but was more recently exemplified in the centrifugal steam governor developed by Watts in 1788 and mathematically explained by Maxwell in the late nineteenth century. The dynamic equations of feedback control now underpin the engineering disciplines of linear, non-linear and stochastic control systems.

During the middle of this century Rosenblueth and Wiener, (1950) and Wiener, (1961) extended the mathematics of feedback control to include the theories of prediction, information and communication required for aiming at a moving target and the new discipline formed by the combination became known as cybernetics. Wiener, (1961), defined cybernetics as the entire field of control and communication theory whether in the machine or in the animal.

Ashby, (1961) extended the notion of cybernetics into a general theory of control in mechanism and system. He defined a system by a listing of variables and relations in a context of interest with a state space that is structurally closed and single valued. For Ashby a system is described by a phase space of variables or operands[1] and the property of closure is a relation between a transformation and a particular set of operands in which the translation does not produce a new element in the set of operands.

Another strand in the development of the usage of system has been the application of the theory of open systems, initially in biology. Kochler, (1938), distinguished between open and closed systems in terms of material or energy transport. Closure here is a different concept from that described by Ashby, (1961, 11). A physical structure may be described as an open physical system and at the same time its formal description may be structurally closed., Ludwig von Bertalanffy, in a series of writings between 1940 and 1970, has extended the theory of physically open systems and advanced the notion of General Systems Theory as a new scientific doctrine concerned with those principles which apply to systems in general and their application to a range of problems

[1] Ashby, (39) drew attention to the ubiquity of system description: "Every material object contains no less than an infinity of variables and therefore of possible systems."

in human affairs[2]. Sommerhoff, (1950), considered the mathematics of control in living systems and formulated the concept of directive correlation[3].

In 1972 Maturana first published his essay on the "Biology of Cognition". This was included in 1973 with his joint essay with Varela on "Autopoiesis, the organisation of the living"[4]. In the first of these essays Maturana stressed the need for recognition, of the role of the observer in the explanation of living systems. He wrote: "Anything said is said by an observer. In his discourse the observer speaks to another observer who could be himself; whatever applies to one applies to the other as well. The observer is a human being, that is, a living system and whatever applies to living systems applies also to him." He defined cognition and organisational self maintenance as central concepts of living systems and wrote. "Living systems are cognitive systems, and living as a process is a process of cognition". Maturana's usage of 'cognition' is not just that of knowledge by perception. For him; "Living systems are cognitive systems, and living as a process is a process of cognition"[5] In the second essay of this volume, Maturana and Varela defined an autopoietic machine as a homeostatic or relation-static system which has its own organisation as the fundamental variable which it maintains. The concept of autopoiesis as a conjunct to agency is vital to this thesis and is described in Appendix A, below. The thesis that autopoietic is identical with living has led to a school of system thinkers concerned to apply the concept to inorganic and extra-organic systems such as synthetic chemical structures and human organisations. The work of this school is summarised and assessed in Mingers, (1995).

Processes of measurement, decision and control can be observed in organisations such as firms engaged in the production of goods and the

[2] See Bertalanffy, (1968), for an overview.
[3] The implications of this work for my application of 'agency' are considered in 3.8, below.
[4] These two essays were published in English together in Maturana and Varela, (1980).
[5] In Chapter 3, below, I modify this usage and place cognition as a central process of agency.

provision of services. Such processes are describable by systems of linear or quasi-linear systems of differential equations similar to those which describe artefacts using feedback control. The System Dynamics movement, initiated by Forrester, (1961), has been based on an attempt to model organisations in these terms and hence derive improved procedures for the management of items such as stock and the timing of management decisions. Current practice in system dynamics has been described by Wolstenholme, (1993), as providing a map or systemic view of the main variables in a problem area and to develop a multitude of analyses of how problems develop and the effects of intervention.

The practice of system dynamics has led in its turn to the development of various methodologies for describing entities such as cultures or organisations in system terms. These are methodologies for the system analyst and can be distinguished from Soft System Methodologies.These are procedures by which members of an organisation in a problem situation, can work together to create system descriptions of their situation and through the process of shared description, learn to recognise the dynamics of the situation or process within which they act and the problems which require solution. These two approaches are explained in Checkland, (1981), and Eden, (1989).

As previously noted, system thinking in philosophy goes back to Aristotle and Plato. Indeed, we may find it astonishing that the term has survived with its usage intact over nearly three millennia. More recently, Broad, (1925), introduced the concept of the emergent characteristics of wholes. While, Aristotle used the term system to refer to the whole which is an organism and saw essence in the component, Broad saw essence as a property of the whole. Smuts, (1926), spoke of 'holism' the consideration of things with organised complexity. Whitehead, (1969), in his "philosophy of organism" came near to system thinking but went beyond the modesty of a descriptive account. Laszlo, (1972b), has applied cybernetic principles to describe the world through its evolution, organisms, their subsystems and their cultures. He defines terms like;

system cybernetics, microphysical cybernetics, bio-cybernetics and socio-cybernetics. Laszlo concludes; "But the systems philosophical framework can remain; it can be held that (i) the understanding of the human being and the world about him is possible in reference to a hierarchy of dynamic systems, defined in terms of their invariances of state and function, and that (ii) the concept of such a systems hierarchy is the optimum framework for the interpretation and integration of analyses of empirical matters of fact". Unsurprisingly, his definition of system is in cybernetic terms: "A natural system is the joint function of four independent properties (a) a coactive relation of parts resulting in ordered wholeness (b) maintenance of a steady state against environmental disturbances (c) Maintenance of organisation against environmental disturbances (d) structural adaptation with respect to subsystems as components and suprasystems of which it is a component".

Laszlo produced his "Introduction to Systems Philosophy in the same year as Maturana and Varela wrote "The Organisation of the Living" and two years after Maturana published his essay on "The Biology of Cognition". The two essays by Maturana and Varela were not published in translation until eight years later so it is understandable that there is no mutual reference between the two works. However, It is surprising that subsequent literature in systems thinking does not draw attention to the similarities and differences between these two theses. In an otherwise admirable account of the development of systems practice Checkland refers to Laszlo but makes no reference to Maturana or Varela. Jantsch, (1980), writing a systems account of evolution in man, culture and the world, uses the paradigm of self-organisation. He acknowledges a debt to Maturana and Varela. For Jantsch, "Mental concepts, ideas and visions become autopoietic levels in their own right".

Recently; several authors, writing of system practice, have sought for and acknowledged, a philosophical basis for their work. Mingers, (1993), writing of information systems, quotes Habermas on critical theory and a communication theory of social action and the later Wittgenstein. At a recent systems conference[6]: Fuenmayor contrasted the writings of Kant, Heidegger and MacIntyre in interpreting the modern discourse of global system thinking. Jackson quoted Heidegger on the 'bringing forth' or 'revealing' of a technological achievement and compared this with the emergence of a social system. At the same conference, Spaul noted that the 'Critical Systems Thinking' commitment to critical and social awareness hangs naturally with Rorty's post modern view of knowledge and contrasted this with Habermas' ‘interest-constitution’ theory.

Philosophers writing in the field of mind and language make extensive use of the term 'system' but generally without explication of their usage. Dretske, (1981), in his work on "Knowledge and the Flow of Information" asks "What makes some systems capable of occupying belief states but not others?" and claims to "use the concept of belief to distinguish genuine cognitive systems from mere processors of information." He speaks also of information-processing systems. He does explain his usage of 'information' but he does not explain what a system is. Dennett, (1983), in applying his concept of intentionality to cognition in organisms describes a vervet monkey as "an intentional system, a thing whose behaviour is predictable by attributing beliefs and desires (and, of course, rationality) to it". He goes on to describe first, second and third order intentional systems in terms of their interactive complexities. He tells us a lot about intentionality but does not explain why or with what limitations, we may describe a monkey as a system. Putnam, (1988), demonstrates that: "... in one sense, any physical system can be modelled as a computer". It is evident that for Putnam, the system is the actual, physical entity and the computer (or automaton) is the model. Kim wrote, in the context of supervenience, of the “world ... as constituting a system, something that shows *structure,* and whose constituents are connected with one another in significant

[6] The 1993 conference of the UK Systems Society, reported in Stowell, et al (1993).

ways." Kim, (1993, 53) His concentration was on the notion of dependence and he did not pursue the conceptual interconnections between system and supervenience. These are considered in 2.8, below.

2.3 Formal Systems

Philosophers have been more precise in their usage in the description of and the development of, the theory of formal systems. Since Plato this usage of 'system' has remained distinct from physical, functional description and the development of geometries and systems of formal logic can be described as the development of formal systems.

In recent years considerable attention has been devoted to information systems. Such systems have been observed at least since the Pharaohs kept records of their stocks of grain and may indeed predate the formal systems of written language. The development and ubiquitous application of the digital computer has led to an explosion of interest in the theory and application of formal systems for the representation, maintenance and storage of information.

The term 'information' is used in two ways; a disjunctive usage is concerned with the probabilities of alternative states of affairs and a conjunctive usage concerned with the combination of propositions known to obtain a holism of inference. The disjunctive usage finds its application in the technology of communication as developed by Shannon and Weaver, (1949), and discussed in Ashby, (1956), Dretske, (1981), and Bar-Hillel, (1964). The conjunctive usage is nearer to our every day usage in information systems and would refer to information in the Magna Carta or the structuring of information in a data base, resident in a digital computer. A description of this usage is given in Kent (1978). As in other formal systems of rules, an information system is realised in physical marks and processes, and a complete description of an information system will be a physical system description of the organisation of the marks and the processes for their maintenance and interpretation.

2.4 System as a Mode of Description

In science or in philosophy, we are concerned to explain. Before we can explain; we and the audience, which may be ourselves need to recognise what it is that is being explained. Prior to the act of explanation is the act of description. The system mode of description is one in which some particular thing in our world is described, not just in terms of its function and form but also in terms of its counterfactuality. That is, how its function and form determine the manifold events that can occur in this aspect of the world. It is in these terms that I wish to explicate my notion of 'system'.[7]

We must distinguish between the act of description and the diagrams and well formed sentences in the language that represent[8] the information conveyed by the act. It is this abstract entity, which is realised in text, within language, that I will refer to by the term 'description'. By locating the description in the text I am not taking a stance in the post-modern, post-structural dispute regarding the determinacy or relativity of the 'meaning' in text, though this could be discussed in system terms. This dispute has been addressed by many writers including Fish, (1980), and Midgley, (1994). My aim is simply to distinguish the text of the description from what it is, whether actual, fictional or hypothetical, which is being described. The concepts referred to by the terms of the text may differ between author and interpreter. It is this risk of cultural ambiguity that the exponents of techniques of systemic cooperative description such as Cognitive Mapping, Eden (1989), seek to avoid.

In a system description the domain of the entities described will range over possible worlds picked out by the counterfactual conditionals within the text of the description. A system description will be a description of something. The thing described may be actually existent in this world and a subject of current

[7] The author presented this concept of system to the UK Systems Society in 1995.

[8] 'Conventionally represent' in the terminology to be developed.

explanatory interest. The counterfactuals will then derive from a physical theory of this thing. Alternatively, the system description may be a detailed or functional specification for an entity not actually existent or existent but imperfectly described. The thing described may be conceivable but not physically realisable within current technology or physical possibility.

This characterisation of 'system' is different from that of workers in the system movement such as Checkland, (1981). In p 110 of his book Checkland writes "Let us start with the physical systems that apparently make up the universe". Although he writes of physical systems as if they are substances and distinguishes the system from its model, in the glossary he defines a System as "A model of a whole entity ... An observer may choose to relate this model to real world activity." Klir, (1969), has written, about systems; "In experimental science; an abstraction to be used when nature is examined from a viewpoint. In engineering science; an abstraction when relations are prescribed and we wish to implement them." It is at Klir's level of generality and abstraction that I will use the system mode of description. The narrative mode of description may describe an actual event or an actual structure, it may describe a fictional or hypothetical event or structure. The system mode of description describes how the properties of the entities and the relations between them, prescribe the sorts of structures or events that may occur and 'may' is a modal idiom. I shall routinely speak of an entity as 'apt for description in system terms' to distinguish the entity from its actual or putative system description. The convention of speaking of such an entity as a system is common and to totally abstain would lead to tedious translation but essentially my usage of 'system' is as a mode of description of entities apt for description in those terms.

2.5 The Counterfactuality of System Description

The notion of counterfactual dependence in our physical world is enmeshed with the notions of causation, theory and law. Everyday statements like "If you had got home in time we would have got to the party" rely for their interpretation on theories of cause and effect. Lewis, (1973), has given a general analysis of counterfactual conditionals. Lewis, (1986), contends that our modal idioms are quantifiers over possible worlds. For Lewis, possible worlds are real entities and 'actual' is an indexical like 'I' or 'here' or 'now'. If in the idioms of counterfactual hypothesis, I say, "Had you got home in time we would then have got to the party" then, in Lewis' terms, I am describing a relation between possible worlds. I could have said, "Had you got home in time we would then have got to the party but actually you were late and we did not". It is in this spirit that Lewis, 92-3, claims that 'actual' is indexical. "... 'Actual' is analogous to 'present', an indexical term whose reference varies depending on a different feature of context...". He does not claim that other possible worlds actually exist. He does claim that, in any text containing counterfactuals, the same logic with the same criteria for the preservation of truth will apply for each alternative text within the combinations of the counterfactuals. It is in this sense that a system may be described as picking out a set of possible worlds in which set membership is prescribed by the properties of the entities and their relations, described in the system.

Lewis, (1986, 1.3), also defines the notion of 'similarity' between sets of possible worlds. If a system description contains or implies counterfactual statements then any state of the entity described is a possible world. Worlds are 'similar' in that the components of the state space vectors that locate them are similar.[9]

[9] Similarity is a vague notion. Perhaps, close to Wittgenstein's notion of 'family resemblance'.

Stalnaker, (1968), describes the formal properties of the *conditional function* and considers how we should evaluate a conditional statement. His suggestion is:

> First add the antecedent (hypothetically to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally; consider whether or not the consequent is then true.

Stalnaker then moves from belief conditions to truth condition by reference to possible worlds. He proposes that truth conditions for a conditional should be formulated in the terms:

> Consider a possible world in which A is true and which otherwise differs minimally from the actual world. *'If A then B' is true (false) just in case B is true (false) in that possible world.*

Following a semantical system for modal logics due to Kripke, (1963), Stalnaker defines a model structure and a selection function. If M is an ordered triple (K,R,λ) where K is the set of all possible worlds, R is the relation of relative possibility which defines the structure and λ is to be understood as the absurd world to allow for an interpretation of 'if A then B' where A is impossible. If α and ß are members of K then αRß reads; ß is possible with respect to α.

The selection function *f* takes a proposition and a possible world as arguments and a possible world as its value, it selects for each antecedent, 'A' a particular possible world in which A is true. The assertion which the conditional makes, then, is that the consequent is true in the most similar world selected.

For a given conditional: If P then Q (1)

a) (1) is true iff there is a relevant set 'S' of propositions true of the actual world α and
b) If there is a set of accessible possible worlds of which S and P are both true which may or may not include α then
c) Q is true in the more similar members of S.

A system description will comprise propositions about α and a set of open conditionals understood in the manner of (1) above. Any system description that we can produce will be a finite axiomatisation involving conditionals of this form.

For an example of this, consider a description of a university. A student writing home might describe the buildings, the lecturers, and the syllabus and leisure facilities. Each of these is a particular description of actual states of affairs in her actual world. She will assume that, when she wakes next morning or returns next year, this will be the same university, although we and she know that there will have been changes. Students and staff will come and go, buildings will be extended or change their use. Within the practices of language, the university is defined, by a holism of function and capacity and by that set of possible worlds in which this university has these functions and capacities and this name or some translation of it.

I do not wish to enter the debate on the reality of these possible worlds. The actual university at any time is comprised of real people, real books, real buildings and an abstract constitution of customs and rules, realised in the marks of memories and text. It is this constitution that is described by the counterfactual statements within a description of an ongoing university. That this ongoing university is real is intuitively appealing. Whether, because of this, we should assign reality to all of the possible outcomes, is outside the scope of this thesis. Possible worlds, whether or not they are reified, that are picked out by a physical system which is realised in the world are possible futures of the actual world described by that system.

If a structure described within a system includes the function of representation[10] then that system will generally include descriptions of the representation and what is represented and what is represented may or may not be a possible world picked out by the system. For example, planning systems will include combinatorial mechanisms for the production of representations of possible worlds contingent on possible actions.

A system description does not pick out an entire possible world. It might not pick out a possible part of a possible world. The conditional terms of a system are components of a theory and this theory may be false. It may be a functional specification for an artefact that is not technologically feasible or is logically inconsistent. Even if the description is of actual states of affairs at some instant and of conditionals that are held true over an appropriate range of contingent events, the description is finite and ideal. A lot of the actual world will be left out and all of the theory is ceteris paribus.

2.6 Cause

In this thesis I rely on two distinct notions of cause. These are 'agent cause' and 'physical cause'. The causal power of agency stems from the capacity of an agent for action that physically changes the world. This notion is a principle subject for this thesis and will be discussed extensively in subsequent chapters. Since my ontological commitment is to the primacy of a physical world; for

[10] Representation is considered in more detail in 2.11, below.

me, agent cause depends on physical cause and this section is devoted to the elucidation of my usage of this term.

My notion of physical cause is Humean in that; physical causal explanation is based on the constant conjunction of contiguous events constitutively described as instances of physical properties that are covered by physical laws referring to these property types and these terms imply counterfactuals relating instances of these properties. I distinguish between events that can be explained in terms of physical cause, spontaneous events and accidents. Although we are physical entities within a real physical world, any state of affairs can be alternatively described and alternative descriptions require alternative explanations.

Physical system descriptions are of physical entities and real physical entities are composed of real physical components. The capacities of such entities are enabled by the structure and capacities of components and this enabling is explained within a functional characterisation of components and these functions are realised by physical events that are explicable in physical causal terms.

Russell, (1919), suggested that the role of causation in explanation has grown smaller with the advance of science and is being replaced with notions of functional relationship: probabilistic correlation relations between events or states of affairs[11]. However, my concern in what follows is with physical explanation by physical cause. One aspect of current usage is to describe the relation between successive contiguous physical phenomena that are

[11] Halliday, (1994), writing on "The construction of knowledge and value in the grammar of scientific discourse" suggests that causal forms of verbal representation have evolved as part of modern science and have penetrated into almost every domain of our semiotic practice. Halliday's treatment of cause is metaphysically neutral: new semiotic identities are created by advances in science and the nominal expression in the grammar construes iconically an objectified entity in the real world. New verbal representations that link these entities are themselves also iconic.

constrained by physical laws that represent the observed regularities of our universe as causal and this is the usage that I will follow. There is one notion of cause as relating to a causal force, as if the way things turn out is the way Nature intends them to go. This notion I wish to eschew. The universe was going on for a long time before there was agency. During that time things happened[12] and if our science has got it at all right, all of these events were constrained by the physical laws that we observe today. It is this counter notion that I am canvassing when I contend that physical causal relations explain events that are constrained by physical laws.

In physical causal explanation the explanandum phenomenon P is described as resulting from the conjunction of:

(i) One or more specified incident phenomena C_i.

(ii) An environment of unspecified incident phenomena E.

(iii) Physical laws; L_i that apply to the incident C_i and constrain the outcome to P.

(iv) That, during the phenomenon, E remains stable within a range such that L_i apply. This is assumption E_{cp}.

Bringing these together, the axiom of physical causal explanation is that:

$$(C_i \ \& \ L_i \ \& \ E_{cp}) \rightarrow P$$

[12] Their classification into things will be in the terms of our agency but what happened then is not changed by our classification now.

where the arrow denotes the causing of P according to this explanation.

The cited laws L_i quantify over instances of C_i within environments that meet the ceteris paribus criteria. In a token causal explanation any of the three terms may be cited as causally explanatory. If in an experiment to demonstrate Ohm's law a student asks 'Why did the current vary in this way?' she may be answered; 'because of variation in the resistance or the potential difference' or 'because it is constrained by Ohm's law' or in comparison with another students error, 'because he did not keep the temperature constant'. Another description of the same state of affairs may include the propagation of an electromagnetic wave along a wire and different laws would be invoked in a different explanation.

The notion of physical cause depends on the notion of succession. Our communal model of our world is of physical substance in a physical space within which phenomena produce phenomena in an ongoing temporal matrix. I will not attempt to justify this model or the notion of physical cause described above. In so far as we understand the phenomena of our physical world it is in terms of our ability to formulate a system of physical laws that are explanatory, coherent and economical. In any description of a stream of phenomena the temporal matrix may be continuous or discrete. If it is continuous then we speak of a causal process. If it is discrete then we speak of a causal sequence of events. However, any closed volume of space-time may be alternatively described and different descriptions in different domains of description will define different events, with different phenomena explained by different laws. Description, explanation, domain, laws and ceteris paribus clauses are interdefined.

In a universe without agency, phenomena are constrained only by physical laws. Although in any description each process or stream of events can be causally explained,[13] any state of affairs taken as cause could have been otherwise. Although; apart from spontaneous nuclear events and ill conditioned equations, Laplace's demon could have predicted all phenomena; for the purpose of causal explanation, all phenomena are essentially contingent. It may be that there are ratchets[14] in our universe and by such ratchets, our universe, our world and life within it have evolved, but they could have been otherwise.

A physical event may be explained by a description of a temporal chain of physical events that are linked by physical laws. Each such explanation will be in a domain of description using terms and concepts appropriate to this level of physical explanation. These terms and concepts may or may not reduce to some more fundamental level. If they do, then bridging statements will be available by which the explanation can be translated to another domain of description, though for some pragmatic reasons, we have chosen this level. It may be that, due to complexity or the frontiers of science, such bridging statements are not available and the laws invoked in this explanation are as fundamental as we can get. Whatever the reason for our choice of domain, we use the notion of physical cause to invoke the law or laws, appropriate to this level.

The description that I have postulated is of a universe without agency. If we assume that the universe began with a 'big bang' then I assume that that event and all subsequent contiguous events were constrained by physical laws and the succession of any two successive contiguous or ancestrally contiguous, events is apt for description as causally linked if they satisfy Kim's, (1970), condition for direct contiguous causation:

[13] If the event is explained as occurring with some probability characteristic of an unstable element then this is a prime cause within a physical explanation.

[14] By a 'ratchet' I mean two successive states of affairs physically causally linked and such that the causal propensity of the latter is, from some point of view, different from the former. Thus, ratchets lead to development, although their occurrence is accidental. An example is the suggested role of autocatalysis in the creation of life. Stewart and Cohen, (1997, 21). Ashby,

> e is a contiguous cause of e' if and only if $e \neq e'$ and e bears to e' the ancestral of the relation of direct contiguous causation - that is to say, $(S)(e' \varepsilon\ S\ \&\ (f)(g)(f\ \varepsilon\ S\ \&\ g$ is a direct contiguous cause of $f \supset g\ \varepsilon\ S) \supset e\ \varepsilon\ S)$.

Not all physical events are contiguously caused[15]. An atom of an unstable element will emit particles and change in its atomic structure at intervals that are random and fall within a Poisson distribution. As far as we know, none of these events is caused by another event within the space time continuum of our domain of description. Although the average rate of emission is law like and the creation of this atom may be explained as caused by a prior atomic event, no causal event can, so far, be found for any specific event of this kind. My criteria for physical causal explanation are conceptual coherence and consistency and I am prepared to take physical laws as describing reality if they preserve truth within their domain of description.[16]

The notion of a causal chain is as vague as the notion of contiguity. Many causal chains may alternatively explain any given event. For any specific causal chain within an explanation, each law invoked will contain ceteris paribus clauses that construe some properties as environment rather than causally constitutive. In another explanation within the same domain of events,

(1956, 196), calls a property autcatalytic; "whose occurrence at one point increases the probability that it will occur again at another point."

[15] The notion of contiguity is necessarily vague. Some laws involving the weak forces of gravitation and electromagnetism explain phenomena across cosmological distances. Other laws such as those involving heat flow and chemical reactions explain only locally. The strong forces that bind the atom are extremely local but their rupture causally explains the release of particles that causally explain phenomena over cosmological distances.

[16] Schrödinger, (1944, 10), points out that physical laws rest on atomic statistics and are only approximate. However, we humans take them as true for the medium sized dry goods that we classify by our agency and recognise their vagueness at the atomic level

environmental properties may be taken as causally constitutive. Each such explanation may be true, the choice depends on the context to the explanation.

In a pair of successive events chosen arbitrarily for causal explanation there may not exist a chain of contiguously causal relata such that the successor event can be explained as caused by the predecessor. The successor event is then causally independent of the predecessor. If for some given event no causal predecessor can be found then the incidence of such an event is effectively random; with respect to any given model of events. It is an *accident.*

Alternatively, a causal ancestor of an event may be but one of so many influences that the notion of direct cause is not explanatory. If the number of causal chains is small then we may speak of contributory cause. If the number of causal chains is so large that aggregate contributory cause is not explanatory but a few causally contiguous events can be identified then statistical analysis over a class of such events may show correlation. In applying such correlation to a theory of such situations we would speak of 'contribution to the reduction of variance'. If the number of causal chains is so great that no causal theory can be found that explains the event then such events are also effectively random. Any token incidence of such an event is apt for explanation as an *accident*[17].

Paul Humphreys is a realist about chance. He describes such situations as instances of 'probabilistic contributing cause'. " ... the characteristic feature of a probabilistic contributing cause is that it raises the chance of the effect. i.e., *it produces an increase in the value of the chance of the effect.* So, *assuming the existence of physical chances*, the *direct* effect of a contributing cause is an

[17] I am an instrumentalist about chance. Gallie, (1957), writes: "Under the influence of rationalising philosophers we have all been tempted to ignore the inescapable reality of what is contingent or irreducibly uncertain in human affairs." We may find aspects of the phenomena that are explained by physical laws, operating within their range, and these are spoken of as 'contributing to the reduction of the variance'. This does not reduce the inherent contingency of the phenomena though it contributes to their explanation.

increase in the chance of some property." [18], Humphreys, (1989, 104). Many of the phenomena recognised in our world are so complex in their aetiology that we recognise the practical impossibility of fully explaining any token event in terms of physical causation. For any class of such phenomena, such as respiratory disease or rainfall, we may establish by statistical analysis positive and negative correlations with other phenomena and further analyse the data to ensure that the variables in our model are causally independent. We also seek for evidence of actual physical processes that lead, by the accidents of actual events in the world, to the phenomena of interest. In these analyses we estimate probabilities and these are measures derived from actual incidence of the phenomena. They are not measures of actual *physical chances*. Within our current understanding of quantum phenomena we may estimate the chance of a physical nuclear event but this is distinct from the relative probabilities that we use instrumentally to allow for the observed incidence of aggregate contributory physical cause.

If two coincident events have no common causal ancestor then they may be described as causally independent. An event may be explained as caused by the conjunction of two coincident events in that their conjunction gives necessary and sufficient conditions for the explanandum event. If these two events are causally independent then the explanandum event cannot be explained independently by either of the incident events. I shall follow a common usage in defining the relation between such co-incident causally constitutive properties as 'orthogonal'.[19]

For a simple example of physical causal explanation: A boulder rolls down a hillside. Examination of the terrain may convince us that it was pushed and this would require explanation. Alternatively, without agent cause, supporting soil

[18] Authors italics.

[19] An example of this usage is Nagel, (1979, 207).

may have naturally eroded, eventually permitting the boulder to roll and we would speak of physical cause. Once rolling, physical laws of gravity and inertia will constrain the path of the boulder. It may come to rest in a stream and dam the stream and without some safe guard, this may lead to a flood which may destroy a village. We may use laws of soil mechanics and hydrodynamics to explain the relations between events. Our explanation will go back as far as we think relevant, it may or may not include the geological accidents that have led to the structure of this hillside.

This account of cause takes no account of mind or agency. Our explanation of the loss of the village may include the lack of a safeguard and this may be cited as agent cause if its lack was intentional. The boulder may have been pushed and this may be cited as agent cause. Each action or inaction, of agency, cited as agent cause, is a responsible act of agency and each such act is a determinate response by that agent, as constituted at that time to events contingent for that agent. As will be argued throughout this thesis, agent cause is parasitic on physical cause since agency is realised in physical phenomena within physical substance. Current realisations of the functions of agency have evolved through the contingent events of evolution. One of the consequences of evolution has been that it promotes more complex systems and this may be because they have greater adaptive capacity, Bechtel and Abrahamson, (1986). The agency of such systems may have complex causal powers but their existence is causally grounded in the contingent events of environment and evolution, constrained only by natural law.

I assume notions of cause and accident, the accidents of cosmic evolution, the accident of life and the accidents of organic evolution. Through physical causation, these accidents have led to the simple and complex agency of organisms on our planet and to the complex social agency of man in linguistic communities. Each act of agency changes the world and this change is physically caused by that act. Some of these changes are within the intention of the act of agency and some are accidental in that they are physically caused but not intended. I will define agent cause as the causal relation between those changes that are within the intention of an agent and that act of that agent. As will be seen, agent effect is the resultant of causally independent events although, from the point of view of explanation by agency, each of these events is determined by prior events in the world.

2.7 Function

Entities may be described as having 'capacities', the things that they can do. Cummins, (1975), considers the role of function ascribing statements in explanation and concludes: "To ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an analysis of some containing system. When a capacity of a containing system is appropriately explained by analysing it into a number of other capacities whose programmed[20] exercise yields a manifestation of analysed capacity, the analysing capacities emerge as functions." Cummins' usage of system is near to mine. Where we differ is that a system description of an entity in the world will include statements and conditionals regarding all items in the world, including the entity, that are relevant to some explanation. If just the entity has some capacity of interest then that entity would be a 'containing system' in Cummins' terms and would be apt for description as a 'system' in my terms.

Earlier in his essay, Cummins defines function recursively.

> "x functions as a ϕ in s (or: the function of x in s is to ϕ) relative to an analytical account A of s's capacity to ψ just in case x is capable

[20] It is assumed that, in this usage, Cummins would roughly equate 'programmed exercise' with Millikan's 'Normal conditions' and does not imply actual programming by an agent.

> of ϕ-ing in *s* and A appropriately and adequately accounts for *s*'s capacity to ψ by, in part, appealing to the capacity of *x* to ϕ in *s*."

To say that the capacity for one function is accounted for by the capacity for another is equivalent to the statement that any function is both of something and for something. All of the systems of interest in this thesis either are or contain, a containing system in the sense that Cummins uses. The system description of the living organism in the world will contain an entity that is apt for description as its own containing system since its function is to maintain itself although it will contain components and component structures that have function for it. Any entity with function within a containing system has that function *of that entity* and *for the entity* described as the containing system. Any function is a two-place relation. In the terms of Cummins' explication:

> If x functions as a ϕ in s then ϕ(x,s) is interpreted as ϕ is of x for s.[21]

That a system is a functional containing system does not imply that the system is topologically contained. The mammalian circulatory system is distributed through the body. The system has function for the organism within the containing system of the organism but the heart and blood vessels have function for circulation within the circulatory containing system. Living organisms can have function for living organisms. Hearts and livers are comprised of living cells. Each cell is comprised of components that have function for it and each cell has the function for itself of maintaining itself. Just as each cell has function within a heart so the circulatory system, which includes the heart and other component organs, has function for the organism, which we describe within the containing system. Although a living organism can be described as its own containing system, as in its autopoiesis, it may also occur as a an entity

[21] The function of autopoiesis is the interpretation of ϕ(x,x).

with function in a wider containing system. Examples are; an organism within an ecology or a wolf in a pack. A human organisation, such as a university can be described as a containing system. Each personal role described within the containing system of the university can be described as having a function for the university and may also have function within the containing system of a department which has function for the university. The university has function within the containing system of a community although the community, like the living organism, may be defined in terms of its self maintenance.

Non living entities may have natural function: as dung has function for a beetle and a dead fly has function for a spider within the containing system of an ecology. Alternatively, they may be artefacts, things intentionally made or modified in order to provide a function for the intentional agent who made them. An artefact may be described within a containing system and its components may have function for the artefact, described as a containing system. In a wider containing system, these components can be described as having function for the intentional agent who uses them.

2.8 System and Supervenience

An entity described as a functional system may be attributed properties relating to its capacities. For example, a particular realisation of a university may, by certain criteria, be deemed good at physical research and *bad* at teaching in the humanities. In modal system terms; in the actual world described by the statements of the system and the antecedents of its conditionals, application of these criteria will lead to these conclusions, e.g. 'This university, as it is now is good at physical research'. The question then arises as to whether the dependence relations between these properties of these capacities meet the criteria for supervenience; if so on what are they supervenient and what is the character of the supervenience? Another issue that should be addressed is; whether within the possible worlds picked out by the system description, these properties meet the criteria for supervenience and if so what can we say about the supervenience of these properties?

The notion of supervenience rests on an assumption of an asymmetric dependency between entities and properties of a space-time region that are defined in one domain of description and other entities and properties that are defined in another domain of description. Supervenience is a modal notion. “Supervenience means that there could be no difference of one sort without differences of the other sort”, Lewis, (1986, 15).

There have been many definitions and classifications of supervenience. Kim, (1993), and Horgan, (1993), give an extensive account of the field. For simplicity in this brief analysis, I will focus on two definitions listed by Horgan, (ibid, 567).

> Weak supervenience pertains only to things that occupy the same possible world; it says that within any world, all things that are B-indiscernible are also A-indiscernible.
>
> Strong supervenience pertains across possible worlds; it says that for any worlds w and w’ and any things x and y (in w and w’ respectively), if x in w is B-indiscernible from y in w’ then x in w is A-indiscernible from y in w’.

In a system description that includes features of an actual physical world; the conditionals within the description pick out possible worlds, accessible from that actual world in virtue of the fact that the system is a true description of the actual world. Causal relations cited between these features constitute a theory and this theory is true if the system is a true description of the actual world.

System theories may be at the level of mechanisms[22] or they may be at the level of system capacities and functional components that enable those capacities. For example, one capacity of a university is education and this is enabled by functions such as teaching, the maintenance of reference material, housing and admission administration. Another capacity could be research and this would be enabled by functions such as; selection of research staff, motivation, the maintenance of reference material and administration of research facilities. Any functional theory may in general be realisable in alternative physical form and such alternative structures would be functionally isomorphic but any actual realisation will have specific properties, attributed against specific criteria, solely in virtue of the physical causal powers of that realisation.

If we consider an actual university in an actual world then it is the properties of this university in this actual world that are assessed by these criteria. Although the system description of this university picks out possible future worlds through its conditionals, the quality attributed lies in the present and may not be maintained in all of these worlds. Considering the concept of discernibility against the concept of a university; within any actual world of universities and quality criteria, those that are indiscernible at the physical level will be indiscernible at the quality level and the criterion for weak supervenience is met.

Strong supervenience requires that indiscernibility be established across possible worlds. For a physical system S any actual occurrence of S is in an actual world and in this actual world the existence of S picks out possible future

[22] Defining 'mechanism' as a nexus of physical cause, seen as a holism in an explanatory context, which may or may not be seen as a component with explanatory function. Malcolm, (1968), questioned whether a neurophysiological explanation of human behaviour is compatible with a purposive explanation. For Malcolm a 'mechanism' is a causal nexus but he does not consider the aetiology of purpose or that 'purpose' may be descriptive of mechanisms that react to events by action that is for their purpose.

worlds. In each of these worlds there may be defined a property A_x. In two such worlds, w_i and w_j, A_x may take the values; A_{ik} and A_{jl} and these will each be associated with physical properties, B_{im} and B_{jn}. If two worlds w_i and w_j are B indiscernible then $B_{im} = B_{jn}$ and strong supervenience is the claim that this entails $A_{ik} = A_{jl}$.

If we are describing a university from the point of view of a quality assessor then the quality criteria are taken as given and *quality* is absolute. Possible worlds picked out by his system description of a university will manifest different values of these qualities and these will be strongly supervenient on the physical properties described if indiscernibility of quality is associated with indiscernibility of physical detail across this set of possible worlds. This system description will contain only those statements and conditionals that are relevant to the capacities and functions of this university. Such a system defines a particular spatio-temporal region. Horgan defines Regional Physical Supervenience, (ibid, 571).

> "There are no two P-regions that are exactly alike in all qualitative intrinsic physical features but different in some other qualitative intrinsic feature."

This notion of regional physical supervenience characterises the relation between the higher level qualities and their physical instantiation but the physical basis of these qualities remains unexplained since the criteria by which these qualities are assessed are themselves candidates for explanation in a wider physical system. To pursue our example of the university: If the system description includes statements and conditionals concerning assessors and their criteria then qualitative intrinsic features such as *a state of attribution of being good at research* would be regionally physically supervenient on the physical embodiment if there are no two P- regions exactly alike in all intrinsic physical features but differing in other qualitative intrinsic features. The essential characterisation of physical regional supervenience is unchanged between these two examples but in the second example, the supervenience of the first is explained. We now speak not of *quality* but of a *state of attribution of quality.*

From the stand point of explanation, the first system description, from the point of view of the assessor, is incomplete. Similar arguments apply to Hare's, (Hare, 1952), examples. If we say "St. Francis was a good man." this statement is made within a community whose members recognise certain criteria of goodness. These criteria may be absolute for that community but the physical basis for that attribution will be within the physical realisation of the mores of that community. If the attribution is of the rightness of an action and 'rightness' has a moral connotation then a similar argument will apply. This is not an argument against moral realism, although this 'attributionism' is consistent with an agnosticism about the reality of moral properties. States of attribution of goodness of a person or goodness of research at a university are real states. If they are physically explicable then they will be strongly supervenient on the physical detail of the system within which such states of attribution occur. This system will include: those agents whose actions comprise the attribution, the objects about which the attribution is made and the mores of practices within which such attributions are made and have conventional meaning for the community. Each of these is realised in physical mechanism.

> I conclude that system properties are strongly supervenient on the physical properties of their realisation only if the system includes those mechanisms by which such properties are attributed.

This distinction between the physical supervenience of attributed states and the physical supervenience of states of attribution is relevant to the to the physical supervenience of mental properties and this will be considered in 4.12, below.

If the attribution is of the rightness of an action for an agent then it will be deemed as right in that it is caused by Normal representations within Normal cognitive processes. The action is deemed as right within a theory of agency. A theory of agency will be described in the terms of a domain of description that classifies the world for the theorist. Since the theorist and the agent share a common world, domains of description for which this theory is true will be linked by bridging laws and the truth of the theory will be independent of the agency of the theorist, although without the theorist it could not be said. This is an argument for realism regarding representational states of agency, such states are realised in physical states that track physical states of the world.

2.9 Alternative Realisation

Within the hierarchical, functional structure of a containing system, components can be defined with function at distinct levels within the hierarchy. In these terms, the heart can be described as having the function of pumping blood for the organism and component structures of the heart, such as muscles and valves, can be described as having function for the heart. If, at any particular level, a component can be fully described in system terms as having a specific function for an entity or another component then any alternative physical realisation of this function may replace that component in this system without changing the system description down to that level in the hierarchy. This is close to Putnam's, (1973b)[23], notion of functional isomorphism. For our purposes:

> Any two components are functionally isomorphic if either can replace the other within a containing system

[23] What Putnam wrote was "Two systems are functionally isomorphic if *there is a correspondence between the states of one and the states of the other that preserves functional relations*" However, Putnam then identified functional states with computationally characterised states and this, he now admits, Putnam, (1992), was wrong.

> without change in the capacities enabled by the function
> of the replaced component.

We have said that the function of the heart is to pump blood. If this is the only function of the heart then any organism with an artificial heart that pumps blood will be functionally isomorphic with one with a natural heart. This is not to say that functional states are identified by their causal role in the world. It is, rather, to say that functional capacities within an entity described as a containing system are identified by their causal role in that containing system. The notions of function and cause are interdefined within a system. If an artefact has a function for an agent then this function will be described in terms of its causal role for that agent in a system description that includes agent and artefact.

Functions are abstract entities described in the text of a functional explanation. Following Cummins, we may explain the actual capacities of an actual physical entity in functional terms and the functional components will be real physical entities with capacities that may be explained in functional terms. However, for any functional component, the capacity that enables the capacity of the containing system may conceivably be realised in an alternative physical form. Indeed, any entity with some physical capacity that could have function in some containing system could, conceivably, have that capacity in virtue of some alternative physical structure. A physical system description could be couched almost entirely in functional terms. For example, some physical capacities say the storage and retrieval of data may be specified. Functional components to enable these capacities may be specified within a functional hierarchy without reference to specific physical realisation which may be in marks on paper, storage in cabinets and agents reading and writing or, the equipment and procedures of a computer based information system.

2.10 Millikan Functions

Millikan, (1984), and (1993), insists that 'proper functions' are those that are necessary for some capacity because of the historical context in which that capacity has been enabled by that function. That is:

> [for] an item A to have a function F as a 'proper function' it is necessary (and close to sufficient) that either; (1) A originated as a "reproduction" ... of some prior item or items that, *due* in part to possession of the properties reproduced have actually performed F in the past and A exists because (causally historically because) of this or these performances. (2) A originated as the product of some prior device that given its circumstances had performance of F as a proper function and that, under those circumstances, normally causes F to be performed by *means* of producing an item like A. (1993, p 13)

In the explanation of natural functions, and this is her concern, I think Millikan is right. However, this leaves us without a term for the functions of alternative realisations. Millikan claims that an accidental double of a mammal has no proper functions because its history is not right, its heart does not have pumping blood as a proper function. I am content to reserve proper function for items with the right history. The gyroscope in a guided missile has the proper function of orientation because it has the right design history and my heart has the proper function of pumping blood because I have the right history. But if I realise the orienting function of the gyroscope with some alternative technology or if my failing heart is replaced with some alternative pump then these devices have their function in their context, though these are not *proper* functions in Millikan's sense.

Millikan also insists on the use of 'Normal' for those conditions under which a proper function has historically fulfilled its function. Again, this is right for those proper functions whose functionality is vested in their reproductive history but it is also right for those functions whose functionality is vested in their realisation of a function within a containing system. My replacement heart will have a 'Normal' context of operation. It may require batteries and fail to operate near strong magnetic fields. Any explanation of my circulation must now include reference to revised Normal conditions of operation for my heart.[24]

2.11 The Function of Representation

In every day usage, one item is said to represent another when, in some context or by some convention, that item stands in for the other. Linguistic and pictorial representations are the most ubiquitous in our everyday usage although the range of representational structures that we use is very large. An extensive literature exists on the topic of representation; ranging from Peirce's work on the theory of signs, to the semantic realism of Fodor, (1985), the psychological usage described by Perner, (1991), and the natural or biological approach exemplified by Dretske, (1988) and Millikan, (1984 and 1993).

My concern in this and subsequent chapters, will be to distinguish between the representational function of states intrinsic to an entity and the conventional representations of symbol, icon and sign. Most writers speak of representational systems. It is difficult to avoid this usage of system as an assembly of components seen as a whole. But in the terminology of this thesis, *representation* (the verb) is a function of a state that enables some capacity of an entity within a containing system that describes and explains the role of this entity in the world. The system may be physical in that it is realised in actual

[24] This issue is controversial. Walsh and Ariew, (1996), distinguish the principal positions. My usage of function is within my usage of system. For me, a function is an abstract entity: a component within an explanatory description. Real physical structures that are apt for explanation in functional terms may have evolved or have been designed.

physical mechanisms, it may be ideal in that it is not actually realised or it may describe actual arrangements in functional terms. In the latter case, it may be that the representational and other functions could have been realised otherwise.

Examples of representational function include:

(i) the function of terms and sentences within the system describing human language in the world;

(ii) the use of natural and artefactual signs such as weather indicators and meter readings within a system describing humans predicting or measuring aspects of their world;

(iii) the function of states within an organism or artefact of regulation, that track those states of the world that are relevant for the regulatory capacity of that entity within a system describing the entity in the world.

In each of these examples we can pick out a capacity that is enabled by the function of representation. The representational function of terms and sentences enable communication through language and perception. Application of natural and artefactual signs enable other representations of the state of the world for the perceiver and, through perception and action, the internal representations of an organism or artefact enable the capacities of agency or regulation[25].

Dretske, (1988, 52), stipulatively defines a representational system, (RS), as:

> "... any system whose function it is to indicate how things stand with respect to some other object, condition or magnitude. If RS's function is to indicate whether O is in condition A or B, for instance, and the way RS performs this function (*when* it performs

[25] In the next chapter I will argue that agency and regulation are functionally equivalent. and that representation is a vital component in the functional enablement of these capacities.

> it) is by occupying one of two possible states a (indicating that O is A) and b (indicating that O is B) then *a* and *b* are the expressive elements of RS and *what they represent* (about O) is *that* it is A (the case of *a*) and *that* it is B (in the case of *b*)."

Dretske divides realisations of representation into three classes that broadly match my three examples above.

Type I systems have no intrinsic powers of representation. Their representational function derives entirely from conventions of use. Dretske follows Peirce in calling the representational elements 'symbols'. In his terms they are "doubly conventional; *we* give them a job to do and then we do it for them".

Type II systems are those of inherent natural or artificial meaning. Their representational function for us, the observers, depends both on an actual correlation between one state of affairs and another, which may be natural, as in tree rings or artificial as in the case of a voltmeter and a convention of interpretation by us. Correlation alone does not constitute representation: I am aware of many correlations that I do not interpret as representations. The position of the steering wheel on my car correlates with the angle of the wheels but I do not normally perceive my direction of travel by looking at the steering wheel. Since such systems require only one act of interpretation, they are singly conventional

Type III systems are those in which the representational function is independent of any conventional procedures extrinsic to the use of the representation for the behaviour of some entity of which the representational structure is a part. The representational function is realised within some entity, a capacity of which is enabled by that function and no conventions or procedures outside that entity play any role in the realisation of representation for that entity. Dretske describes these as natural[26] systems of representation. "Natural systems of representation ... are ones which have *their own* intrinsic indicator functions that derive from the way the indicators are developed and *used by the system of which they are a part."*

Where Dretske distinguishes the extrinsic nature of type I or Type II representation from the intrinsic nature of Type III representations, Dennett, (1982/3), contrasts the *explicit* representation of interpreted structures with *tacit* representations that represent within the processes of an entity. He characterised explicit representation in the terms:

> "Let us say that information is represented explicitly in a *system* if and only if there actually exists in the functionally relevant place in the system a physically structured object, a formula or string or tokening of some members of a *system* (or 'language') of elements for which there is a semantics or interpretation and a provision (a mechanism of some sort) for reading or parsing the formula." [27]

Dennett is not using 'system' in my strict sense of a description. There are, apparently, two 'systems', in my sense of the term, to which reference is made in this passage. The first is a physical system, which includes representations, what is represented, the users of the representation, and interpretative mechanisms by which the users relate the representation to the represented. The second is a system of conventions within which the physical representations are interpreted by those who are party to the conventions. This, latter, system would be realised in the "provision (a mechanism of some sort) for reading or parsing the formula." which would be, at least partially, within the cognitive

[26] Since Dretske is not concerned with artefacts of agency, for him, all such representations are *natural*. Thus in this context *natural* can include the artefactual.

[27] In Dennett, (1987, 216, author's Italics.

processes of those party to the conventions. Within the theory that I shall develop, representation within a physical system of usage of this kind should be seen as 'explicit' since, following the conventions of representation and interpretation are intentional acts of the agents who use the representations within the first system, mentioned above.

An example would be a system of cars, drivers, roads and traffic lights. The function of the system is transportation and the function of the representational states of the traffic lights is to represent the rights of way at road junctions. The semantics is the usual relation between colour and right of way, and the mechanism is the perception of colour by drivers and their agency in following the conventions. Since the system includes the represented, the representation and the users of the representation, a complete explanation of the system will include the semantics of the system and the mechanisms by which the semantics are implemented. In such a system the function of the representation is explicit within the conventions of users who maintain it because the users of the representation are aware of its conventional nature. Other examples of explicit representation are in language and computation. In a flat language, without semantic ascent, the conventions are not themselves explicit but are manifest in the intentional acts and interpretations within the practices of the language. A computation is an intentional act and interpretation of the representations of computation is conventional within the context of the computation.

Sometimes implicit representations will depend on explicit representation. In our traffic light example a photograph of a driver disobeying a signal will implicitly represent the rights of way at the time through the inferences of location and timing of the photograph. If this photograph is used by the police, then the photograph will explicitly represent the crime through the interpretations of another physical system including police and a public prosecutor.

Dennett contrasts explicit and implicit (or derived) representation with 'tacit' representation. Explicit representation represents under an interpretation while tacit representation represents within a function He invokes Gilbert Ryle's, (1949, 58), distinction between 'knowing that' and 'knowing how'. Knowing that is essentially propositional, explicitly representational in that use requires semantic interpretation. The representations of knowing how are essentially procedural, tacit in that use is independent of semantic interpretation. In Dennett's words; "For the whole point of tacit representation is that it is tacit! States of such a system get their semantic properties directly and only from their functionally defined roles." (ibid. 223)

The extrinsic-intrinsic distinction is one of physical domain. The representational character of colour in traffic lights is vested in the conventions of the community that see and activate the signs. The lights are perceived by members of the community but are not a part of any member of the community. Similarly, propositions are produced, parsed and interpreted by members of the community but are not a part of any member of the community. Intrinsic representations, on the other hand, are used by the entity of which they are a part. They are intrinsic to that entity.

The explicit-tacit distinction is one of logical domain. The conventions that govern the interpretation of traffic light or sentences are available for articulation and inspection by those using them. The physical processes through which an entity represents its environment for it are a part of what it is to be that entity and are not available to the entity. As I shall argue, interpretation is an act of agency but representation within agency is a part of what it is to be an agent.

These distinctions are vital. All of our descriptions and explanations are couched in the explicitly meaningful symbols of language and it is in these representations that we interpret the intrinsic and tacit but meaningful, representations within all agency, including our own. The intrinsic representations of autonomous agency will be considered in 3.5. In 4.12, I will argue that propositional attitudes are explicit representations of attributed relations between actual or putative representations within the integrated agency of humans and other higher vertebrates and the actual or putative states of the world that they represent.

Computation depends on extrinsic representation. In Chapter 5, I will argue that, although some of the cognitive processes of agency may be procedurally isomorphic to some computational processes, cognition is not computation. Computation is an act of agency and the intrinsic representations within agency are not available for interpretation by the agent.

2.12 Malfunction and Misrepresentation

Within a system explanation of a capacity, the function of an entity is explained as the way that Normal functioning of that entity contributes to the exercise of that capacity. An actual realisation of that function in the physical domain will be subject to the accidents of a contingent world. A heart may not pump blood because a clot has lodged in a valve. A term in language may be misunderstood because of noise or because parties to the exchange did not share the same concepts. The damaged heart may pump but badly, a misunderstanding within language may cause minimal embarrassment.

The notion of function proposed by Millikan is normative while that proposed by Cummins is explanatory. These have been seen as opposed by some writers, including Millikan but need not be. If a system description that is set out to explain a capacity of a particular entity or of a natural kind, includes that some function enables this capacity, then if the explanation is of a Normal capacity, it will explain this in terms of Normal function and this may be a 'proper function' in Millikan terms. If hypothesised possible events lead within theories to hypothesised loss of function in some entities that Normally have that function then a system can be described in which counterfactuals concerning the occurrence of such events and their consequences are included. We may be led to formulate such systems because narrative descriptions of such occurrences exist or because within our current theories, such events are possible. Within such system descriptions, possible events antecedent in counterfactuals, have consequences that are apt for interpretation as malfunction. If one Normal consequence is that some state is representational then it is at least possible that the consequent state will misrepresent.

The relation between 'misrepresenting and malfunctioning' is discussed in detail by Karen Neander, (1995, 125), in an essay with this title. She illustrates the analysis with a by-relation diagram of functional causal influence in a frog striking at things flying past.

contributed to gene replication
↓
by
helping to feed the frog
by
↓
helping the frog to catch flies (prey? food?)
↓
by
detecting small, dark, moving things

Neander suggests (p. 129), that "the frog doesn't misrepresent as long as its representation R is caused by something appropriately small dark and moving." Her justification for this is that, (p.130), "it is *by* detecting small dark moving things that the frog detects frog-food and flies." This would seem to imply that, for her, the frog does not misrepresent if any misrepresentation is not the result of malfunction in the frog and this is reasonable. However, since the cognitive processes and representations of the frog contributed to species survival by helping to feed the frog then, under another description[28], small dark moving things are intentional for the frog since they may be represented as food. If, under this description, a small dark moving thing is not food then those intentional states or representations of the frog that represent this as food, misrepresent. In general, the Normal function of representation will be consequent upon the Normal functioning of other explanatory components. However, if a state misrepresents, this may be the consequence of malfunctioning of other components or of ambiguity in the world for the frog. This distinction is relevant to the consideration of misrepresentation within agency, which will be considered in 3.7, below, and to misrepresentation within computation, which will be considered in 5.4, below.

Representation in the frog is of type III and for the frog, issues of truth and falsity do not arise. If we see a thunder cloud and conclude that it will rain but it does not, then the proposition in which we state our conclusion is proven false. However, in type II terms, the signification of the thundercloud for rain was a malfunction, for the forecaster, of representation and our conclusion of rain to come was a misrepresentation. Though, had we said that it looks like rain, we would have spoken truly. Propositions, representations of type I, may be held to be true or false in so far as they relate to states of the world and if they are false then they misrepresent. However, a statement is an act of agency and its representational function is for that agent. If he intentionally lies then the misrepresentation of the lie is functional for him, although its falsity may be dysfunctional for a listener. Alternatively, a state may falsely represent because

[28] This is Anscombe's (1957) notion of intentionality 'under a description' and will be met again in 3.2, below.

of an error of perception or ambiguity in the world for the perceiver and misrepresentation, falsity and malfunction will go together. A true statement may be misheard and the resulting type I misrepresentation will be a malfunction of communication.

2.13 Conclusions

Although there is an extensive and respectable tradition of systems thinking within Western history some of the usage within this tradition is, in my opinion, incompatible with my development of systems methodology for the explanation of, in particular human, agency. For this reason I take system to be a mode of explanation of the behaviour of physical entities that is characterised by the use of counterfactual terms. Systems pick out possible worlds. A system may or may not be truly realised in physical form. If it is, then it will pick out worlds that are actually possible; that is, accessible from the actual world in which it is realised..

The behaviour of such entities is explained by theories that are based on physical laws so that all resultant phenomena may be taken as physically caused. It is an axiom of my thesis that all cause is, at root physical cause: phenomena constrained by physical laws and initiated by events that may be contingent. Agents cause events by their actions that result from the physically realised and functionally explained capacity of agency.

Within such descriptions we explain capacities of physical entities in terms of the capacities of their components. Components within entities are described as having function for the entity in virtue of their enabling capacities of the entity. Functions are abstracta, which may be realised in physical entities or other components with function, which in turn, may be realised in components or physical entities and the hierarchy will rest on a basis of physical causal explanation.

A component entity or structure may have function in terms of what it does. A state of some entity or structure may have function in terms of what it represents. Components may malfunction and states may misrepresent. Malfunctioning and misrepresenting are both described within the abstraction of a system description in which the counterfactual terms within a description of Normal functioning are augmented by counterfactual terms that pick out the abnormal consequences of abnormal incidents.

3

SYSTEM, INTENTIONALITY & AGENCY

3.1 Intentionality

The intentionality of mental states and events derives from their property of being *directed at* or *about* objects and states of affairs in the world. Discussion of intentionality goes back certainly to the medievals and some notion of intentionality could be derived from Aristotle's discussions of the practical syllogism. For Aristotle, man is a rational animal and our concern is to understand and perfect this rationality[1]. Also, for the medievals, intention is the distinction between the natural existence of the insensate or merely brute, and the mental existence which distinguishes man.

For Aquinas, intentionality is that our thoughts are about things and the form of the thing in a thought is a concept. A form of something present in someone's mind as a concept is, for Aquinas, obviously present in a very different way from its presence in an external thing. Aquinas called this mental mode of presence 'intentional being'.

For Brentano, the fundamental property of consciousness is intentionality. All and only mental phenomena exhibit intentionality Every subjective experience is understood as an act of consciousness referred to some object. Brentano was concerned with the ontological status of intentional objects. If objects can be described as intentional then they cannot be intentional in themselves, de re, as

[1] A contention that will be argued for in this thesis is that Aristotle was wrong in this regard. The behaviour of man as agent is as rational, though far more complex than the other higher vertebrates. What distinguishes man is his capacity to rationalise. As a member of a linguistic community, his behaviour as agent may be capable of more complex rationality due to this rationalising.

this would either lead to the vicious regress of intentional perception of intentionality or to a dual aspect theory of things. The alternative is that intentional objects are objects of thought and this was Brentano's choice. If objects of thought are real then we need a duality of stuff. If they are ideal then we have the Platonic duality of form and stuff. If they are nominal then they are within the description and this is close to the position that I will develop.

For Husserl, every act of consciousness is intentionally directed towards some object. He wrote: "Conscious processes are also called intentional; but then the word intentionality signifies nothing else than this universal fundamental property of consciousness to be conscious of something."[2]

For Sartre, intentionality is constitutive of consciousness, its directedness on to outer objects that are prior to consciousness. He wrote: "To say that consciousness is consciousness of something is to say that it must produce itself as a revealed-revelation of a being which is not it and which gives itself as already existing when consciousness reveals it."[3]

Anscombe, (1957), distinguished intentional actions as those that are answers to the question: "why is he doing that?". Of particular relevance to this thesis is her notion of a hierarchy of descriptions of an action. Her conclusion that an act can be seen as intentional under an aspect of a description was extended by Davidson, (1971) in his definition of agency.

Recent approaches to intentionality range from the instrumental account of Dennett, (1987), to the realistic representations of Fodor[4] and the biological functionalism of Millikan, (1984). Lyons, (1995, 1) defined intentionality as

[2] Cartesian meditations, (p 33).

[3] Being and Nothingness, 1943 ed'n., Routledge, 1991 p.xxxviii.

[4] See the introduction to "Meaning in Mind", ed. Loewer & Rey, (1991), and Fodor, (1987).

"covering those characteristics of mental activities on account of which those activities are said both to have a content that contains information about something beyond the content and the activity, and to involve a particular sort of attitude towards that content. Moreover, it is a peculiarity of mental content that it is necessarily 'perspectival'."

Lyons, (op cit.), presents a cogent argument for a bottom up explanation of human intentionality. Since he is concerned to explain human intentionality, his account is concerned with the physical basis for the aboutness of the world for humans. He starts with the aboutness of brain function and distinguishes this from the aboutness of the behaviour of the pre-linguistic child and distinguishes this from the aboutness of the propositions with which we attribute aspects of aboutness to one another and ourselves. As will emerge in this chapter and the next, my conclusions are close to those of Lyons. Where Lyons is concerned to explain human intentionality, defined in his terms, I am concerned to generalise the very idea of intentionality and to demonstrate that the resultant generalised notion is a systems notion, the attribution of which need not be restricted to humans, vertebrates or even organisms.

For each these writers, intentionality is a property of consciousness and consciousness is a phenomenon to be accepted or explained. If we take intentionality as a property of consciousness then we will need to consider what sort of thing is consciousness[5]. If intentionality is a property independent of consciousness, we must enquire of what it is a property and what sort of property it is. I will argue that intentionality is a property of an action or state of an entity and is essential to the agency of that entity. This is not to suggest that consciousness is not intentional but it is to suggest that the combined notions of intentionality and agency are more fundamental.

[5] The explanation of consciousness is not a subject for this thesis. Where, as in chapter 6, I need to consider consciousness I will rely on Dennett, (1991).

Whether it is a property of consciousness, of thought or of sentences, intentionality is aboutness.[6] We may speak of the objects of thought or the objects of consciousness. Thought is about something and we are conscious of something. Sentences are about things, whether actual, historical or fictional. Sentences are sequences of marks within a syntax, through which they are given meaning for us by our practices of use. It is in our practices of use that sentences are about something and so sentences are about something for us. The thoughts we think are our thoughts and we are aware of them in our consciousness. As we recognise that our thoughts are ours we also recognise that other's thoughts are theirs. 'Consciousness' is the term we use for our awareness of our own thoughts, memories and sensations: the awareness, which we can share with others. In this sense, our consciousness is essentially ours, the objects of our thoughts or the objects of our consciousness are objects for us. Our intentionality is about-ness and it is for-ness and it is about things for us. If we wish to characterise intentionality as a system property it is obvious that one entity which can be described within a system as having this property is a human. This does not pick out a system kind, unless we assume that humans are the only kind with intentionality. The question to be asked is: whether the fact of possessing intentionality can be used of itself to pick out an entity that is apt for description as a system kind? The answer is that we do already use a term for just such an entity and this term is ' intentional agent'

3.2 Agency

Agency is to do with action. In philosophy we are concerned to distinguish human agency in terms of action that is by an agent and for an agent. Davidson, (1971, 46), summarised this in the terms:

[6] Definition in: "The Oxford Companion to the Mind", ed. R.L. Gregory, Oxford, 1987.

> a man is the agent of an act if what he does can be described under an aspect that makes it intentional.

This usage of 'agency' must be distinguished from the other common usage of an 'agent' as a substance that acts. In chemistry an agent is the active component of a reaction in distinction from a catalyst that enables a reaction or an inert substance that has no relevant causal power. In human affairs, a man is the agent of another if he acts on that person's behalf. In biology we may say that 'wind is the agent of pollination'. These notions of agency focus on a particular change within a causal chain and define agency as the property of that which has causal power for this change. My usage of physical cause as explanatory within a system does not relate events in this way and this usage of agency is not relevant to my thesis. Throughout this thesis 'agency' is to be understood as 'intentional agency' and 'agent cause' is change caused by an intentional act.

For Davidson, (1963), agents act from reasons and reasons cause actions. Reasons, for him, are mental states such as beliefs and other attitudes. It is that agents have these states that cause an agent to act. Thus Davidson, (1970), suggests that; although physical cause requires physical laws and reasons cause actions, there are no psychophysical laws. Davidson is concerned with human agency, human action and human rationality. If we assume, as he does, that mental events are events of a human agent then we have reason to consider the apparent paradox of how such events can cause action without a physical causal connection. For most of the rest of this chapter I intend to bracket this question. My reason for this is that I will argue that mental talk is not best interpreted as talk about mental states or mental events of or within a human or other, agent. To establish this argument, I must first generalise the definition of agency. I will develop the argument that our rich conception of human agency can be seen as an elaboration of an austere conception of agency which stems from its definition in terms of action and intentionality.

My amendment to Davidson's criterion is slight, I propose that agency, intentionality and action be co-defined in the terms:

> An entity is the agent of an act if what it does can be described under an aspect that makes it intentional for that entity.

By an entity I mean any substance that we may distinguish and describe. The distinguishing and describing are our actions. They are actions of theorists who are agents, also in the world and of the world. An entity has agency under our description if; within this description, its actions have an effect on its world and this effect is by the entity because it is for the entity. Actions of agency are physical events, physical changes of the agent that cause physical changes in the world that includes the entity. States of an entity are intentional if they relate to states of affairs that are for the entity. Thus the concepts of agency, action and intentionality are logically bound. Necessarily, an agent has intentionality if its actions are for it. An entity with intentionality but without agency would be in the world and of the world but would not affect the world in a way which reflects that its world is for it. Although such an entity may exist, unless it is our artefact and we have access to its states, we cannot be aware of its intentionality. A digital computer may have states that are about states of affairs in the world, though they are not of its world, but it does not act. The states of any measuring device are representations, for the user, of states of the world, but the instrument does not act. Such artefacts are not intentional in themselves since their intentionality is vested in our interpretation which is manifest in our behaviour in our world. States of the brain are about states of affairs in the world, within our theory of our agency but the brain does not act.

I will take it as an axiom that there is some way that the world is, which is independent of agency although its description by us must be dependent on our agency. The project of science is to describe the world, using the terms that reflect the way the world is for us and theoretical terms, grounded in these terms. However, every agent in the world, the world that we describe in the terms of our science and which we believe to be independent of agency, is also, in its world, the world that is for it in virtue of its intentionality. Any agent is in the world and it is of the world. Also, it is in its world, that cross section of the world that is, intentionally, for it.

The world of the agent is causally related to the real world. The set of the objects that are for it may overlap the set of objects of the world of a theorist, within some domain of objecthood in the theory of agency of that theorist. Since the theorist is an agent her objects are a cross section of the real world. Her ontology of that world is created by her classifications of the world through her agency. Each agent classifies the real world of actual physical stuff for it. Agents who describes other agencies in their theories will couch those theories in the terms of their own ontology which derives from the classifications of their own agency.[7]

3.3 Functions of Agency

An entity has agency if its actions can be described under an aspect that make them intentional. The actions of an entity are intentional if they change things in the world and this change is for the entity. Agency is a system capacity that is enabled by the intentional functions of perception, representation and action. If we can establish criteria for realisation of these functions then we can use these criteria to derive a minimal functional description of an intentional agent. This will then enable us to pick out examples of entities in this world to which we can attribute agency and intentionality.

Any token realisation of agency will be a physical entity in the world and we must distinguish between the entity that is the agent and the entities that

[7] This ontological dependence is illustrated in the example of 4.3 below.

comprise the world of the agent. As a physical entity it will have causal relations with its world. Changes in the agent will cause changes in the world of the agent and changes in the world will cause changes in the agent. This theory of agency may be summarised in five functional definitions.

> (i) Actions of agency are those agent changes that cause changes in the world that are intentional for this agent
>
> (ii) Perceptions of agency are those agent changes that are caused by the world and are intentional for the agent.
>
> (iii) Representations within agency are those agent states that are changed or maintained by perception so that they are intentional for the agent.
>
> (iv) Cognitive processes are those causal processes by which perception changes representation, representations change representations and representations cause action so as to further the goals of the agent.
>
> (v) Goals of agency are those possible worlds that are picked out by the cognitive processes of agency.

The defining characteristic of agency is intentional action. Action causes change in the world that is for the agent. Action is caused by change in states of the agent and these states are themselves changed by events in the world. Those states of the agent that cause action by the agent that is for the agent are caused by events in the world that are relevant for the agent. These states I will define as representational states of agency. They are representational since they re-present the world for the agent.

The agency of an entity is manifest in a process. This process can be described as; events in the world causing changes in the agent which cause changes to the world that are for the goals of the agent. The component process within which events in the world cause change to the agent I call 'perception' and the component process within which changes in the agent cause change in the world I call 'action'. Those states of the agent that, through perception, are changed by events in the world and that, through action, change the world I call representations. Those processes, generally within the agent, by which perceptions change representations, representations change representations and cause action that is for the goals of the agent I call 'cognitive'.

3.4 Goals of Agency

E. Nagel, (1979), distinguishes three explications of goal directed processes. Firstly, the goals of intentional action in which an organism can be described as goal-directed[8] only if it is legitimate to ascribe intentions, desires and beliefs to the organism. Secondly, the program view of goal directed processes in which the process is constrained by some code to continue to some end. Thirdly, the system property view according to which an entity is so structured that it will adapt itself towards some goal state against variations in its environment.

To take the first of these explications is to seek a functional account of 'goal' that will lead to realisation in actual states of an entity that represent propositional attitudes that are about actual states of the world and are concerned with goals that have actual import for the entity. Later in this thesis we will be concerned to explicate the folk psychological practices of a community of linguistic agents in terms of physical states of the community but this is not relevant here.

[8] McFarland, (1989, p.43), distinguishes goal seeking systems from goal directed systems, where the latter refer to explicit goal representations. In this usage Nagel makes no reference to goal representation but the distinction is relevant and I shall speak of goal seeking processes.

To consider the second of Nagel's explications: A program may have a goal and if the coding of the program is an act of agency then this will be the goal of an agent. Alternatively, the program may have evolved, as the coding of DNA programs the growth of an organism. It would be reasonable to think of such growth as the goal of that program but this growth is not within the goals of the agent organism. Also, it is my contention that the evolution of DNA is not the outcome of an act of agency.

The system-property view of goal-directed processes as described by Nagel is the one that will be argued for in this thesis. Nagel's exposition is based on the work of Sommerhoff, (1950). Sommerhoff identifies goal directed behaviour by its plasticity; that the goal may be reached by different paths and from different initial conditions and persistence; that the continued maintenance of the entity in its goal-directed behaviour is enabled, by changes within the entity, against disturbances within the entity or its environment. In particular, the controlling variables that maintain action towards the goal must be orthogonal. By 'orthogonal' we do not mean that their values will not be correlated as they obviously are. What is meant is that neither is causally dependent on the other or jointly causally determined except through their co-operation in control[9]. Nagel's examples are of homeostasis within the circulatory system and of steam engine speed control through a centrifugal governor. In each example, controlling variables; blood water content and kidney activity in the first and engine speed and valve opening in the second are orthogonal as defined. For kidney homeostasis, the blood water content can be seen as the environment to the kidneys and for the governor, engine speed can be seen as the environment to the governor. The goals of agency of an organism are within the environment of the organism. The Sommerhoff criterion is that the events in the environment against which the organism is pursuing its goals are essentially orthogonal to

[9] This usage of 'orthogonal' was defined in 2.6, above.

the structure, in virtue of which, the organism can be described as pursuing its goals. Actions of the organism will correlate with events in the environment and Sommerhoff coined the phrase 'directive correlation' for this linkage. This causal independence or orthogonality between goal and perception in the analysis of intentional action is critical:

> Apart from actions of an entity that are programmed towards a goal, goal-seeking behaviour is characterised by input events in the world, which are essentially unpredictable for the entity.

It has been argued that: if 'goal' is to be associated with 'final cause' then, since goals are in the future, effect can precede cause, (Taylor, 1966). This is not my usage of goal in this account of agency. In the description of a theorist, an intentional act divides the world. For the theorist, the world of the theorist is classified *for* each agent by the intentionality of that agent. This classification is in the concepts of the theorist and in this sense the classification is *for* the theorist but any token classification is an aspect of the theory of agency of a token agent. It is the acts of that agent that establish its world *for* it and this world is manifest *for* the theorist in the classifications of the theorist. Perception, representation, cognitive process, goals and action define agency in a theory of an theorist, who is also an agent in and of the world that contains the agent, whose acts are the subject of the analysis and explanation of the theorist.

The goals of agency are described within the theory of a theorist. In one such theory of a short-term goal described by a possible state of affairs, the goal may be described as if it were in the future. But this is a possible future within the counterfactual theory held by the theorist and picked out by the system of agency in the world. Fundamental goals of any organism are to survive and to reproduce, since organisms that exist are those, whose ancestors have survived and reproduced. Again, for any token organism, its survival and reproduction are possible futures described within the theory of agency. Falk, (1981), gives a regulatory account of agency that is very similar to the above, although his terminology is very different. He describes regulatory systems as 'natural feedback systems' and generalises the notion of 'subtraction' to accommodate all complexities of goal directed processes and writes "so the negative feedback system, complex as it is, is the minimum system with states that can have a representational function for the system."

It is not of the nature of a goal of agency that it be represented. Although, in the complex integrated agency of the higher vertebrates or a target seeking missile, actual or possible states of affairs may be represented in states of the agent. However, goal states such as survival or the achievement of reproduction may not be represented within the system. Analysis of the system comprising the organism in its environment will reveal that the cognitive processes of agency Normally pick out possible worlds that, within the theory of agency, are apt for description as survival or reproduction. Nourishment is not represented in the representations of the agency of a bacterium that bias the aggregate motion of its cilia or the neural states of a frog striking at a passing speck. In those cases where goals are represented the representation may be explicit, as in a student deciding to work for a degree or co-ordinates in a target seeking missile.

Taylor, (op cit), argues that purposeful behaviour is uniquely characteristic of agents but he claims that it cannot be understood or even described using only the concepts of physical science. In Chapter 7 he argues that there can be no possibility of a causal analysis of intentional action unless, "all events including those that are the acts of men are causally determined by other events". For him the *I* in '*I* move my hand' is transcendental to any causal analysis. Purpose has a fundamental place in his ontology and he sees cause as irrelevant to purposeful explanation, (221).

McFarland, (1989, 54), opposes the notion of goal representation. He contends that: "... the behaviour of individual animals (and people) is guided, not by any goal-representation, but by myopic hill-climbing behaviour". My use of goal is more modest than the usage that McFarland is concerned to oppose. A theory of agency that includes the perception of a potential gradient, whether gravitational or chemical, may include as goal; climbing up that gradient. This is how E. Coli proceeds towards its nourishment and sometimes it will fail. However, the complex agency of an advanced vertebrate may include learned responses and more complex goals may be appropriate within a theory of agency that explains the behaviour of an entity in its environment. Goals are not represented, the structure that realises the cognitive processes of agency is such that actions of agency pick out possible worlds that are described as goal states within a theory of agency.

Natural agents that are observed are those that exist and their existence is consequent on their survival and the survival of their ancestors against the events of their environment and that of their ancestors. These events may be presumed to include accidents from the point of view of the agent, events orthogonal to the current states of agency, since we do not presume any higher agency. Since these natural agents exist we presume that the cognitive processes by which their natural representations lead to intentional action are, at least, not inimical to their survival. If these processes have developed over the life of the agent then this process of development is, at least, not inimical to its survival. If these cognitive processes or their process of development is inherited then we presume that these processes were, at least, not inimical to the survival of its ancestors.

As theorists of the agency of ourselves and other natural agents we deduce from our current observations and fossil evidence that the cognitive processes of natural agents and their ontogenetic development have evolved so as to lead to action that is, at least, not inimical to their survival. Since each agent and its ancestors have survived in an environment of limited resources, we infer that cognitive processes and the processes of their ontogeny have evolved that are, ceteris paribus[10], beneficial with respect to the survival of the agent. Species evolve in environments, and behaviour that is appropriate for the survival and reproduction of the individual against the contingencies of its environment will be selected for. Such behaviour can be interpreted, in a theory of agency of the individual, as *for* the goals of agency. In a given theory the goals attributed may be to avoid danger, to obtain food or to reproduce.

If teleology is the doctrine that structure and behaviour are determined by the purposes they fulfil then this definition of goal is not rooted in teleology. Structure and behaviour have been determined by the filtering of selective processes against events of the past. Any perception, representation or act of agency can, in principle, be described by an theorist, in the terms of some domain, as about aspects of the environment: this is what it is to be intentional. In these terms the theorist can describe the specific goals of an act within the goals of survival or reproduction as the pursuit of prey, avoidance of a predator or partnership for reproduction. The theorist possesses the concepts of these goals. The acts of agency are physically caused by physical states of the agent that have been caused by the environment of the agent. It is within the theorist's theory of agency that we separate this process into perception, representation and cognitive process. Agents act in this way because the selective processes of evolution have filtered out those entities that do not act in this way. The behaviour of agency can be explained in terms of perception, representation, process, goals and action but this process has not been designed to meet these goals.[11] Our use of the notion of 'purpose' is a component practice in our communal theory of mind and the evolution of these practices is considered in 4.11, below. Looking ahead to this analysis, the aetiology of 'purpose' is. Firstly, primitive agency has evolved that classifies the world for it by its

[10] Cashing in the 'ceteris paribus' clause would take us into contentious areas of evolutionary theory that are not relevant to this argument.

[11] This point is also made by Falk, {1981, op cit.).

actions. Secondly, complex integrated agency has evolved that recognises objects and action that are relevant in its world. Thirdly, linguistic communities of such agents have evolved that use signs to refer to these objects and describe behaviour as purposive with respect to the objects that have been classified by that behaviour.

Goals may conflict. The behaviour of an organism or artefact may be apt for explanation in terms of multiple goals and action will be the result of processing of representations that involves a payoff between opposing goals. Normal cognitive processing will have evolved or been designed, to result in action that is generally appropriate for this type of organism or artefact, whether solitary or within a group. However, token actions may not be best for goals such as survival of the token entity if such a goal were to conflict with inherited or learned conventions of a group.

3.5 Representation Within Agency.

Agency, in the stipulative usage of this thesis, depends on representation. Whether human cognition depends on representation is a matter of current debate. Van Gelder, (1995) suggests that cognition may be explained in terms of dynamic coupling He considers the example of a centrifugal steam governor and contends that: since the operation of this mechanism is dynamically determined, its power of control is not apt for explanation in terms of representation. However, van Gelder sees the representational approach to cognition as "... a matter of rule governed manipulation of symbolic representations." He argues that cognition is not just computation. I share this view: chapter 5 of this thesis is devoted to it.[12] However, the type III

[12] More recently Ramsey, (1997), contends that: "... for a significant class of models there is no compelling reason for thinking that connectionism's own terms should include a notion of internal representation." I agree; cognitive science need have no truck with representations. But a theory of agency must involve the agent in the world and describe the world as it is for the agent. If a neural process is called *recognition* within a theory then its output state represents what is *recognised* within that theory.

representations within agency are not symbolic; they are states of an agent that depend on states of the world that are relevant for the agent. They are *intrinsic* to the agent in the sense defined by Dretske and their interpretation by a theorist is *tacit* in the sense defined by Dennett. These concepts were distinguished in 2.11, above.

Recall our five definitions in 3.3, above. Action, perception, cognitive process and representation are functions for the capacity of agency and agency is a capacity of an entity described within the containing system of an agent in the world. Actions are agent changes that are caused by states of the agent and these changes cause change in the world that is intended by the agent. Perceptions are processes of agent change that are caused by changes in states of the world that are intentional for the agent and the processes of perception cause changes in states of the agent. Thus, agency implies that there exist states of the agent that are causally linked to states of the world that are relevant for the agent and that the causal powers of these states of the agent are necessary for agency. Not sufficient, since although agency implies the existence of these states, it has other implications, such as a world apt for classification, processes apt for description in terms of action and perception, and autonomy.

Such states are intentional since they are about states of the world. If the processes by which such states are for the agent in that they lead to acts that are for the agent in the context of perceptions of the world, then these states must, in some sense track the states of the world. We theorists, in our functional theory of the agency of this entity, can call these 'representational states' since, within our theory, these states re-present the world for the agent. I conclude:

> To effect its agency each agent must represent its world, for it, in a state of itself. Such representation is an intentional relation between a state of the agent and a state of affairs in the world.

This purpose for us of this theory of agency is to explain how an entity described in terms of agency functions Normally in the world. The states of the agent postulated in such a theory represent the states of the world for the agent in so far as they mediate, for the agent, between how the world affects the agent in perception and how the agent affects the world in action. That the state represents is independent of any interpretation by the theorist, though what it represents may be described within a theory. The capacity of an organism to achieve its goals of survival or reproduction or ambition in an unpredictable world does not depend on the interpretation of its representations within agency by a theorist. The function of such representations for the theorist is to explain the behaviour of the agent. Their function for the agent is to cause appropriate action in the world.

This theory of agency postulates the functions of perception, representation and action within a causal functional nexus in which perceptions cause representations and representations cause representations to cause action in the world that is for the goals of the agent in the light of its perception of the world. This causal functional nexus will vary from agent to agent but each agent type and each token realisation of a type of agent, can be described in terms of some causal structure of cognitive process that maintains the intentionality of agency for this type or token of agency. This is a thin, austere conception of agency but it is the conceptual core of the agency that is common to living things and artefacts of regulation. Representation is the bridge between perception and action. Actual representational states are about the world, as it actually seems from the point of view of the agent

3.6 The Content of Representations Within Agency

If perception presents the world for the agent and action changes the world for the agent then representation re-presents the world for the agent. The way that our world is for us can be represented in language within the representations of propositions: these are type I representations in Dretske's classification, outlined in 2.11, above. The terms of these propositions represent the content for us of states of our world. It is in sentences made up of these propositions that, within our theory, we describe the states of the world represented within an agent. Any representational state postulated by us in our theory of the cognitive processes of a token agent in the world will be postulated as having content for the agent and that content is what that state represents for that agent. If we can explain the behaviour of an agent by a theory then this theory will be apt for description within sentences of our language. If the theory is true then these sentences will be true. The content of these sentences will be made up of concepts, concepts possessed by the theorist who states this theory.

An agent, qua agency, need not possess concepts and the content of its representations is generally nonconceptual[13]. However, the actions of agency may include linguistic transactions within membership of a linguistic community and people[14] who are agents within that community will generally satisfy the possession conditions for the concepts of their shared language.

The function of representations within agency consists in their causal role in leading, Normally, to appropriate acts of agency within the cognitive processes that define this type or this token of agency. In any true theory of the cognitive processes of a given realisation of agency there will be some instantaneously isomorphic relation between a representational state and the world state or possible world state represented. The dynamic on-going agency, of the realisation of the theory, may not be apt for explanation in terms of static, causal isomorphic relations between the representation and the represented. The content of such representations is generally nonconceptual for the agent, though

[13] The relevance of nonconceptual content is demonstrated by an example in 5.3, below. The nonconceptual nature of the content of perceptual experience and sub-personal computational states is defended by Bermudez, (1995).

[14] The distinction between person and agent is explored in more detail in 6.7 below.

not for the theorist. However, the cognitive processes of articulate rational agency may include explicit representations of beliefs attributed. For example, “I stopped because the light turned red.”

We may observe a cat, a dog and a tree and theorise that; the cat perceives a dog and a tree. In our theory the neural states of the cat will represent the state of the world that we call 'a dog there' in some neural processes that represent that state of the world for the cat; and similarly for the tree. Acts of the cat will be appropriate to a dog and a tree within the limits of this situation (functions are not always fulfilled). Each of its representational states will be caused by and functionally related to[15] some of its perceptions. Each of its actions will be caused by and functionally related to some of its representational states.

An actual entity will have the capacity of agency in the world only if it has states that causally track states of the world so as to enable intervention in the world that furthers the goals of its agency. In simple realisations, such as artefacts and unicellular organisms, the representational content of such states can be interpreted within our theory of the cognitive processes of the agent. Of a simple vertebrate we can say; this state of affairs caused this pressure that excited this neuron that excited that neuron that activated that muscle. The austere agency of bacteria, parts of plants and simple creatures such as the Californian Sea Hare and the cockroach[16] can be at least partially explained in these terms. That the agency of more complex creatures is capable of explanation in these terms is a matter for ongoing research. Actual cognitive processes may turn out to be a blend of serial and parallel combinatorial processes. The hypothesis of the this thesis is that whatever structural form these processes take they will be apt for interpretation as realisations of combinations of representations of the world as it is for the agent.

[15] 'Caused by' as the realisation of 'functionally related to'.

[16] Ritzmann, (1993), cited in Clark, (1997).

Akins, (1993) has given a detailed neural and cybernetic explanation of the agency of the Mustached bat as manifest in its hunting. She writes, (p.151):

> " ... what science suggests is that the sonar system of the bat is probably *not* concerned with the representation of three dimensional objective particulars. ... There being, that is, no particulars *for it*, we should not ascribe to the bat perceptions of those particulars ... Because there are no objects that the bat perceives, there are no objects for the bat to perceive *in a certain bat-like way.*"

Akins, in this passage, appears to be arguing against a symbolic usage of representation. In a theory of the agency of the bat, the bat's neural states represent and what they represent is a state of the world. This state may be described in our theory as a moth or a landing place or a cave layout or as combinations of Doppler shift, azimuth and amplitude. Earlier, (p.150), she writes:

> "The bat's auditory system seems to be designed for sequenced non-integrated information processing."

Alternatively stated, the behaviour of these complex creatures can be explained in the functional terms of agency but this agency is realised in quasi-autonomous cognitive processes. Although these processes combine to facilitate survival of the bat, their agency is not co-ordinated to this end and there is nothing that it is like to be a bat. The goals of the bat that are functional for its autopoiesis are served by the totality of its cognitive processes but these are not integrated cybernetically. Each example of agency that is realised in a bat can be functionally described in terms of perception, representation, cognition and action. We may describe the perception as 'of a moth', and the act as 'following, sending and catching' but the content of the representations is most robustly defined as of physical properties like Doppler shift, amplitude and angle of azimuth. In this domain, perception is of these properties and each action is a physical 'hill climbing'[17] response to a perception

The world of the bat is very different from that of the cat. The neurological mechanisms of the cat will provide for large scale integration of visual, audible and olfactory data and it is tempting to conclude from this that the representations of the cat are of real objects out there, though we may have no justification for this[18]. A Normal token cat will respond to stimuli that are relevant for it and in these responses it will track objects in its world. The perceptual mechanisms of a cat are transducers just as are ours and the bats, and representations of the cat will track such states of the world as size, speed, shape, smell and sound. Combinations of such representations may be 'recognised' and specific behaviour may have adapted to previous experience. In this case the agency of this cat will be manifest in its actions about these perceptions of recognition. We, in our observations of the cat, may interpret these perceptions as classification by the cat of its world. We may name the classes 'picked out' in terms that denote aspects of the world for us and we may translate these terms into others that denote aspects of the world that appear to us to be *for the cat* in our theory of the agency of the cat. From this analysis we may explain the behaviour of the cat in terms of concepts like prey or food source or comfort. These concepts are possessed by us but not by the cat. The conceptual theory established may be sound in that it explains and predicts the behaviour of the cat but the intrinsic representations of agency that are realised within that cat are not representations of these concepts for the cat. It may be that neural states of the cat correlate with realisations of these concepts: indeed,

[17] Using the terminology of McFarland , (1989).

[18] Dennett, (1991,43) denies animals consciousness since he insists that consciousness is only having a "centre of narrative gravity". However, "Human beings are not the only creatures smart enough to suffer." (449), If 'being conscious' is not the right term for 'being sufficiently cybernetically integrated to manifest a multi-sensory world view', then we need a term and the Dretske, (1993, 272), notion of 'thing consciousness' may apply.

this is likely for a mature cat in a stable environment. However, although these two functional theories trade in the same coin they are not dealing with the same goods. If, within our theory of the agency of the cat, we speak of perception of prey, representation of prey and action of pouncing on prey, then we gain in interpretation but we lose in realisation. We, not the cat, have decided that 'that is prey'. I contend that our attribution of conceptual content to type III representational states should be seen as a theory of attribution and not as a functional explanation of phenomena realised in the world within the cognitive processes of the agent.

Further, I will argue in 4.11 below, that our attributions of beliefs and desires as representations with conceptual content should also be seen as within a theory of attribution that we call folk psychology. Also, I will argue that the functional representations of folk psychology are not realised within the cognitive processes of the human agent. The content of intrinsic representations of agency within a human agent is generally nonconceptual. A representation may involve the recognition of phoneme structures within their parsing, it may involve the recognition of scripts evoked by such parsing. However, for the agent that is animal and human, these are combinations of intrinsic representations caused by perceptual events, past and present.[19]

Extrinsic and conventional representation within language must be distinguished from intrinsic and tacit representation within agency. The content of those representations in language that deal with representations within agency is wide in so far as it is concerned with internal states and their representations of an external world. The content of statements of propositional

[19] Models of the neural processing involved in such parsing may deploy the symbolic representations of serial processing or the sub symbolic processing of neural networks. The partial stages of convergence within a neural network are not apt for interpretation as representations of agency, although the recognitions to which they converge and some stages in the serial processing may be. However, the symbolic representations are representational for the cognitive scientist. These distinctions are considered in Chapter 5, below.

attitudes is also wide. Such statements are attributions by which we may explain the behaviour of agents who are sufficiently like ourselves to make the attributions of our own folk psychology plausible. We might want to say that the frog believed there was a fly but be inhibited by our knowledge of the simplicity of the mechanisms of the frog's cognitive processes. However, the fact that our own cognitive processes include learning, are far more complex and can only be mapped on to the world, in a theory of agency, by the positing of many goals does not change the fact that they are mechanisms. We share a continuity of realisation of intrinsic representations in states of neural circuitry with other vertebrates. Our intentional activities can be functionally explained in examples of the same theory of agency that we may use to explain the actions of any living thing.

Conceptual content is within language. Peacocke, (1992, 23) wrote: "Possessing a concept is knowing what it is for something to be its semantic value." Possession of a concept by a human agent implies that the agent has intrinsic representations within agency with content of something (the semantic value) and intrinsic representations within agency with content of the well formed strings in language that denote that something. Concepts are abstracta, realised in the practices of language. Representations with conceptual content are realised only within sentences of language spoken by people to people about things, concrete or abstract, in their world.

Ethologists have a different concept of 'concept'. Allen and Hauser, (1991), define 'concepts' as

> "[C]apable of explaining complex abilities to generalise over variable stimuli, to rapidly produce appropriate responses to the common features underlying these stimuli and to modify behaviour when it is discovered that perceptual stimuli are unreliable guides to underlying features."

Such *concepts* are those with which an observer may construct a theory of the agency of a complex organism in the world. In such a theory of agency, representations will be comprised of such concepts but in the Peacocke sense, these are concepts possessed by the theorist.

3.7 Misrepresentation Within Agency

The function of a representational state of agency is relative to the explanation of agency that includes that state and this explanation is within a theory of agency that explains this type or token of realisation of agency. If this theory is true then it will explain the Normal functioning of agency in terms of representing states and states of the world that are represented; this usage of 'Normal' is that of Millikan, (1984). However, the Normal functioning of the processes of agency is contingent on events in the world, within or outside the physical boundary of the agent. It may be that, due to some neural flaw, a representational state fails to track events in the world and this state may now be that appropriate to some other state of the world. Alternatively, it may be that no conceptual interpretation exists for this state. In so far as this state can be said to represent then it misrepresents.[20] It may be that the organs of perception, the transducers that realise the function of perception for this agent, are faulty or that some physical extension of agency, such as a clear view of a state of affairs, is impeded. If so, then any representational state that functionally relies on this perception will misrepresent. Intrinsic representations within agency may or may not be realised by binary processes that cause unequivocal representations. The tracking of representations may be vague but this need not invalidate a theory of agency by which agents act in a world in ways that are for their goals by perceptions of that world that cause

[20] Misrepresentation is considered again in Chapter 6 in the context of the computational theory of cognition.

representations of that world. Such representations have content for the agent. Their representation will be tacit if their interpretation is in action rather than language. Their content will be nonconceptual for the agent if they are not explicitly interpreted by the agent. If the content of a representational state is not that which is Normal for the cognitive processes of that realisation (or embodiment) of agency then that state misrepresents. The state may not cause inappropriate action but the cognitive processes leading to that action will not be Normal.

Intrinsic representations within agency will perform their function Normally if :

(i) the perceptual processes that maintain them are functioning Normally

(ii) the cognitive processes that maintain them are functioning Normally

(iii) the processes by which representations cause action are functioning Normally.

In animals that have evolved and in artefacts with complex agency, Normal cognitive processes can be expected to be truth preserving. Each causal process by which representations combine to lead Normally to another representation can be compared with the introduction and elimination rules for the logical operators. For example, in combinatorial terms a cognitive process may be equivalent to an AND gate. Its occurrence will be equivalent to & introduction and behaviour based on both conjuncts will be equivalent to & elimination. Another may be an isomorphic translation between representations in different physical media. Its occurrence will be equivalent to $\supset$ introduction and behaviour based on the inference will be equivalent to $\supset$ elimination. It is by evolution or design that the cognitive processes are *for* the goals of the agent. If this *for*-ness is robust to a variety of relevant events in the environment then we may expect that successful cognitive processing will be processing that Normally leads to survival or the achievement of design goals and will comprise representational processing that preserves truth. Abnormal processing, due perhaps, to neural or component failure, will not be systematically truth preserving and may lead to misrepresentation[21]. The processing itself will be topic neutral. It is because of the relative stability of environments, that through the accidents of evolution, cognitive processing will be, generally, apt for interpretation in terms of truth preservation.

This is not to say that the interpretation of Normal representations by a theorist in a theory of agency will Normally yield content that is true. Deceptive strategies by a guided missile may involve representations of the false. An animal may have evolved to behave in ways that consistently overestimate the threat of an antagonist. But the need for adaptive response to a contingent world implies that truth preservation in Normal cognitive processing will be advantageous for design or evolution. Since such processing is designed or has evolved to cope with a range of contingent events we would expect its logical interpretation to be topic neutral.

3.8 Agency, Regulation and Cybernetics

In technology we say that an entity is regulatory if some state of the entity within its world is maintained at or near some goal state against variation in that world. Regulation may be within an entity in that the proximal cause of variation of this state is within the entity. Alternatively, it may be within the world in that the entity physically changes its relation to the world, which is the environment of the entity, in order to achieve or maintain some relation to the world against change in the world. It is the latter form of regulation that I wish to compare and contrast with agency. This is the regulation which Sayre, (1976, 54), dubbed 'heterotelic'. It is the process by which an entity maintains some

[21] It may be that, by chance, the content of the resultant representation is Normal but I will rely on a notion of 'justification' in Normal representation and avoid such complications.

state of the entity in its world by active causal intervention in that world. The entity is in the world and it is of the world. Vital states of the entity may include energy levels, which are depleted by operation of the entity. The counter-entropic maintenance of order within the entity requires importation of energy from the environment and this importation requires principled intervention in the world, which is the environment of the entity.[22]

Regulation is about the world and it is about the world as it is for the entity. Successful regulation requires perception of the world and representation of the world as perceived. It also requires a structure by which the state of a representation leads to the act required to change the world in a way appropriate to its goals. Some of our artefacts include in their functional specification the capacity to act on the world to maintain certain properties against changes in the world. We call such artefacts regulators and the theory of such artefacts is the discipline of cybernetics.

> Since the requirements of regulation are functionally isomorphic to the requirements of agency I claim that agency is functionally equivalent to regulation and that the theorems of cybernetics apply to all realisations of agency.

The term 'cybernetics' was first introduced by Wiener in 1948 and derives from the Greek for 'steersman'. Wiener's interest at that time was in the problems of tracking, control and prediction associated with anti-aircraft artillery and cybernetic systems were seen as open systems designed to track a moving target or maintain some state of a system against a varying environment. In a seminal essay in 1943, Roseblueth, Wiener and Bigelow linked the teleological concepts of 'behaviour; and 'purpose' to the engineering concept of feedback but only in

[22] Many authors have considered the connection between entropy, cybernetics and life. Brillouin, (1949), gives a convincing and accessible account.

the sense that purposeful behaviour requires negative feedback. This thesis has been attacked by Richard Taylor, (1950 and 1966), on the grounds that the purpose of an artefact with feedback can only be the purpose of its designer. For Taylor 'purposes' are transcendental to the physical embodiment of agents. "[T]he concept of purpose is as basic a category as that of cause and effect. Neither can be reduced to the other." (1966, 256).

Ludwig von Bertalanffy, in a series of books and articles from 1929 to 1968, has developed the general systems implications of living organisms. Many writers including Nagel, (1979), have discussed the relevance of cybernetics to biology and organic behaviour. Shannon, (1948), introduced the notion of information in telecommunication and this has been linked with the counter-entropic implications of the maintenance of structure by purposeful or teleological, control. Rosen, (1991, 244), concludes that: "[A] material system is an organism if and only if, it is closed to efficient causation." And: "Biology becomes identified with *the class of material realisations* of a certain kind of relational organisation, and hence to that extent divorced from the structural details of any particular kind of realisation." (245) Rosen's *relational theory of systems,* (117) is close to the methodology of functional analysis deployed in this thesis and his notion of a causally closed system is closely analogous to the notion of agency expounded here.

3.9 Contingency

Recently, Corning, (1995) wrote, "Self determining systems are also cybernetic systems. And cybernetic systems are controlled by the relationship between endogenous goals and the external environment". He quotes Powers, (1973), as describing organic goal seeking systems in terms of error correcting systems. That is, systems which can be described mathematically in terms of a tendency to oppose an environmental disturbance of an internally controlled quantity. This is the classic concept of negative feedback control and Sommerhoff, (op cit.), opposed its application to explaining the endogenous behaviour of an organism within its environment. He wrote (p117) "The control of goal-directed activities in servomechanisms and other automata is commonly effected by means of error signals which inform the control units about the magnitude and direction of the discrepancy between the desired output state and the actual one. It is frequently taken for granted that this must also be the case in the motor control functions of the nervous system. ... In so far as the overt acuities of living systems are goal directed they may, of course be viewed as error-reducing or error-eliminating activities. But this does not necessarily mean that their control mechanism depends on the generation of explicit error signals of one kind or another."[23] Sommerhoff also described the principle of directive correlation. He distinguished between goal seeking and equilibrium seeking and argued that the actions of agency (though he spoke of goal seeking) are correlated with the events in the world that provoke them but these events are orthogonal to the states of the agent. He called this sort of correlation 'directive correlation'. In his words: "In the case of the directive correlations the variables concerned in the focal conditions are orthogonal variables. But in the equilibrium case the variables are not orthogonal." In his example of a pendulum, the pendulum is not responding to an unpredictable environment with appropriate behaviour. Regulation and agency depend on the essential orthogonality of perception to current representation. The statistical notion of correlation is defined in terms of the extent of correspondence between the ordering of two variables. it is neutral with respect to causal role. Directive correlation is concerned to elucidate the causally correlative implications of agency. Since agency is concerned with action that is for the agent in the context of the world, the causation between states of the world due to agency

[23] This conflation between 'negative feedback' and 'regulation' has confused discussion in this area since the phrase was used by Rosenblueth et al, in their essay of 1943. If what is fed back is extended to include information and the notion of 'negative' is extended to include 'relevance to deviation from a goal' then the terms become synonymous but at a loss of semantic precision. Sayre, (1976), accepts this loss. Bertalanffy, (!968, 150), distinguishes between kinetics and thermodynamics in open systems and feedback and information in cybernetics but he equates

are essentially circular (or spiral) when agency impacts on the world. Representational states of agency are causally correlated with the states of the world that they represent through perception. States of the world are causally correlated with intrinsic representations within agency through action. Although states of the world will be statistically correlated through acts of agency, they are epistemically contingent for the agent[24] in each act of agency.

This notion of environmental orthogonality is vital: it is a matter of a priori definition. If at least some of the changes of the environment of an entity were not orthogonal to that entity in the way described then the behaviour of that entity would be programmed by its constitution. The very idea of regulation implies an orthogonality of environmental change to agent structure. If we think of representational content in epistemic terms then at least some of the perceptions of agency are epistemically contingent for the agent. In a complex organism acts of agency will be woven in with actions that are planned. The actions of a predator such as a lion, engaged in hunting, are to an extent, stereotyped by learning and inheritance but at least some of the actions will be occasioned by the unpredictable. The very choice of that zebra will be contingent on that zebra being there, in a way that can be singled out by the lion at that time and the evasive movements of that zebra are in detail unpredictable for the lion.

The behaviour of examples of austere agency such as single cells, homeostatic mechanisms, reflexes, tropisms and simple control systems is characterised by a limited repertoire of response to a limited range of environmental variation. Each performance is to some degree, unpredictable for the agent; if it were not

cybernetics with the feedback cycle and relegates the feedback model to "secondary regulation".

[24] Thanks are due to Bob Hale for discussion that led to this insight. The notion of 'epistemic contingency' seems to have been introduced by Taylor, (1966, p. 45). My usage of this notion is much stronger than Taylor's since I hold that agency entails epistemic contingency.

then its function would be of program rather than control. But without learning, the behaviour of the entity can be predicted, by a theorist, to remain within a limited range. The richer agency of organisms with nervous systems and complex artefacts such as robots and guided missiles is characterised by a rich variety of response so that a rich variety of unpredictable environmental change brings forth a rich variety of behaviour.

3.10 Variety

Complex environments require complex representations. Ashby, (1956), considered the theoretical requirements for a system to maintain complex states against variation in a complex environment. His law of requisite variety[25] states:

> In a regulator R controlling against variation in an environment D, the variety of outcome in R cannot be less than the variety of D divided by the variety of R.

Ashby drew an analogy between the regulatory requirements of an organism coping with a varying environment and a mechanism, which might be designed to restore the information lost by a noisy channel in a communication system. Shannon, (op cit.), showed that if the information rate lost due to noise is N then the information capacity of a correcting mechanism must be at least N[26]. This can be seen as analogous to variety in regulation. In Ashby's words, "R's information capacity as a regulator cannot exceed R's information capacity as a channel of communication". Organisms that have survived have a complexity of possible response that matches the complexity of their environment. Once an

[25] Ashby's proof of this theorem is summarised in appendix 2

[26] Shannon writes in terms of conditional entropy and does not use 'N' but the argument is unchanged.

organism has evolved to relate to relations between events and in our case, has evolved to cope with entities that are representations of representations of relations between events the information becomes incommensurable but the basic principle remains. For an agent to be viable in its environment the representational content in the agent, measured as information must be at least equal to the content in events in the environment that are relevant to the agent. This can also be seen as the design challenge for the designer of any artefact in which the function of control is delegated to mechanisms. We can study tabular representations of information input at our leisure. In the real world of control in an environment of epistemic contingency the representational and computational mechanisms are processing information in real time. Information rates can be measured in bits per second or bandwidth, but this is still the mathematical description of the requisite complexity and processing capability of an agent in an environment of complexity and change.

I draw three inferences for agency from cybernetics.

> (i) If agency is not a playing out of a prescribed programme then the perceptions of agency that are caused by changes in the world of the agent are necessarily at least sometimes epistemically contingent for the agent.
>
> (ii) For an agent to be viable in its world the variety of its representations must be at least equal to the variety of world states that are relevant for it.
>
> (iii) The ability of an agent to handle the variability of its environment, in space and time, can be characterised in the measure of information. in these terms: the information capacity of cognitive function viewed as a communication channel must be at least equal to the information rate of relevant change in the environment.

3.11 Realisations of Agency

To avoid the circumlocution of 'entities whose behaviour is apt for description in terms of agency', I will speak of all such entities as agents. All agents then are either natural agents, including us and are living things whose agency is realised in biological form, or artefacts of agency in which the realisation of their agency derives from their function for us. Following Cummins, (op cit.), I take the explanatory role of function as relative to its necessary contribution to the capacity of a containing system[27]. For a living agent this is its capacity for intentional action. The agency of a living thing is manifest in the capacity by which it maintains its autopoiesis in a changing world. Although other species use tools, as far as we know, we are the only species that has produced artefacts that satisfy the minimal functional requirements of austere agency. If we design a thermally controlled chamber then we will produce a specification that includes the measurement of temperature, a supply of heat and a mechanism by which changes in temperature leads to the release of heat in such a way that the chamber is maintained at a desired temperature. Maintenance of temperature was our choice for the capacity of this system and since we designed it to do this, we may speak of this capacity as its goal. We may call such a device a regulator and our functional description of this device would describe the measurement as perception, a state of the device as a representation of temperature and the controlled release of heat as an act of agency. Any artefact that regulates can be described in these terms. This is an austere functional characterisation of regulation but any artefact that realises a richer regulatory capacity can also be described in these terms.

Within the axioms of this thesis, living things have not been designed to have goals. External explanation of the behaviour of complex organisms in their

[27] Nagel, (op cit. 292-4), takes Cummins as espousing a teleological view of 'function'. Nagel equates function and teleology, Cummins does not mention teleology and he writes of capacities rather than goals.

world is in terms of their realisation of agency with permanent or transient goals and internal explanation is in terms of their realisation of autopoiesis. Their existence is explained by the accidents of evolution and autocatalysis.

Within my austere characterisation of agency all living things have agency but all agents are not alive. A living cell has agency: it is characterised by action, which is by it and for it. Within the closed containing system description of its autopoiesis[28], its goal is the maintenance of its own organisation. The cell may have function within an organism, described as a containing system and that function may be within an organ, such as a heart. The presence of the heart can be explained by its function within the containing system of the organism and the heart can be described as a containing system, within which each muscle cell and each neuron has function. The function of the heart is to pump blood for the organism and the function of blood is to transport various materials for the organism. A muscle cell of the heart is alive, a blood cell is alive, a neuron is alive. An active heart, pumping blood in a living organism is not an agent, neither is it alive. It is not an agent since it has no perceptions, its actions are not for it they are for the organism. In the functional characterisation of 'life' as implying self maintenance of substance and form, as opposed to the everyday notion of life as characterised by its biological context, it is not alive since it is not autopoietic[29].

Homeostatic mechanisms within organisms meet our criteria for agency. The human endocrine system includes a feedback mechanism whereby the circulating blood hormones (from the endocrine glands controlled in this way) inhibit the hypothalamus and the anterior pituitary. Such a feedback mechanism

[28] Autopoiesis as the system characterisation of life and its interconnection with agency is described in Appendix 1.

[29] Varela, et al, (1974) gave a six point key for identifying an autopoietic system, which is quoted in Appendix A. The heart fails the fifth criterion, the components of the boundary of the heart are not produced by the heart.

enables the levels of hormones within the blood to be maintained within close limits. This mechanism is not alive since it is not autopoietic, it is apt for description in terms of agency since its perceptions result in actions that are for it in terms of its capacity (to control blood hormone levels). Reflex mechanisms such as a knee jerk or a blink can be similarly described but such mechanisms are not alive although such mechanisms could be described as having natural function for agency.

The brain is organisationally closed, Varela, (1979) and is self organising, Edelman, (1992) but it is not autopoietic since it does not maintain its own boundaries or its own components, though it does maintain its own organisation. The brain does not possess agency either within an open system description of the organism in its environment or within an open system description of the brain in the organism. In neither of these does the brain act for the brain. Thus, although the representations of the brain are intentional, the brain is not an agent since its representations, within the open system of the organism in its environment, are not for it; they are for the organism.[30] In a functional explanation of human agency the brain is the principle organ by which representations change representations within the processes of cognition and this is realised within our neural structures and the electrical and chemical processes of human cognition.[31]

3.12 Conclusions

I have argued that if our concept of human agency rests on action that is intended by the agent under an aspect of a description then agency can be functionally explained in terms of perception, representation, cognitive process, goal and

[30] Functional components of the brain may develop and operate by a "reentrant processing"., Edelman, (1987),.and could be described in terms of a 'local agency'. But they do not act in the world of the agent with the brain.

[31] Lyons, (1995, 161-4) writes of brain-level intentionality. However, although the brain is the main site of representation and cognitive processing in the human realisation of agency, the

action. This austere characterisation of agency gives a functional theory that explains the goal seeking behaviour not only of all organisms but also of all of our regulatory artefacts.

The interpretation of intrinsic representations within agency by a theorist will depend on the theorist's description of the world of the agent. It will include things of the world of the theorist or properties of the world that are recognised by the agent in a given domain of description. Thus the content of such representations is conceptual for the theorist but is generally nonconceptual for the agent.

In a Normally successful agent the processes of cognition will be realised within a mechanism and the processes of that mechanism may be explained within a theory. If the theory is true then Normal cognitive processing will follow that theory. If some token processing event is abnormal then processing subsequent to that event will be abnormal. States caused by such processing may or may not be interpretable as representing within a theory of agency. In so far as they can, then their conceptual content attributed by an theorist in that theory will misrepresent. It may or may not be false but it will not be justified within the system of justification presumed by the theory. With respect to the survival of the agent or achievement of its goals, such changes may be beneficial or harmful, depending on luck and criteria. By and large, they may be expected to be harmful since Normal processing will have been designed or evolved around the survival of that agent type at least until the achievement of its goals.

Events in the world that drive the processes of agency include some that are essentially unpredictable for the agent. To be otherwise would be to be programmed. Actions of agency are determined by events perceived and the

intentionality of the representational states of the brain is within the theory of the agency of its owner.

current constitution of the agent. The current constitution of the agent includes the current cognitive processes that realise the current goals of the agent. Thus, although a theorist with knowledge of the current constitution of the agent and able to predict events in the world could predict actions of agency, these actions are necessarily, to some degree unpredictable for the agent since they depend on epistemically unpredictable contingent events.

The representational states of agency are real states that realise a representational function for the agency of the entity. They are not states of belief though I will argue that, through the functionality of practices of belief attribution within a linguistic community, the extrinsic representations of belief statements are closely associated with intrinsic representations within agency.

4

LEVELS OF AGENCY

4.1 Introduction

Our concept of agency derives from our experience of the rich variety of acts of human agency. These range from our everyday decisions regarding what to wear or eat or stepping round an obstacle to more salient choices regarding such issues as career or mate or the local football club. However, as argued in the previous chapter, these manifestations of agency are complex, integrated examples of the behaviour of entities each apt for description as a system with capacities that can be explained in simple functional terms. The terms are those of an entity that acts in the world so as to maintain its goals in the world in the context of its changing perceptions of that world. The mechanisms by which these acts are chosen we call representation and cognition. If each action of agency is intentional then the content of that action can be construed as the goal of that action and 'goal' and 'intention' are synonymous in this usage. But this is not my usage of goal in this context. The goal built in to a central heating system is to keep a house warm and it will have been designed to pursue that goal against random variations in the environment. One of my goals is to stay alive and I will pursue that goal against contingent variations in my environment.

Realisations of agency within our world include all organisms and some entities that are combinations of organisms. Artefacts of regulation also meet the functional criteria of agency as do artefacts that regulate by combinations of regulators. We humans extend our agency by our use of artefacts and this combination may manifest agency within another aspect of a description[1]. In this chapter I will outline the range of complexity of actual realisations of agency in our world and will classify these by the distinction between organism and artefact and by the dimensions of complexity and combination. To illustrate the range of this classification, consider the simple example of a man driving a car. Put aside the complexities of navigation and control and consider the two quite separate issues of control that are concerned with fuelling the car and keeping the engine cool.

In the first example, I drive my car into a filling station to get petrol. My decision to get petrol was based on my reading of the fuel gauge, my beliefs about its reliability and my knowledge about cars and filling stations[2]. To explain the relation between petrol in the context of the car, and the car in functional terms, we need a wider functional containing system. The petrol has function for the car, the car and the petrol have function for me and I decided to get petrol. My agency is manifest in; my perception that the gauge is reading low, my representations regarding this and its implications and my cognitive processes that maintained the representations and caused my actions of stopping at the garage and getting petrol. Running low on petrol is an epistemically contingent event for me as I rely on the gauge to tell me when I need petrol.

For a second example, we may note that the engine of the car is cooled by water, which is circulated round the engine by a pump. The heat generated by the engine varies in a way which depends on variables such as speed and terrain that Normally are causally independent of (orthogonal to) the water temperature. In order to maintain the engine temperature within a close range,

[1] As an example consider the use of a burglar alarm. The act of seeking an intruder is caused by the perception of the alarm, which is caused by the burglary, but the setting of the alarm was a previous act of agency.

[2] This is in the vernacular of folk psychology. The relationship between belief talk and representational states of agency is discussed in 4.11 below.

the manufacturer has fitted a device which increases the water flow when the engine temperature rises and decreases it when it falls and this arrangement keeps the engine temperature reasonably constant. A valve lifts when the water gets hotter and this state represents the water temperature for the device. The entity, which is the device, has no input from my decisions in driving the car and without intervention from me, the right decision on water flow is made. Hotter water is an epistemically contingent event for the device as the device relies on water temperature for its operation. The water flow has function for the engine, the engine has function for the car in its function of moving me around but the decision to vary the water flow was not made by me.

Although getting petrol had function for my car in its function for me the car had no perception of its lack of petrol or of petrol in the filling station. I had these perceptions and I was the agent of the act of getting petrol. The device in the cooling system perceived the engine temperature via the water temperature and was the agent of the action of opening the valve. I had no perception of the engine temperature and had no agency in that action, though others had agency in delegating that agency to a device. In each case the appropriate action was taken by an entity, based on facts in the environment of that entity and causally available as action by that entity for the goals of that entity. In each case we can describe a functional containing system which includes every component which has function in the explanation. When I got petrol that action was an expression of my agency. The petrol was for my car, for me. When the thermostat in my car opened to cool the engine a little, the function of that action was for my car, and the function of my car was for me. In that description, the function of the agency of the artefact was for me but I had no agency in that action.

These are but two examples from a vast range of complexity and co-ordination. In what follows I will illustrate this range by examples of agency in our world. The first two sections describe examples of simple agency in which questions of complexity and co-ordination do not arise. More complex agency may be realised in the complexity of cognitive processes for a single agent. Alternatively, it may be realised in the co-ordination of agency in which some of the autonomy of an agent is subservient to a larger agency within which and for which, the subservient agent has function. Examples of this have already been considered in 3.6, above. In later sections, these will be extended to a consideration of the agency of social groupings of agent organisms that are themselves co-operative groupings of living cells and to social groupings of agent organisms some of which are also apt for description in terms of agency.

4.2 The Agency of Simple Organisms

Not all organisms manifest agency by purposeful action within their environment. Several bacteria and amoebae display behaviour which is random with respect to their environment and take nourishment and harm from the environment as it falls, but all can be described in terms of mechanisms which maintain their internal states and metabolic and reproductive rates by some form of homeostasis. Homeostatic regulation meets the functional criteria of agency but here as in the cooling system of a car, the environment is the environment of the controlling mechanism and is within the entity.

The "lactose system" describes the mechanism which govern the synthesis of three proteins in E. Coli. In this system a regulator gene directs the synthesis, at a constant slow rate, of a repressor protein, this repressor interacts with the DNA to inhibit production of the three proteins, the repressor protein is inactivated by a molecule of a an inducer galactoside which is itself inhibited by the three proteins[3]. In this homeostatic, regulatory mechanism the metabolism of the cell is maintained through the operation of the DNA of the cell and protein need is represented by the presence of the galactoside. The same organism controls its location in its world by the heterotelic regulation of

[3] This process is discussed in depth by Monod, (1971).

agency. The movements of E. Coli are generally a random tumbling but "The normal swimming behaviour is modified by chemotactic attractants or repellents that bind to specific receipt proteins and affect the frequency of tumbling by increasing or decreasing the time that elapses between successive changes in the direction of flagellar rotation." Here, perception is realised by the binding of an attractant and the activation of three appropriate proteins, representation by the level of these proteins and action by a change in rotation of the flagellar motor[4]. This species of bacterium has survived because its regulatory behaviour leads to nourishment and avoids harm and this is the how this organism maintains its essential autopoiesis by its agency in the world.

Since such simple creatures as amoebae and bacteria lack nervous systems their behaviour is simple. This is an advantage in illustrating that their actions are regulatory and meet the criteria of simple agency but they do not illustrate how more complex organic agency would involve more complex representation and cognition. To isolate a neural state as a representational state is to describe a physiological state with a particular functional role in the shifting pattern of physical states within the nervous system. This is just possible for a range of invertebrates that includes the garden slug, the pond snail, the medicinal leech and the sea hare[5]. Lockery, (1989), describes the gill-withdrawal reflex in the sea hare and how second order conditioning of this reflex may be realised within the neural circuits. If this conditioning illustrates Normal or potentially Normal learning for the sea hare then the neural circuits described are realisations of representations of learning within agency.

The agency of complex organisms depends on the co-ordination of groups of cells into organs that have function within a theory of agency for each organism, as the gill is an organ with function for the sea hare and a muscle has

[4] This account is taken from Alberts, et al., "Biology of the Cell".
[5] This list is taken from Lockery, (1989)

function for a vertebrate. Each muscle cell is a living organism: it has agency. But, within the co-ordination of various proteins, following programmes realised in DNA, this agency is subordinate to the function of the organ for the agency of the organism. The agency of a muscle cell is not sensibly described as autonomous within its environment. Its autopoiesis depends on a supply of nourishment that is not acquired by its agency and its agency is limited to its perception of certain proteins and through its representation of these, its actions of contraction or relaxation according to these commands. A bacterium such as E Coli, flagellating away in its intestinal soup, may be less complex in its agency but its agency is autonomous within that environment.

We attribute autonomy to an agent within a description. If we were explaining the dynamics of an ecology we would describe how the agency of members of a species is limited by and maintained by the agency of other species. Each transaction of agency for each token agent is a unique transaction based on perception contingent for that agent but the scope of such transactions is limited by the structure of the ecology. An agent is autonomous under a description of its agency within an environment in so far as its actions are only caused by its perceptions of that environment. Autonomy is a system notion, as function, agency, regulation and autopoiesis are system notions.

4.3 The Agency of Artefacts

I have already discussed the agency of simple regulators such as a thermostat in a room or a car. Such devices are limited in their scope and do not demonstrate the distinction between states of representation and processes of cognition. Consideration of more complex devices such as robots and guided missiles will require this distinction but the technological complexity may obscure the simplicity of the issues involved. A mid course is to describe a functional specification for a device that is technically feasible and at the right level of complexity of goals and environment. We can then consider possible realisations of this specification and some of the philosophical implications of its agency.

Consider an environment consisting of sources of sound and light at the periphery of a circular arena. In this environment we place a motile device equipped with a sensor for sound that will produce two signals; one a measure of amplitude and the other a measure of angle of source relative to the device, and a similar sensor, also producing two signals, for light. These signals are fed into an integrator, which is designed and built to produce a navigational output such that the device will steer towards light and away from sound. For each source detected, the device would produce three signals: a quantity proportional to brilliance of light or volume of sound, a quantity proportional to the angle of the source relative to the orientation of the device and a binary signal indicating whether the source is of light or sound. The sensors would be designed to rotate until a signal is detected and then track the signal as the device moves.

If more than one sound or light source may exist then the device can only pursue these multiple goals effectively by the provision of more sensors and a complexity of programming to integrate their output or by faster rotation of the sensors. Either will require the provision of a 'learning ' capability to identify different sources as 'objects' in the world. Ashby's criterion will apply and the device will require variety sufficient to match the variety in its world.

The device would be designed to move at a constant velocity in the direction given by the summation of the input vectors. For the simple case of one sound of volume v and one light source of brightness b, at angles α and β to the orientation of the device, the device would move at an angle θ where

$$\theta = \tan^{-1} \frac{b \sin\alpha - v \sin\beta}{b \cos\alpha - v \cos\beta} \qquad (1)$$

Also, the device is designed to sense when its forward progress is impeded and stop until it is free to move. Thus, the device will follow a path that maintains the relative weightings between light preference and sound avoidance and constraint on motion. If the sources are constant in amplitude and location then the device will come to rest at some balance point or against the boundary wall. If the sources move then the device will move, if unimpeded. The device is then *free* in the Lockean sense. Some implications of this usage are explored in Chapter 6.

What such a device would be useful for is a matter for speculation. What is important for our purpose is that the device has its own goals, albeit given by us in its specification. Its actions are intentional under an aspect of a description. We may describe the device in terms of its sensors and their outputs and the causal processes by which these outputs lead to turnings of wheels and their rotation. In another description we may speak of the perception of sound and light and location and the action of moving toward light and away from sound. In the latter description this action is intentional. In a theory fully understood by us, it perceives and it represents and its acts are the outcome of cognitive processes on representations.

We can use this specification as a test bed for the theory of agency. This is not an exercise in artificial intelligence or artificial life. Any realisation of this specification would be necessarily artificial and the notion of intelligence has not been invoked. Any realisation of this device would not be alive since it does not meet the criteria for autopoiesis.

The functional specification for this device was based on an environment of sound and light and it was required that the entity be attracted to light and repelled by sound. Attract and repel are not of themselves intentional, magnets attract and repel. The intentionality of the device, under an aspect of a description, comes from its functional specification, which implies that its behaviour in this environment should be capable of description in intentional terms. Within that description, the Output State from each transducer is a realisation of an intrinsic representation within agency. The output of the sound transducer is a vector of two voltages. One of these is nomically related to the angle between the front of the device and an air pressure wave front, the other is nomically related to the instantaneous average power density of that wave front and it is these properties that are causally efficacious in further processing. The functional role of this state in the agency of the device is to represent the direction of the sound relative to the device so that the device can retreat from it. Not, where the sound is, as location has no relevance for the device within its intentionality. Its intentionality is necessarily indexical since that is how it is defined.

The functional specification left little room for ambiguity since sound is hard to confuse with light. However, we could specify a second device of greater complexity. Suppose we specify that red light or high pitched sound repel and blue light or low pitch sound attract. Then the intentionality of the device would consist in four distinctive behaviours, depending on the combinations of colour and pitch. The output from the sound sensor is now a triad of direction, volume and pitch. Pitch now has a place in the world of the device. In each token event of sound arriving at the first device the sound will be at some pitch but this pitch is irrelevant to the device. The world of our first device did not contain features distinguished by pitch.

We now have sufficient complexity for the issue of misrepresentation to bite. For the second device, a sound may be misrepresented if a sound source that is high pitched in its normal implication for the device causes a low-pitched action. For each source detected, the device would produce four signals: a quantity proportional to brilliance of light or volume of sound, a quantity proportional to the angle of the source relative to the orientation of the device, a binary signal indicating whether the source is of light or sound and a binary signal indicating the hue or pitch. The last of these of these would be discriminated by a filter and this could misinterpret.

The importance of history in function has been stressed by Millikan, (1984, 17-38). For Millikan, to be a direct proper function is necessarily to be a member of a reproductively established family and reproduction depends on a similarity of causal history and a Normal explanation is an explanation of how a particular reproductively established family has historically performed a particular proper function. For our second device; what the pitch Normally causes is determined by a design history that has established a Normality of outcome. Whether boundary cases represent or misrepresent will depend on what is Normal, which will depend on the design history. There is no issue of disjunction here since the functional specification did not include ambiguity. Were we developing a theory of agency for a strange device then our question would be of what is Normal for this device.[6]

We could also apply Fodor's, (1990), criterion of asymmetric dependence. This criterion relates the possibility of a token representation misrepresenting to the inferential dependence, within a theory, between Normal representation and misrepresentation. If the misrepresenting causal relation is inferentially dependent on the normal representing causal relation but not vice versa then the former relation is said to be asymmetrically dependent on the latter. In Fodor's, words, p. 182: "All that's required for 'cow' to mean *cow* ... is that some 'cow' tokens should be caused by (more precisely, that they should carry information about) cows and that non-cow-caused 'cow' tokens should depend asymmetrically on these." Normal operation of the device is, that high pitched noise is noise defined as high pitched by its significance for the device. If there

[6] See Dretske, (1986), in Stich and Warfield, (1990), for a similar argument.

had never been a misrepresentation and the notion was undefined the notion of Normal representation would not be affected but without the notion of Normal representation, misrepresentation is not defined. The logical dependence is asymmetrical and the criterion applies. Fodor appears to find use of this criterion incompatible with the teleological notion of Normality since under such theories: "... what's special about false tokens is that they can't happen when circumstances are Normal". The two theories appear to be in harmony. The falseness of a false token depends on the Normality of true tokens and this can be seen in the asymmetry of the dependence.[7]

In our first and simpler device, the integrator could be realised in various ways. In one realisation, R_1 say, the integrator could comprise some mechanical arrangement by which vector forces are added and the wheels turned until there is no angular displacement between their direction and the resultant. In another realisation, R_2, the two angles could be interpreted as a number of degrees of arc and the two amplitudes interpreted as numeric quantities. The formula, (1), for resultant angle would then be calculated and the wheels turned by that angle.

The designer can say, of R_1, that a mechanical state of the mechanical integrator is caused by a state of the sound sensor together with a state of the light sensor and that this combination causes the wheels to turn, thus. If the equipment is working properly this is true, if not it may be false. The designer can also say that this equipment meets its functional specification in this environment because that mechanical state represents the amplitude and direction polarity of the light signal together with the amplitude and direction polarity of the sound signal and that these events cause the device to move appropriately. He can also

[7] This neglects Fodor's distinction between information and meaning. The notion of 'meaning' has no role in a cybernetic theory of agency, the role of 'meaning' in a linguistic community of

say that these states represent that there is sound here and light there, of these magnitudes and that this causes the device to act by going towards light and away from sound.

In R_2, there are particular electronic states that cause action and are caused by perception. The states may be of magnetic devices and can be interpreted as numbers in a binary notation. These numbers can be said to represent under the interpretation of a theory. There could be a radio link from the device to a large time sharing computer that performs the calculation. This computer may reuse memory constantly and each instantaneous representation will be a unique usage of silicon chips under some application of virtual memory in the computer but its representation is still of the same phenomena for the device in the environment of the device.

Within either realisation, misrepresentations will propagate through the cognitive process. Normal operation is maintained by Normal transduction at sensors and Normal processing, where Normality is historically defined. In Normal processing a blue light of brightness B at an angle α (call it condition p) and a high sound of volume V at an angle ß (call it condition q) will be represented by a token representation of p and a token representation of q. This will cause a token representation in the integrator of p & q and will lead to an action which may be described by us as moving in the direction given by (1) above. This action is represented by the output from the integrator. This output will consist of two conditions of mechanical position, voltage or digital code, depending on the realisation of the integrator. One will represent the angle to turn and the other will represent whether or not to move depending on whether the acceleration sensor had detected an impediment to forward or backward movement.

social agents is considered in 4.10, below. 'Information' is a quantitative measure of the probability weighted latent variety of a representation within a theory of agency.

The content of representation within the device is nonconceptual for the device. This is expressed in physical states that represent conceptual content for us in our theory of the design of the device. We are confident of the truth of the theory because we designed the device. The conceptual content of this information is within our theory of the world for the device. It could be articulated in sentences including mention of 'wave fronts at the sensors', or 'sources at the arena boundary'. The content of action by the device is also non-conceptual for the device. It, also, is described in terms of physical states that have conceptual content for us in our theory of the design of the device. It could be of 'turning by an angle θ and moving forward' or' moving toward a source'.

The physical processes that realise cognition in the agency of this device can be described in statements of our theory of the device — where we, the theorists, conceptually express the contents of the processes. Two questions emerge: Firstly, since these physical processes can be described in terms of changing representations, are these processes apt for description as syntactic? Secondly, since at least some of the processes are formally isomorphic to a computation are they apt for description as computational.

It may be that the analytical methodology of syntactical research will be of use in explaining a cognitive process but this does not make the process syntactic. Syntax is the formal description of the conventions of a language. A language is a set of social practices within a grouping of agents and each linguistic act is an act of agency by a member agent which is given semantic form by the syntax of that language. Acts of agency derive from the cognitive processes of agency which are either those that have evolved as functional for agency in this sort of organism in this environment or those that have been designed to make this sort of device work in this environment. Syntax derives from social acts of agency that are communicative. There are no signs in the cognitive processes of our artefact since there are no agents within such processes for whom such signs would have meaning. We may describe the cognitive processes of agency in terms of codes or quantities, as I did when considering representation within the integrating mechanism. Such states are represented as codes in my theory of the cognitive processes of agency of the artefact. They are not codes for the artefact. For the agent artefact they are states that cause states that cause action.

If the cognitive processes of an agent are such that the causal processes by which representations maintain representations will Normally preserve the truth of the semantic content as interpreted in a true representational theory and the actions Normally caused by those representations are apt for interpretation as for some goals attributed to the agent then the actions of that agent can be interpreted as Normally rational in the context of the pursuit of those goals

We can get yet more philosophical mileage out of this simple specification. Suppose that we had two devices of the second kind, one in each of two arenas. We then arrange things so that each is in an identical position with identical sources of light and sound identically placed in each arena. Each is identically specified regarding weightings of brightness and volume, and colour and pitch preferences. If one is realised mechanically and the other by a time sharing computer then their representations would be very different but each would have the same nonconceptual content for the entity. Each would have the same conceptual content for the observer in terms of incident physical phenomena and each would have the same conceptual content, for the observer, in terms of location of objects in the world of each entity. For an observer, aware of the two arenas, they each represent different things with different names in different places. The inner representations of each device are functionally identical though physically very different. The wide content semantic values of each representation will depend on the particular arena inhabited by the artefact. There might have been two independent constructions with neither constructor of an artefact knowing of the other. The two constructors might have chosen different frequencies and different volumes and a different size of arena but used the same terms and names for features of the environments. There are at least three levels of indexicality; that of the artefact, that of each independent constructor and that of a common observer. If the independent constructors were in separate linguistic communities but of the same kind of realisation of agency then by their agency they would classify their worlds alike though in different but inter-translatable terms. Each of the two artefacts are in worlds of light and sound and despite the different realisations of agency in the artefact, each constructor would *mean* the same thing by their terms for angle, pitch and hue in their conceptualised theories of the content of representation. This *meaning* is not in the head of either artefact; it is within the practices of the linguistic communities of the constructors. The conceptual world of each constructor includes the features that are within the attributed content of intrinsic representations in the theory of agency of each artefact. If the two constructors were of different realisations of agency then the world of one might not include the same features in its theories of agency as the other. There may be little or no common conceptual content.

We see that analysis of the behaviour of even a simple artefact with agency can yield insights into our use of the concept of representational content. More complex artefacts with agency will deploy greater cognitive complexity and may learn.

4.4 More Complex Artefacts

Robotics has begun to play a role in our everyday lives. Complex items are shaped, cut, welded, painted and moved by machines that have been designed to react to different things in different ways without a predetermination of incident event or outcome. The agency of such devices is limited by the simplicity of the world for them that our designs prescribe[8]. One machine may recognise a new item and its boundaries and spray just the item. Another may

[8] The 'frame problem' of artificial intelligence arises from this self-evident fact.

run along a track and branch according to various criteria and stop if it meets an obstacle. Within the technology of Knowledge Engineering, we now program our computers to 'learn' from history and use guidelines to reason from this to the implications of new input. We can program computers to play chess at grandmaster level. I will consider the implications of this for the computational theory of cognition in Chapter 6. We can construct guided missiles that will home on a heat source and compensate for intentional evasion by that source. The intentionality of this evasion may be that of human agency. Within the limitations of this context, as with chess, the agency of the artefact may defeat that of the human.

Isaac Asimov[9] has written many novels, within the genre of science fiction, around a possible impact of the development of robotics on Western civilisation. In these stories, the final design of Asimov robots emulates humans sufficiently closely to pass as human in action if not in appearance. They are programmed to speak a public language and to learn. They are also programmed to a morality of deference in action for the protection of natural human kind. His novels in this genre have explored some of the problems of attribution of conscious morality and self-consciousness that the interactions of such artefacts with a natural human community would pose. Whether such robots are physically practicable is uncertain but their realisation is not negated by any of our current 'laws' of science. A new robot, straight from the production line, would manifest a rich agency with a variety of regulation equal to, if not greater than, that of an adult human but *it* would not have learned. It would have no experience of the world but its actions in the world would be based on cognitive processes, realised in an artificial medium but functionally isomorphic with human cognitive processes that have developed ontogenetically through years of interaction with the world of things and others.

[9] All of the Asimov robot stories were published under the title: "The Complete Robot", Harper Collins, 1993.

According to Asimov, such robots would then begin to learn. Their 'positronic' brains would be plastic to the retention of memory and the evolution of subsequent behaviour so that their subsequent agency would be that of the agent at the instant, continually modified by its experience.

Analysis of Asimov robotics in terms of agency and intentionality would be a thesis in itself. Meanwhile, within our current technology, increase in the complexity of our artefacts is generally limited by application needs, cost and size. All of our artefacts with agency are realised within the causal structure of our physical world. All agency of artefact, from the simple thermostat to the most complex interplanetary learning probe is linked by a continuity of technological development within the physical laws of our world. We have developed talking robots that will tell humans to get out of their way but this is a simple parroting of speech. We have developed programs to parse speech but not to use it. Neither of these developments has much to do with artificial language but this issue must wait until we consider language and agency. Meanwhile, we should return to agency in the natural world.

4.5 More Complex Organisms

As with all organisms, the regulation of plants is a combination of homeostatic and heterotelic controls. Generally, plants are not motile. However, "plants are able to carry out movements and to respond to external stimuli. They turn their leaves to the light, their roots grow downwards, blossoms can open and indeed many insect eating plants make fast grabs for their victims", von Frisch (1964, p. 239). Hart, (1990, 2) distinguishes between two distinct types of mechanisms in plants. *"Turgor movements* are due to reversible changes in the sizes of of special cells which bring about the movement of a lever arm such as a leaf blade or a reproductive structure. *Growth movements* result from particular patterns of differential growth within and between organs and, although necessarily slow, are no less dramatic or crucial to the life of the plant."

Turgor movements are brought about by osmotically driven changes in the volume of special cells, or of cells in special regions. Exogenously induced turgor movements, those relevant to our account of agency, enable plants to carry out rapid and repeatable movements and are involved in; protective roles, in processes of food collection, in both general and specialised situations and in reproduction. Growth *tropisms* involve the formation of a growth hormone, called auxin[10]. The phototropic, geotropic and haptotropic bending of stems, roots and tendrils comes about by unequal growth which is caused by unequal distribution of auxin.Which in turn, is caused by an unequal distribution of light, turgor, moisture or whatever environmental feature is appropriate for this change in the relation of this part of this plant to its environment. The major forms of movement are summarised in the Table below, from Hart, (op cit., p. 3)

Type of Stimulus	Turgor change	Growth
Endogenous	Nyctinasty (leaves) Ultradian rhythms (leaves) flower movement (sunflower)	Nutation (all organs)
Temperature		Thermonasty (petals)
Light	Photonasty (leaves) Heliotropism (leaves)	Photonasty[2] (flowers) Phototropism[2] (aerial organs)
Mechanical	Seismonasty (Mimosa leaves) Thigmonasty (floral parts) (insect traps)	Thigmonasty (most tendrils) Thigmotropism (some tendrils)
Chemical		Epinasty (petioles, shoots) Chemotropism (fungal hyphae, pollen tube?)
Injury		Traumatropism (all organs)

In tropic growth and in the reversible movements of turgor we can pick out the functional components of our regulatory system. *Perception* is realised in the way in which the incident phenomena affect the auxin production in this part of the plant and the spatial differentiation of auxin levels between cells in this part of the plant *represents* that incident phenomenon for that component of the plant. Due to the structural organisation of the leaf, the auxin differentiation stimulates appropriate differential growth and *action* is that differential growth. In the Venus fly trap, *perception* is realised in the way in which pressure on the touch sensitive hairs on the surface of the leaf, generate an action potential of about 50 mV in the base cells of the hair. This potential *represents* that pressure for the plant and *action* is the trap closure caused by the turgor loss in the motor cells along the hinge of the trap.

We must consider what we mean by the entity. A leaf turns towards the sun and this turning benefits the plant by enabling photosynthesis within the leaf. In the same plant, a root reaches downwards and this reaching benefits the plant by enabling ingestion of water and minerals and the alighting of an insect may cause a trap to close to feed the plant. These processes are causally independent. Each can be functionally described in terms of perception, representation, and action, and their realisation depends on the evolution of the plant. In each, the relation between the representation and the action enables the autopoiesis of that plant in terms of the nonconceptual content of the representation and that enabling is manifest in the structural arrangements by which representation leads to action. But it is the agency of the leaf and the root and the trap. The autopoiesis of the plant is not manifest in the regulatory behaviour of the component parts. It is manifest in the design of the plant, the causally independent regulation of leaf and root and trap and these have function for the autopoiesis of this species of plant within the functional containing system that includes the whole plant and its reproduction in its environment. The leaf satisfies our criteria for agency. It is intentional in that it has representational states, which are about a world that is for it in the context

[10] In some tropisms, at least, changes in the cell membrane and electrical activity play a part. Hart, p. 21.

of its desire for photons that realises its function of photosynthesis. Its actions depend on these intentional states in a way that is rational in the context of its photosynthesis and this rationality is manifest in the design of the leaf. It is this design that can be thought of as a representation of the goal of the leaf to obtain photons. A similar argument can be developed for the root in the context of its downward growth and for the Venus' flytrap in the context of its rational behaviour for catching insects. It would seem that our austere characterisation of agency is satisfied separately by separate components of a plant though not by the plant overall. As with the bacterium; for the leaf, for the root and for the trap the nature of the regulation is reflex.

Each individual regulatory mechanism of a plant is independent and each has the requisite variety needed to cope with the variety of its environment. For the autonomous living plant, there is no overall regulatory structure such that the regulation of the plant as a whole has the requisite variety to control against the variation of the total environment for the total plant. This is not to say that the plant is uncoordinated, the regulatory mechanisms of each component are co-ordinated by the DNA of the plant through which the evolved regulation of the species is maintained, but this is the co-ordination of a programme. Co-ordination of the organism to enable the organism to respond as a whole to the variety in its environment requires some method by which information, from the environment, obtained by perception across the surface of the organism is available for computation for action which is holistically for the organism. However, plant cells lack the capacity for specialisation that has led to the evolution of neural communication in animals. Chemical signalling in multicellular plants is limited by the presence of the rigid cell wall, which restricts access of large molecules to the plasma membrane.

Much of the homeostatic regulation in animals is mediated by chemical intracellular communication. "The physiological response a cell makes to a given signal is governed by its particular specialisation. ... The response is determined chiefly according to the functions a cell has become differentiated to carry out" Lawrence, (1989, 248). The differentiation of the neuron has permitted the development of; "[N]ervous systems which allow an animal to sense and respond to external stimuli and communicate between different parts of its body far more rapidly than would be possible if communication were restricted to chemical diffusion between and within cells", (ibid, 448). The mammalian peripheral nervous system is functionally differentiated into an autonomic nervous system, which includes the various homeostatic regulators that are under neural control and a voluntary nervous system comprising the nerves supplying skeletal muscles which are under 'conscious' control.

Ashby's principle of requisite variety establishes a necessary harmony of complexity between an agent and its world. In evolutionary terms we may see variety as one dimension of an ecological niche. One spectrum of phyla, past and present, is from simplicity to complexity. A worm that desiccates dies, if another worm can respond to a moisture gradient to seek out dampness then more of its descendants will survive. The cognitive complexity of the species will have increased to maintain a necessary harmony with the increased variety of the world for its members.

As the cognitive variety of vertebrates has increased so has the plasticity of the neural system and its importance for the ontogeny of the individual. The peripheral nervous system is separated by interneurons that intervene between sensory and motor surfaces. This severs the one-to-one relations between perception and action and vastly increases the range of states open to an organism. Although the interconnection of the interneurons remains constant their specific activity and through the blood supply, the general activity of the organism, changes their state. This results in behavioural changes, some of which may be described as learning. Through the interneurons the nervous system connects together distinct sensory modes of the organism. The organism can thus react to relations between events as well as to the events themselves.

Since the organism learns it can react to relations between events over time as well as sensory modality.

Explanation of the physical realisation of the complex agency exhibited by animals with nervous system is made difficult by the vast range of complexity. For simple organisms and artefacts we can demonstrate by example that conceptual content is for the theorist, not for the agent. Animals sufficiently like us for us, as theorists, to consider the attribution of propositional attitudes with conceptual content to their intrinsic representations within agency, have nervous systems too complex for us to begin to consider actual realisations of such representations, as we can for simple organisms and complex artefacts.

Although we are not able to build a systematic description of the cognitive processes of a complex vertebrate we have sufficient clues to consider some generalisations. Although we differ radically from other vertebrates in our use of language and the complexity of our reasoning and tool use, we are sufficiently like our anthropoid cousins and other higher vertebrates to use such terms as recognition, avoidance and pursuit reasonably in describing their and our behaviour. A baby, a lamb and a puppy will each recognise their mother. Although different sensory modalities are involved very similar sorts of neural syndromes are involved. In each case an innate agency is realised in a rapidly evolving neural structure. Each act of agency involves a neural event as a cognitive component and each such event changes the cognitive structure of the agent. This lamb or this child is a continuing entity in its world and the world of the observer but each agent at the instant may be markedly different, as neural structures develop to realise cognitive processes that give a richer agency.

The neural structures of a particular sort of frog are probably not very different between one frog and another, of the same sex. The neural structures of a given frog won't change much with time or maturity, although frogs may learn and the cognitive processes of this frog today may not be quite the same as this frog tomorrow. The cognitive processes of animals as complex as cats and apes certainly do develop through infancy: adults can and do learn. Our description of the agency of a kitten will be very different from our description of the adult cat though they will be different instances of the same cat. Each kitten from the same litter will be a distinct individual, though with very similar cognitive and sensory mechanisms. Each sensory event will cause cognitive events and these will include changed neural states that represent that sensory event for the agency of that kitten. These processes will lead to actions and the interneural states of this kitten will change in ways that are *normally* for this kitten in the context of its history of sensation and action[11]. That they are Normally for the kitten depends on the evolutionary history of the species that has led to this pattern of development. The cognitive processes within the agency of the kitten of today are realised in its neural structures as they are today. The cognitive processes of the kitten of tomorrow will be realised in its neural structures, as they will be tomorrow.

Saving accidents, the kitten of today will lead to the cat of next year but the cognitive processes of that cat will be realised in the neural structures of a mature cat and what these become will be determined by the events of the year, combined with its genetic inheritance. The cognitive processes of that cat will be no more mysterious than the cognitive processes of the artefact of agency described in the previous section. They are far more complex and are realised in a rich combination of adaptive processes. They are also essentially unpredictable because the detail of the realisation depends on a variety of contingent events that is unique for each cat. But the realisations of perception, representation, process and action will be in the physical cellular substances of

[11] I have not capitalised 'normally' here. Millikan defines a 'Normal' explanation as "an explanation of how a particular reproductively established family has historically performed a particular proper function. A cat or a person establishes its own transient norms within the continuity of its development and the notion of a reproductively established family fails. These norms may or may not be functional for survival but they are what is normal for that individual at that time.

nose, ears, eyes, neurons and muscles. Natural and intrinsic representations within the agency of the cat are caused by sensory input; chemical densities, incident pressure waves at the ears and patterns of light and shade. The actual cognitive processes by which these cause action will depend on the history of the cat.

That a cat recognises a location is determined by a history of acquaintance. Because of this history some gestalt pattern of sensory input will cause this cat to act. The cat, like us, has many reflexes that could be described as examples of simple agency. But the agency that is of interest is described by complex actions of the cat that are responses to aspects of its environment that we can describe as features and even as objects for the cat.

A cat may be in a field. This cat may never have hunted but it will have a set of inherited responses by which particular sensory combinations cause particular hunting like behaviour. If it had hunted or been shown hunting behaviour by its mother then these responses would have been modified for this cat. The neural causal sequences, by which these sensory combinations cause this behaviour, in this cat, at this time, are the cognitive processes by which this aspect of the agency of this cat, at this time, is realised. These causal processes will include neural states that we could correlate with different aspects of the sensory combination that caused that behaviour within that event. If our neurology were up to it we could say that one state represents that sort of smell, another represents that sort of sound and another that sort of pattern of light and shade. This is the complex, organic equivalent of the representations of angle, volume and pitch in our proposed artefact of agency. In a cat, this agency has not been designed. It has evolved and its Normal proper function for the autopoiesis of the cat has been honed over millions of generations of cats in environments in which this sort of behaviour sometimes leads to food or reproduction.

Although the cat may be in a field, it is not a 'field' for the cat. Though the inherited and acquired cognitive processes of the cat may enable it to distinguish being in a field from being in a house and a gap in a hedge from its cat flap. We observers may describe the behaviour of the cat so that its actions are of agency under various descriptions. We may speak of it seeing the bird and catching a bird or of muscular contractions following sensory stimulation. Each description will be apt for attribution of truth within its own domain.

All of the thought experiments that we considered for our artefact of agency can be considered for the cat. In considering these for the artefact we gained confidence in our reasoning by the knowledge that we could actually make such a device and that it would behave thus and so. For the cat, as of now we have not the technical capability to make an exact duplicate and place it in an environment identical with respect to its agency. Even if we could, there is a sense in which we would not have made a cat. Although the conditional statements within a true system description of the copy and its environment would pick out an indiscernible set of possible worlds, the goals of the agency of that cat at that instant would be our goals. The goals of Same Cat, the one we copied; ranging from appetite and comfort to doing something about the new cat next door, have an aetiology that is a logical part of having been Same Cat with its unique experience, growth, damage and inheritance and they are its goals. Next day and next year, Same Cat and Copy Cat will be distinct qualitatively as well as quantitatively. To the extent that its current goals have been modified by its history, the agency of Copy Cat is realised in a different mechanism from that of Same Cat. They are different entities in different actual worlds; even if their environments were indistinguishable, the system that describes them would pick out different possible future worlds.

We must distinguish between two different distinctions. On the one hand, we may distinguish between entities of different kinds but each apt for system description in terms of agency or apt for system description in terms of the complex agency of a cat. On the other hand we may distinguish between stages in the becoming of an entity apt for system description in terms of learning or maturing agency. In our theories of the agency of such an entity we are describing the same thing over different realisations of agency. This distinction is more marked in the metamorphosis of a caterpillar into a butterfly. The realisation of agency in a caterpillar is profoundly different from that of a butterfly. They might be joined in a system description of the processes of such metamorphosis but within such a system, each creature of the instant can be described in system terms that pick out the possible future worlds that include this entity. As long as it lives, a constant capacity of the entity described is its agency: in general, the detailed functional specification of this agency and the detail of its realisation will change.

The dynamics of agency due to learning and adaptation has profound implications for our attributions of personality and we will consider some of these in the next section. Chapter 6 is devoted to the metaphysical implications of this dynamics for our attributions of freedom and responsibility to one another.

I opened this chapter with an example of my agency as manifest in my action of getting petrol for my car. This was an action of agency, intentional under a description. It is an example of a rich agency involving a complex interdependence between my perceptions of the petrol gauge and the filling station, my experience, the design of the car and all of the interactions between my agency and that of others in my community. The goal of agency maintained by that action is having enough petrol in my car. As a child I knew nothing about cars or petrol and my cognitive processes had no traffic with representations of running out of petrol. My current agency includes a realisation of such representations and my goals can be said to include not running out of petrol. The neurology of this is very complex and it is dubious whether we will ever be able to identify realisations of such representations but this does not make them less real. The distinction between representation and process may blur but this does not make it less real. My current agency is integrated, complex and dynamic but it is still apt for description in terms of perception, representation, cognitive process and action and each of these is intentional.

4.6 Learning and Adaptation

I have argued that a necessary accompaniment of life in our world is agency. Life itself requires autopoiesis but the counter-entropic realisation of autopoiesis against the variation of an entropic world requires that the living thing act in the world in response to events in the world and by that action classify the world for it. These, I have argued, are the criteria for agency. Within our taxonomy of zoology we divide the sorts of living things into species; sorts of living things that maintain a continuity of classification by their interbreeding for reproduction and their similarity. The realisation of agency differs, often dramatically, between species. It also differs between current members of a species and between current members and their ancestors.

That species differ in the way their members realise their agency is self-evident. Most birds see prey and predators and act by flight, grasping and tearing with the beak. Moles smell prey and predators and act by tunnelling and biting. Differences between members of a current species will be less dramatic. In simple organisms the differences may be negligible but in more complex organisms, particularly in those with complex nervous systems, these differences may be marked and will be due to the different genetic inheritance and the different developmental history of each.

Local differences between members of a species due to genetic inheritance will be random and distributed around some norm if the species is stable. If the dynamics of evolution are relevant then these differences may be significant and we would speak of the trend in differentiation as adaptation. Successive generations of this species in this environment will differ from one another, however marginally, in some components of their realisation of agency. There may be small differences in some sensory acuity such that intrinsic representations within agency represent some different aspect of the world. There may be small differences in cognitive processing such that action results from some new combination of aspects of the world. There may be some change in resultant action such that the world is changed differently. Over enough time, there will be some combination of all three. These changes will change the environment and this will also influence the genetic trend.

One class of changes of this kind within all animals except sponges has been the development of the nervous system. Abercrombie, et al[12], define the nervous system as: "A mechanism which co-ordinates the various activities of an animal with each other and with events in the external world by means of messages rapidly conducted from part to part. ... It consists of numerous nerve cells which have branching thread like processes. .... The nerve cells and their processes, linked by synapses, form a system which permeates the whole body." We can distinguish between; the peripheral nervous system which consists mainly of nerve cell processes running directly to sense organs or effectors and mediates in reflex processes which depend on inborn nervous pathways, and the central nervous system which co-ordinates the activities of an animal.

Reflex processes can be described in terms of a simple agency, functionally similar to the turgor of parts of plants. Although such processes have function for the autopoiesis of the animal, their agency is autonomous and they do not directly contribute to the rich agency of the animal enabled by the central

[12] From "The Penguin Dictionary of Biology"; M Abercrombie, C J Hickman & M L Johnson; 1980.

nervous system. All animals of the same species will differ to some extent, if only in physical dimensions but our interest is in how each individual of a species may differently realise its agency within its environment. Each act of agency is a unique event that can be described as intentional under some description. For a given species of animal, within a reasonably stationary environment, we can define a class of sensory events to which the reactions of animals of this kind define a class of action. In the terminology of agency; an event within this class will Normally cause cognitive processes that will Normally cause actions within that class. We may then define:

> Animals within a species are of *similar* agency in so far as they Normally produce similar actions from *similar* events,[13]

By defining similarity in this way we may bring out the way in which variety and learning produce 'dissimilarity'.

Simple animals within a species are very similar. Two Normal frogs or two Normal bats will act in much the same way within the same environment. Frogs may learn a little and bats may learn a little but their agency, the way they react for them, to events in their world, does not change much. Although the world of each token animal may be quite various in terms of space and prey and this variety is matched in the complexity of its agency, relatively simple organisms such as frogs or bats do not adapt to fresh variety in their world and each remains much the same as its conspecifics.

Colonies of social insects can and do, vary in character but each individual insect within a caste within the colony is very similar to its neighbour.

[13] The notion of 'similarity here is close to the Wittgenstein notion of 'family resemblance'.

Behaviour of the colony is co-ordinated by complex pheromones[14], stemming mostly from the queen, though produced by each insect.

Observation of complex vertebrates shows a marked diversity between individuals. This can be seen in studies of social animals such as chimpanzees where researchers name individuals and comment on their distinctive characteristics such as ingenuity and dominance. Studies of life histories of identical twins in Western civilisation today have concluded that about half of our dissimilarity is due to our genetic inheritance.

Whether the variety is due to genetic diversity or learning or environmental pruning, there is a clear correlation between individual variety and cognitive complexity. This has reached its peak in humankind but we can see and respond to, idiosyncratic behaviour in chimps and cats and dogs. Thus, any theory of cognitive processing that we deploy in our account of the agency of creatures that can learn and display a cognitive idiosyncrasy between individuals will not apply in detail between one individual and another. In describing the agency of our artefacts, however complex, we can describe how representation and cognitive processing is realised in this sort of agent. Each act of perception is unique and is epistemically contingent for that agent but the action of that agent to that perception is predictable, at least in principle, for the theorist. Similarly for simple organisms, the cognitive processes of the bat, as described by Akins, (1993), show a complex co-ordination of reflex processes that catch prey and avoid obstacles but leave no room for idiosyncrasy.

Animals as complex as the higher vertebrates are born with more variability and this variability can be enhanced by learning within a lifetime of experience in a changing environment. The behaviour of each animal can be described in terms

[14] These are chemical substances which, when released into its surroundings by an animal influence the behaviour or development of other individuals of the same species.

of its agency. Within the terms of this description, the actual sensory organs are the mechanisms that realise perception for this agent. The actual interneurons and hormonal secretions (in so far as they have a role) are the mechanisms that realise representation and cognitive process for this agent and the actual motor nerves, muscles and claws are the mechanisms that realise action for this agent. Another member of the same species or this member at another time will be similarly described, but the realisations of cognition will be different since significant parameters of the system differ. It is, however slightly, a different realisation of agency in this kind of entity due to its different experience and inheritance, and the agency of this species is sufficiently complex for this to matter.

Agency involves an agent with the world. Cognitive complexity is relevant in a complex world. This is the moral of Ashby's theorem of requisite variety. Organism with complex cognitive processes and the ability to learn have evolved through adaptation to greater environmental complexity and unpredictability. The agency of such organisms is idiosyncratic in that, within a species in an environment, each agent at the instant is dissimilar to its previous self and its neighbours in regard to the cognitive processes that lead to its actions of agency. This is one way in which my austere characterisation of agency must be enriched to describe our agency. Another way is by our social behaviour and this is the subject of the next section.

4.7 Social Agency

The agency of an organism will include the role of other organisms in its world for it. These roles are ascribed in our description of its agency and terms such as prey, predator, mate and symbiote are within our theories of ethology. Each agent will act in response to its perceptions according to its current cognitive processes. By these descriptions we classify interactions between organism as within the symbiotic, parasitic and predatory relations of an ecology or within co-operative and competitive behaviour of sexual reproduction or within the co-operative and competitive behaviour of a social group. The agent interactions of an ecology describe the evolutionary niche inhabited by each member agent and each agent will manifest a regulatory variety appropriate to the variety of that niche. All animals reproduce by some sexual behaviour and goals for that behaviour will be implicit in the cognitive processes of each agent. This section is focussed on the interactions between membership of a social group and the agency of its members. To this end we should consider what it is to be a member of a social group.

A co-operative group of agents of the same kind is defined by its co-operative practices. If the perceptions of each member of a group of agents include recognition of other members and lead to action that is for the group then these actions will become practices that characterise this sort of group of this sort of agent. Groups of this sort will persist if the practices favour the survival of the members and some copying mechanism exists by which these practices can persist between successive generations. Alternative realisations of such groups will persist if they are successful within the normal dynamics of evolution.[15]

In the groupings of cells that comprise an organ of a complex organism, the DNA of the organism prescribes membership of the group. Each cell still has agency but the goals of its agency are no longer totally determined by the autopoiesis of the cell. The cell perceives and it acts but some of what it perceives can be described as commands; to contract or to relax, to fire at a synapse, to die or to reproduce. Each organ can be described by its function for the organism and by the way the practices of its member cells enable that

[15] This is not Social Darwinism, at least, in its current pejorative usage. My argument is that types of human groupings have evolved by the accidents of success just as types of other anthropoid groupings and insect groupings have evolved. That they exist is evidence for their survival. If a kind of social group has survived in virtue of some of its practices then, to that extent, the aetiology of those practices is explained.

function. An organism, functionally described by its organs and their functions is apt for internal description in terms of autopoiesis and for external description in terms of agency. This is an extreme manifestation of social agency in which agent autonomy is totally submerged in a larger agency.

The practices of species of social organisms without productive language range from the instinctive and total obedience of the social insects to the loosely structured consensual behaviour of a tribe of chimpanzee. The agency of an organism may itself be subservient to its agent membership of a social group. Some of the group practices may be inherited and some acquired by learning. Some of these practices may involve goals of agency that compete with the goals of individual survival or reproduction.

Groupings of social insects such as ants are maintained by inherited practices which include mechanisms for the selection of a queen and her maintenance, and dietary practices that differentiate castes. The genes by which such practices are copied through the generations do not determine the caste of any female but they do determine the allometry of a colony and thus the characteristics of the caste system as a whole.[16] Within a caste, member ants are very similar and one theory of agency will apply to each ant. Genetic differences between colonies will lead to substantial variations in differentiation between castes and the content of pheromonal communication and thus to variation between colonies.

In more neurally complex animals social groupings are more diverse. Lions live in prides whose enduring centres are closely bonded females and their young. Adult males live in separate groups, often as pairs of brothers. Adult males and females hunt together, with the females taking the lead role. Some of these

[16] Wilson, (1992, 78-80).

practices will be maintained genetically. Other practices will be maintained by copying and repetition of individual and co-operative movements. Although environmental pressures and the need for intra-species recognition will reduce variation between lions in a pride, enough variation will occur for dominance hierarchies to develop between members.

Many species of primate live in social communities bonded by complex practices of co-operation that maintain the group and its members but permit marked variation between members. Each agent member of a group is a unique individual marked by its own life experiences. They recognise one another and make use of the differences between others in furthering their ends within the community. They will deceive one another and grieve for the death of another, (Seyfarth and Cheney, !992). Not only is each agent member marked by its own life experiences but also these experiences have been within the group and the goals of each will be at least partially the goals of the group. The agency of each monkey or ape is approaching the rich characterisation of human agency in that, although each free act[17] is an act of agency, the perceptions, representations and cognitive processing that cause that act are distributed over a lifetime of experience.

4.8 The Agency of Social Groups

Each agent organism is a co-operative group of agent cells whose agency is subservient to the practices of that group. We can consider whether a social group of organisms such as a colony of ants or a pride of lions or a group of monkeys is apt for description in terms of its agency. In 4.2 above I defined agency in the terms: 'An entity is the agent of an act if what it does can be described under an aspect that makes it intentional for that entity.'

[17] What it is for an act to be 'free' is discussed in depth in 6.8 below.

Ants within a colony communicate by pheromones. An ant meeting prey, problem or predator will send out signals that cause other ants of appropriate castes to act in ways appropriate for the benefit of the colony. This signalling is an act of that ant but it is not an act of the autonomous agency of that ant since it was not *for* the ant. The act was *for* the colony, just as the synaptic impulse of a rod cell in an eye is not an act of agency for that cell but is an action of the eye which is for the organism. Actions of an ant can be distinguished between those that are in response to its own local perceptions and those that are in response to the stimulus of pheromones. A soldier ant will respond to distal stimulation by these pheromones, to come to the site of an attack and is programmed by its instincts to attack anything that is living and not another colony member. The actions of the soldier ant are actions of the colony and these actions are caused by representations within the colony, realised in levels of specific pheromones and the states of the ants releasing them. The goals of the colony are manifest within the practices of the colony and are realised in programming by the DNA of its members. Due to this programming, perceptions of individual ants cause the release of specific pheromones that are perceived by other ants and cause action by other ants that is for the goals of the colony. States of the world for the colony are perceived by member ant and represented by proximal levels of pheromones. These representations cause action by member ants, and this action is for the goals of the colony.[18] Our functional characterisation is complete. The acts of a colony of ants can be described under an aspect that makes them intentional for that colony. The agency of individual ants is not entirely subservient to the agency of the colony since each ant will act reflexively to attack when proximally stimulated but we can clearly distinguish between acts of the colony and acts of an ant. We may conclude that

[18] This in contradistinction to the independent agency of the organs of a plant. The agency of the organ is for the good of the plant but its actions are for the goals of the organ. The transmission of perceived information by pheromones is analogous to the transmission of perceived information by neurons.

> An ant colony is apt for description in terms of agency.

"Vervet monkeys frequently grunt to each other during normal social interactions. ... Vervet monkey grunts are strikingly different from their alarm calls: alarm calls given in response to different predators are easily distinguished acoustically.", Seyfarth and Cheney, (op cit., 339). The agency of each member of the monkey group is partially subordinate to its membership of the group but the rich individual agency of individual monkeys is largely autonomous. Alarm calls given deceptively by individual monkeys will lead to a selective ignoring of that monkey's calls by other members. Giving and responding to alarm calls could be described as perception by the group and action for the group. Selective discrimination could be described as learning by the group that modifies the cognitive processes of the group. Current neural states of members that represent for the member agents could be described as representations for the group. The problem with this model of agency is that no coherent theory of agency of the group emerges; the theory shifts with each event. Some of the actions of members are concerned with maintenance of the group but generally actions of members are for themselves or one another. We cannot clearly distinguish between acts of the group and acts of monkeys. We may conclude that:

> A tribe of monkeys is not apt for description in terms of agency.

Human groups may be similarly described. Each human agent lives in several groups and generally, no single group can be described as autonomous in maintaining itself through the perceptions, representations and actions of its members.[19] This distinction is vague. We may clearly distinguish at extremes such as a hive of bees or an organism of cells and a total plant. The distinguishing feature will be the existence of some co-ordinating mechanism, some medium of communication and institutional practices that realise the integration of agency across members of a group so that actions of members of the group are caused by representations within the group and are for the goals of the group.

However, humans do construct artefacts of agency that are made up of human agents. Such artefacts include the entities that we call corporations or states or universities. Such organisations are defined by their goals, which may be set out in a mission statement or charter and a constitution or system of laws that set out the practices of group membership. The constitution will include procedures to ensure the continuation of the organisation by appropriate response to external and internal events. It will also include functional posts to be filled by human agents with specific terms of reference and provision for the monitoring of their actions in these roles and their replacement on retirement or death or failing to meet the terms of reference. The construction of such an artefact will be an act of human agency but the actions of the human agents within their prescribed roles may be described as acts of the artefact. Letters of instruction to human agents may be signed "Clerk to the Court" or "Minister of State for Defence". It is difficult to deny agency to such artefacts. Such acts of human agents can be described under an aspect that makes them intentional for the organisation. Just as we can see the theory of agency of an inorganic artefact of regulation through its functional specification so we can see the theory of agency of an organisation through its functional specification. The inorganic agency of an artefact is realised in hardware. The organisational agency is

[19] Searle, (1995, 25), discusses 'collective intentionality' which he defines in terms of collective forms of the propositional attitudes. This is an attributed intentionality, distinct from the intentionality of the representations of the agency of a social group.

realised in people[20] and the physical marks of rules and constitutions. In each, the processes of agency are actual, physical causal processes between states of physical, fallible components. The goals of the inorganic artefact are realised in the physical location of components or the marks of programs in digital equipment. The goals of the organic artefact are realised in the marks of constitutional procedures in text on physical media or within the memories of the community.

> I conclude that the natural groupings of human agents are not apt for description in terms of agency but artefacts of human organisation can be if appropriately designed and realised.

4.9 An Example of Social Agency

In chapter 3 intentionality was defined in the terms:

> An entity is the agent of an act if what it does can be described under an aspect that makes it intentional for that entity.

If the entity under consideration is a grouping of human agents then its intentionality will be manifest in acts of agents that are apt for description under an aspect that makes them intentional for the group. As with a colony of ants, the actions of human agents are to be explained as actions of a group when these actions are programmed by the rules of the group. in virtue of which the causal consequences and antecedents of these actions are best explained as behaviour of the group. For such programming to have effect, communication will exist between the human components of the group and this communication

[20] The distinction between person and agent is spelled out in 7.7 below. By a person here I mean the ongoing agent at the instant whose personhood is attributed by the organisation. It is the organisation who will appoint people to posts.

can be interpreted as the cognitive processing of intrinsic representations within a theory of agency of the group.

Consider a university engaged in admission procedures for the coming academic year. In what follows we will assume that the entity described is the university and that at least some acts of the university are apt for explanation as acts of agency. Such acts will be physical events involving components of the university. Such events will be physically caused by representations of states of the environment of the university, within procedures of the university that are such as to produce action that is for goals of the university in the context of these acts. These representations will have been caused by events in the environment of the university that are epistemically contingent for the university. One example of such acts is a series of letters sent to applicants advising them whether or not they are offered a place. Under an aspect of this description of the university, each posting of a letter is an act of the university. Under other aspects, it is also an act of the person who posts the letter, of the official who signed the letter and of the committee who agreed the selection or not, of this candidate. Any of these may or may not be agents under some description, though we will be focussing our attention on the university

A rough description of these procedures is that letters of application are received which contain details of qualification. Although the admission procedures have been established in the expectation of receipt of such letters, each letter is an epistemically contingent event for the university. Each letter is an intrinsic representation, within the agency of the university, of a candidate and his qualification. A clerk may produce schedules of numbers of applications for subjects within band of qualifications and such schedules will be representations. Within the procedures, numbers of applicants and their grades will be compared with current criteria for numbers and quality in each subject and cut off levels will be established according to the criteria and the number of applicants. These levels will be applied to the lists of applicants,

applicants will be selected and the appropriate letters sent.

In this very simplified model of the process the goal of agency is to admit qualified students. This goal is not represented within these processes of comparison[21]: it is a possible world picked out by the system description of the university and this system has been designed so that Normally, these procedures lead to the admission of qualified students. The staffs carrying out these functions are human agents within a linguistic community. For each human interchange of information, the representations are explicit, recognised as representational by the people concerned. Within the procedures of the university these people are following rules. With respect to the functions of agency of the university as realised, these acts by these people are component processes within the cognitive processes of agency of the university

A planner who initiated the procedures and criteria may have designed this minor clerical mechanism. Representations, such as the letter format and criteria, would have been explicit for him and for policy makers who agreed the criteria. However, this does not make them explicit within the entity comprised of people following rules and it is the behaviour of this entity that we are explaining in terms of agency. When human agents use these representations in a programmed manner they are not representations for these agents; they are tacit representations of the agency of the university.

Within a larger system, the total number of places available is given to the university by an external agency and this event sets in train complex procedures of allocation of numbers between departments. The criteria followed in selecting applicants may depend on the number of applications received for each department and the form of this dependence may depend on the total

[21] It may be that this goal is explicitly represented in a mission statement but this is not necessary for its role in the agency of the university.

number of applications to the university. Criteria for allocation of places to departments may depend on other contingent situations such as the relative quality of departments, current staffing levels and cross departmental teaching loads. If the events of setting admission criteria for each department are apt for explanation as acts of agency of an entity then this is in virtue of mechanisms within which, external contingent events are perceived and represented and these representations cause action that is for the attributed goals of the entity.

One goal that we can attribute to the university is that of continuing to exist. Another is of at least maintaining its size. It may be assumed that; if as a contingent input, an intake of N student is permitted then procedures will be established to allocate all of these between departments and this may include procedures to ensure that any unused allocation is redistributed. Representation of the total permitted intake will be an explicit representation in a formal communication between senior members of the university and the description of procedures to be followed will be explicit representations in communications between people. These communications may explicitly refer to goals of the university.

In our description of the complex mechanisms that realise the agency of the entity described, representations of fractions allocated to departments and the representations of procedures to be followed are parameters in the specification of a mechanism. If contingent inputs such as numbers of applicants and their quality are represented in text and this is read by staff independently of their programmed procedural role *then* these representations are explicit for them. If these same people are reading this data and acting in accordance with the goal seeking procedures established then the representations are of the agency of the entity within which these people are components and the representations are tacit in the sense of Dennett, (1982/3).

The term, 'following a rule' can be used in two distinct contexts. Firstly, if one is carrying out a long multiplication and one has had the appropriate tuition then, normally, one will set out the sum and proceed according to the rules learned. Similarly, if we program a computer to do such sums then we are setting out rules for the computer to follow. Our decision to do the sum or program the computer is not necessarily an act of our agency, we may be acting within some larger program, but it is, structurally, within our autonomy. Representations within such activities are then explicit: signs can be interpreted as numbers and numbers interpreted as representations in other acts of our agency

If alternatively, one is engaged in some activity in which the following of rules subsumes our autonomy; as when following established procedures within a university, then we do not interpret the representations that cause us to act. The rules that we follow may be complex and involve extensive computation but these acts of computation will be applied within the procedures, they are not interpreted by the staff whose acts are those of components, albeit complex and unique, of the mechanisms of the university. This is not to deny a quasi-autonomous agency to some human agents with Normal function within these procedures. The procedures may include a person or a group of people deciding on a procedural change or on some action outside normal procedures. Such people will be using explicit representation and carrying out autonomous act of computation. Their acts will be of their agency, intentional under an aspect of this description that makes their goals, hopefully, congruent with the goals of the university. Resultant acts, such as letters to candidates and response to the Funding Council will be acts of the agency of the university. Only in so far as such acts are for autonomous goals of the human agents concerned should they be construed as acts of human agency.

4.10 Personhood

Perceptions of complex agents in a social group will include recognitions of other members of the group. Most basic of these will be the mutual recognition of mother and child. Adult recognitions will include recognition of other adults as dominant or weak, as friend or adversary. These recognitions are lasting, Seyfarth and Cheney describe a young male chimpanzee dying from grief for his mother's death. It is apparent that other advanced social species and us have, within ourselves or our social practices, a theory of mind. By a 'theory of mind' I do not imply a complex attribution of beliefs and desires or the recursions of emotional involvement but that intrinsic representations within agency include representation of another monkey or human as 'that monkey' or 'that human'. If this is so and the evidence is strong, then each social agent is aware of other agents as ongoing identities. The practices of the group will include treating others, not just as agents of the instant but as ongoing agents of whom one has a continuity of experience, that is, as people. Whether or not one expects others to treat one as a person, this continuity of recognition implies the attribution of personhood to others. In the higher apes this appears to be one sided. We with language, Normally attribute personhood to others, and ourselves expect others to attribute personhood to us and enjoy a self-conscious appreciation of our own personhood. [22]

4.11 The Impact of Language

How language evolved is a matter for speculation. Parts of the human brain are specialised for the production and interpretation of speech and how this has evolved is a matter for debate[23]. My aim in this section will be to explore the effect of language on the manifestation of human agency within social groups.

[22] The psychological syndrome known as 'autism' may be due to a defect in this process of recognition. See Carruthers and Smith, 1996, for an introduction to this theory.

[23] See Beaken, (1996), for a plausible account of the social evolution of the practices of language.

I will take as the basic entity to be described; a linguistic community. I will assume for simplicity that; a linguistic community is comprised of normal humans, that all adult members have the same intension for each extension denoted by terms of the language. Also, that their world is stable so that the set of extensions denoted by the language is stable. Such a language would be static and could be described in terms of Fregean sense, each adult member would have the same lexicon of terms and each would have the same intension for the extension of that term. Such a language is ideal, all natural languages are in flux and at least some human belongs to more than one linguistic community, but my aim is to construct a simple and practical theory around this ideal.

Millikan, (1984), has demonstrated that a language can be described in functional terms. If, despite her reservations regarding the use of Cummins functions[24], we take a linguistic community as a functional containing system then, within this system the practices of language can be explained as having function for that social group. The group is comprised of agents and membership of that group has function for each member described in another, equally true, containing system. Members of an ideal linguistic community share a common agency. They do not, in general, share a common realisation of agency, but their agencies are functionally isomorphic with respect to those of their perceptions that are shared in the language. Quine makes this point in "Word and Object", Quine, (1960, p 8). For a term to have the same intension for two speakers its place in the life of each speaker must be sufficiently similar for communication to be about the same state of affairs in the world for each. Each, by their agency, classifies their world for them. For common signs to commonly denote, the classifications must be sufficiently similar.

[24] In Millikan, (1993, 20), she argues that the notion of function that relies on explanation within a containing system does not recognise " that basic sense of function that hooks function to purpose. This functional realism leads to a teleological account of evolution but a functional account of language does not depend on this.

As the cognitive processes of advanced vertebrates include recognition of things in their world so the cognitive processes of linguistic humans include recognition of the symbol and the thing in the world which the symbol denotes. Man is a social animal, living in groups bonded by shared practices. As with other social species these practices are concerned with co-operation in such tasks as hunting, gathering, carrying, defence and caring for young. It is likely that the practices of language have evolved around practices of communication to improve the efficiency of this co-operation, Beaken, (op. cit.).

Within the processes of cognition, some complex of perception and representation causes each act of agency.[25] This causal process can be summarised in a simple causal statement.

$$\text{x perc that p} \rightarrow \text{x rep that p} \rightarrow \text{x rep that q} \rightarrow \text{x does a} \qquad (1)$$

Where 'perc' abbreviates 'perceives' and 'rep' abbreviates 'represents' and '→_' abbreviates 'physically causes'.

An equally simple statement for an elementary linguistic transaction would be:

$$\text{y perc p \& y hears (x say P)} \rightarrow \text{y rep that p} \rightarrow \text{y rep that q} \rightarrow \text{y does a} \qquad (2)$$

Where ‘p’ and 'q' are some shared perceptions or 'categorical representations' of the world and ‘a’ is some shared practice appropriate to dealing with p and q. I will routinely use upper case to denote symbols and lower case to denote things

[25] It is by such complexes that the acts of agency categorise the world of the agent for the agent. Harnad, (1996), defines "categorical representations "that..."preserve and encode only the invariant sensory properties shared by all the members of a concrete perceptual category".

in the world, 'a' to denote action and 'p' or 'q' to denote a state of affairs or event.[26]

Perception of the world may be of lack as well as of incidence. This is not 'desire' since at this elementary stage the community may not refer to the propositional attitudes. A member asking for food is not referring to his desire he is producing the symbol that means 'give me food' and recognition of that symbol is within the practices of this community. This is in accordance with Beaken's theory that language originated in the organisation of labour.

The actions of each member of this community are still the actions of agency. Perceptions in a linguistic community include recognition of others as individuals; dominant or subservient, friendly or hostile and recognition of symbols made by these individuals. Actions of agency include producing symbols. Just as each action of a solitary agent is caused by intrinsic representations that have been caused by perceptions of the world so, each action of a linguistic agent is caused by current and past intrinsic representations within agency. Some of these will have been caused by perceptions of the world and some will have been caused by recognition of symbols that, within the conventional practices of the language, denote things in the world.

Actions are in the world and cause change in the world. Whether an action is a physical manipulation or the production of a physical symbol it is an act of agency caused by intrinsic representations that have been caused and maintained by the perceptions and cognitive processes of the agent.

[26] The mechanisms of these extremely complex processes span several disciplines. Despite their complexity, they are physical processes realised within the perception, cognition and action of the human agent.

At some stage in its evolution language will have become self-conscious. Members will refer to things that have been said. This is the stage of semantic ascent, (Quine, op cit., 271-276). If we, self-conscious observers sharing the same categories of representation within our agency, could observe a human community with a 'flat' language; that is, a community with linguistic practices that refer only to things in the world, we might say "By that symbol he means that" or "Having seen that event or heard that symbol he believes that ...". We might also say of an utterance by a member "That is true" or "That is false" but we would not hear members speak of meaning[27] or belief or truth. We, the observers, may say that the ontology of this language is only the ontology of things.

Suppose now that, within this community a new practice evolves. When an agent hears P and sees that not p, two sorts of response would be useful and thus likely to survive as practices. One response would be to say, perhaps expressively:

$$P * p \qquad (3)$$

The other, perhaps equally expressive, would be

Not P[28] (4)

It may be that some expectation of p in our, the observer's, explanation would lead us to make this distinction. For the agent these are two states of affairs not to be simultaneously accepted. Psychology recognises the phenomenon of

[27] 'Meaning' here is entirely in use. I define a linguistic community by the practices described by statements: (4), (5) and (6). 'Meaning' is the name given to the relation between the two recognitions, correlated by the practices of language.

[28] This usage of 'not' would be distinct from the negation of the statement in (1). This negation would have been, x sees that not p → x says Q, where Q is the symbol for the absence of p.

cognitive dissonance[29] as an unpleasant feeling due to opposite perceptions of the same event. Also, there would be evolutionary advantages in rejecting anomaly. Whatever the mechanisms, symbols have evolved to represent the concepts of meaning and falsity and the practices of reference to text have developed. Two principal practices may be summarised in causal form. Firstly:

y perc p & y hear (x say P) → y say and indicate 'P * p'. (5)

Where the symbol * is a term in the language that denotes a relation of meaning between symbol and world. We observers may comment that * means 'means' but for the community this is an additional practice. The classifications of their ontology now include things that are items of text. They are explicitly aware of the relations between their usage of text and things in the world. The second practice would be:

y perc p & y hear (x say P) & y rep(P * p) → y say P # (6)

Where '#' denotes that P is being properly used in virtue of its Normal usage and the perceived state of the world. For we observers, the '#' is a term in the language that denotes a relation of truth between usage of symbol and perception of the world.

If meaning has become tacit within the community then statement (6) can be simplified and translated:

y perc p → y say 'P is true' (7)

[29] See Festinger, 1956,

This describes the practice of warranted assertion but the right hand side is not logically entailed by the left. We are describing a practice, not making metaphysical assertions. Normally, these practices will involve trust in other's statements so that:

$$\text{y hear (x say P is true)} \rightarrow \text{y rep p} \qquad (8)$$

And truth is seen, like trustworthiness, as a property of P. The trusting members of the community may now say 'p if P is true' and we have one half of the disquotational schema explicitly recognised within the community. However, trust may fail. Individual members of the community may prefer to rely on the warranted assertion of statement (7). For them, P is true if p is perceived: they will say: 'P is true if p'. This is the other half of the disquotational schema. Deflationists contend that 'P is T iff p' is completely descriptive of truth. To say 'p if P is true' is to take an trusting role in the linguistic community. To say 'P is true if p' is to take a challenging role.[30]

Meaning and truth are concepts of relations between agent, language and the world.[31] In a flat language without explicit semantic terms, an agent will say “P” because it has representations appropriate to perception of p or a need for p. If one wishes to ask another about p, he might say, questioningly; "P?" The other might reply affirmatively and the first agent now has a revised representation of states in the world. Also, the first agent will now have a revised representation of the representational states of the second. The practice of naming that state of another which is concerned with its representations of the world would lead to an attribution of a property to that other and this is the

[30] This distinction touches on wider issues regarding the realist anti-realist debate that are beyond the scope of this thesis

[31] Whether semantic practices in humans evolved together with flat speech is a matter for research. Any functioning language requires a grammar and a syntax; practices that enable the productivity and comprehensibility of complex communication. But neither grammar nor syntax need be self conscious in the sense that they are topics in the language.

property that we call 'belief'. The articulate attribution of belief requires the ontology of text. The attribution of belief to an agent is a separate practice from the attribution of truth or meaning to an item of text. Their relative priorities are arguable but each requires semantic ascent and the practice of reference to text.

Human language is about the world of the human agent. With the cooperative practices of language man can study the phenomena of his world. Using the extended ontology of a metalanguage, man can study the phenomena of language. Within the functional theory of systems and biology, man can study his own agency. Human agency can be explained in terms of the world and language can be explained in terms of social agency in the world, but language cannot be explained as an aspect of the world or in terms of human agency, the three are explanatorily inseparable.

4.12 A Theory of Mind

In the example above, the interrogated agent may have perceived p without articulating P. Does this justify the statement that it believed that p in virtue of its representation of p? Those neural processes that we functionally describe as: *its intrinsic representations of agency, leading to action appropriate to p*, may be quite independent of language. If, by some chance, it were denied language but had the relevant experience of p, should we then say that it believed p but without the concept P? If we take this view then it is difficult to deny belief to any agent of sufficient cognitive complexity to act according to the content that an observer would attribute. We would say that my artefact, of Section 3 above, believes that a high pitched sound is incident at a given angle. 'Belief' is now a term that we may apply to any representational state to which we can ascribe content in propositional terms. We may be aware of misrepresentation, as in our perception of the Muller-Lyer illusion but the representation that leads to action would be our real belief.

Alternatively, we may hold that the state of belief is only possible for an agent capable of articulating that belief. This is the stance taken by Davidson, (1982), By taking this stance we deny the possession of a belief state to any agent currently lacking the term for a concept but otherwise fulfilling all of the possession conditions for that concept.

As a third alternative, we may deny representational status to belief and locate the attribution of belief within the practices of language and this is the alternative for which I shall argue. Whether we attribute *belief* to other creatures is then a matter of choice, based on whether our theory of their agency implies that such attribution will explain their behaviour. Whether we attribute the *attribution of belief* to other creatures is another issue and will also depend on whether we consider that their social life is mediated by a suitable theory of behaviour. That other social species use a theory of mind appears unlikely. Seyfarth and Cheney, (1992, 342), suggest that chimpanzees behave so as to alter or control other individuals' states of mind, but that "Even apes, however, seem to have difficulty attributing specific mental states to others." In our human communities children seem to acquire a theory of mind from about four years and can articulate an attribution of belief by about four to five years, Perner, (1993, p. 184).

Languages that survive will be those that work for their speakers. Linguistic communities that survive will be those of which membership favours survival of the members and whose practices favour survival of the group. We would expect that the semantics of successful languages would favour the preservation of truth. We would also expect that practices of attribution of propositional attitudes would favour those attributions that correctly link the attributed attitude to the Normal representations of a Normal agent of those states of the world to which the semantics of the language refer.

4.13 Hypothesis for the Status of Beliefs

I suggest that belief is not a physical state of an agent with direct causal powers. The attribution of belief is associated with representational states of agency that have causal powers but this association is through the practices of language. If an agent is so constituted that she will perform action 'a' because the intrinsic representations within her agency have content 'p' and she is a Normal speaker, possessing the appropriate concepts, then her honest utterance will be P. Others in the community who share possession of her relevant concepts and have a theory of her agency may attribute to her the belief that p. Due to cognitive dysfunction and consequent misrepresentation, her action may not be 'a' but the attribution may still be justified. We can formulate attribution conditions for belief in the terms:

> The practices of belief attribution within a human linguistic community are such that a speaker who meets the possession conditions for the relevant concepts[32] may honestly attribute 'belief that p' to another speaker if and only if he has evidence that that speaker, if cognitively Normal and in possession of the relevant concepts, wishing to tell sincerely and able to speak, would honestly[33] affirm P.

This is a description of a practice in a language. If the attributer has no evidence to expect the other to honestly affirm P then he has no grounds for attribution. If he has such evidence then he is justified in the attribution. If the attributer has evidence that the believer is deluded then this may affect the grounds for attributing belief. If the speaker is unaware of the delusion but has the

[32] This is the Peacocke, (1992), notion of 'possession conditions'.

[33] By 'honestly' I mean to discount any distinction between conscious and subconscious awareness as well as a lack of prevarication. If the attributer has no perception of subconscious denial then the fact of such denial cannot affect his grounds for attribution.

perceptions appropriate to attribution then he meets the attribution criterion but his attribution may differ from the self-attribution of the putative believer.

Perner, (op cit, 169), describes the acquisition of the concept of knowledge in children. He states that this proceeds in two stages. "Children start by using knowledge as a theoretical construct within a *theory of behaviour* that explains success and failure. Later, as children acquire the concept of representation they re-conceptualise what they know about knowledge within a *representational theory of mind.*" Perner is writing here about concepts within language, possessed by people and deployed in linguistic acts. In such acts people talk about beliefs held by themselves and others. From this re-conceptualisation: "... The importance of informational access is fully appreciated, relegating successful action to its place as a 'mere' consequence of knowledge."

Nothing in the above suggests that knowledge or belief are physical states of the holder. Children learn to report; firstly, on what they can say or do and secondly, on their justification for what they can say. On page 148, Perner quotes examples of knowledge talk in children. In one example a child of fifteen months is asked, "Where is X?" and replies "I don't know" (with a palms up gesture). One interpretation could be that the child was reporting on a personal state. Another, more plausible interpretation is that the child was following a practice: If you see or remember seeing, X and are asked "Where is X?" you say "There" and point or you say where you saw it If not, then you say "I don't know". At this stage the language of the child is flat and the child is reporting on the world. Semantic ascent and the textual character of knowledge come with insight into representation. As Perner stresses, beliefs attributed are representations and the acquisition of skills in handling representation and misrepresentation are important milestones in development. But these are representations of convention, explicit within the conventions of language. There is no claim that these are intrinsic and nonconceptual representations within agency.

Is there then a fact of the matter regarding a belief? It is common ground that belief statements are semantically opaque. A proposition embedded in an attribution of belief has truth conditions with respect to states of the world. But this does not bear on whether that proposition is truly believed. Since Freud, we distinguish between conscious and sub-conscious belief and accept that our self perceptions are corrigible. It may be that, with respect to some given domain of discourse, the content 'p' of the representational states of agency of a human agent at the instant may precisely match the content of some sentence 'P'. Also, that these states are consciously available to that agent and that the concepts within that content are equally possessed by all parties to that discourse. Under these conditions it could be said; “that agent truly believes P and P means that p", where p is the representational content and P is that sentence. In so far as beliefs attributed have propositional content and in so far as real representational states have propositional content, these contents may coincide. One could then speak of a true attribution of belief. However, due to intrinsic misrepresentation within agency, the content may not be true and this matching of content may have occurred despite the attribution being dishonest or without proper cause.

4.14 Support for the Hypothesis from the Supervenience of Intrinsic and Conventional Representations

In 2.8 above, I concluded that system properties are strongly supervenient on the physical properties of their realisation only if the system includes those mechanisms by which such properties are attributed and that this amounts to Horgans' notion of regional physical supervenience. A system of agency will be described by a theorist, and the theorist may attribute properties within the functional description of the agency of an entity. For example, a representation may misrepresent and misrepresentation is a system property of a representation. Of the possible future worlds picked out by the existence of this entity those, between which physical properties of the entity are indiscernible, may or may not misrepresent depending on physical states of the world and the truth of the theory of agency. Completeness of the physical system requires that possible states represented include all possible states of the world as classified by the agency of the entity within the theory of a theorist. Misrepresentation requires that the theory be true[34]; the truth of a theory is a relation between the theory and the world and indirectly, a relation between the world and the theorist.

If an action of agency is attributed a *quality* by the theorist, such as that an action is reprehensible, then this attribution is made against a context of practices within a community. These practices are realised within the agency of members of the community and the memories and texts they share. A complete system description will include the community and its practices and in these terms, the *quality* of an action will be regionally physically supervenient on the physical states of the agent and the rest of the community, and the physical states of its texts and memories.

A similar argument applies to the attribution of a belief. Although the representational states of an agent member of a linguistic community are strongly supervenient on the physical properties of the agent and its world, an attribution of belief is made consequentially on recognitions within a linguistic community, shared within the practices of that community.

> The property of being a state of attribution of a belief is regionally supervenient on the physical states of; the agent, the linguistic community making the attribution, their marks in language and the world they share.

[34] In Fodor's sense of asymmetric dependence.

If belief statements were replaced by physical statements, as Churchland, (1981), recommends, then in the consequent flat language, members of a linguistic community would have no way of expressing the theory of such replacement. Any theory of belief or attribution must be described by a theorist, who will be an agent capable of description by that theory The theory must be couched in a metalanguage including the denotation of object and text since it must include statements of belief. There seems to be no principled way in which belief talk or representation talk can be eliminated except by a decision by fiat to cease the practice of belief attribution. But, why should we do that? Folk psychology is not a theory, though it can be discussed within a theory of attribution. It is a practice that has function for the community that practices it.

Whether explanation of belief and representation in terms of physical mechanisms that pick out possible future states by their current actual states amounts to reduction is a matter of definition. Such states do not reduce in the Nagel sense of bridging laws. If, in the liberal sense suggested by Horgan, (ibid, 575) "a metaphysical position will count as reductive merely by virtue of asserting (i) the causal completeness of physics and (ii) the thesis of regional supervenience." then my position emerges as one of reductive materialism but without bridging laws. All of the functional models described are multiply realisable. Even within a realisation, the causal powers of specific functional properties do not correspond with the causal powers of specific physical properties in a type-type correspondence. Even the humblest control system must be explained as a holism of function. My preference is to deny reduction and suggest that physical explanation through regulatory mechanisms is neither eliminativist nor reductionist but requires an alternative term Perhaps Smuts' term 'holistic' is appropriate.

4.15 Conclusions

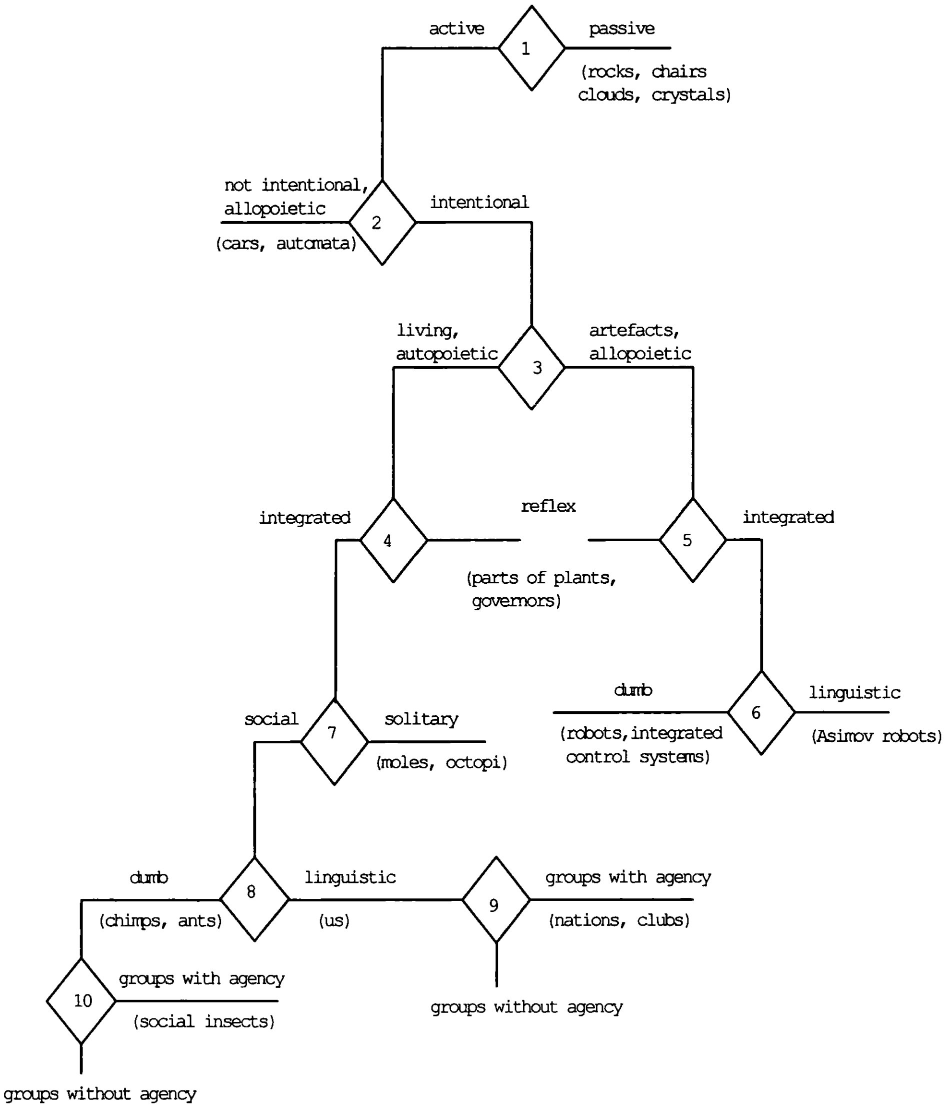
Building on the concept of agency defined in Chapter 3 above. I have described how all living things meet the criteria for explanation of their behaviour in terms of agency. I have also described the agency of those artefacts to which we human agents delegate the tasks of maintaining or achieving goal states against the variability of an unpredictable world.

The range of complexity of realisation between the austere agency of a bacterium, plant part or thermostatic control and the rich agency of a normal, living human is vast and the advent of learning and language lead to emergent phenomena such as personhood and the attribution of belief. My task has been to show that these phenomena, though complex in their manifestation, are not mysterious. The agent at the instant, continually changing in intrinsic representations within its agency and cognitive process and the person, recognised as an ongoing entity within a social grouping of agents, each themselves recognised as persons within the group, is the same physical entity. The same intrinsic representations that are causally maintained by perception and cause action within each human agent are components in the holisms of perception recognised by each person of others and themselves that lead within the practices of language to attributions of belief.

The diagram at the end of this chapter shows how things in this world may be categorised by their intentionality. Whether non-living things can be autopoietic is a matter of debate within the systems discipline that need not concern us here. Only the Asimov robots are not actual things that we can recognise in our world today. The distinction between integrated and reflex agency, at Junctions 4 and 5, is vague. In her description of the cognitive processing of the bat, Kathleen Akins remarks that "the bats auditory system seems designed for sequenced *non-integrated* information processing". The behaviour of reflex agency can be described as functional within the design of an entity, which is autonomous in virtue of its topology and its DNA. However, some of the behaviour of higher mammals is clearly based on integrated information processing and this processing will be generally functional if the creature is to survive. I doubt if there is a clear borderline. The agency of a chimp is clearly integrated in this sense and that of a tree is clearly not. Less advanced vertebrates such as bats are somewhere in between.

I have concluded that, although none of our artefacts of agency compare with ourselves in the rich diversity of their agency, the only distinction of principle between them and us, with respect to agency, is in the aetiology of our functional organisation. We natural organic agents owe the rich variety of our agency to our evolutionary history. Although we can speak of our organic structure as realising the functional requirements of perception, representation, cognitive process and action, our function is for ourselves. Our agency supports sour autopoiesis and that is what it is to be alive. The functional containing systems in which we are best explained are of our selves within our worlds and the functional hierarchy is recursive. Our artefacts have function for us, It may be that we could break that link and create autopoietic artefacts but we would still have made them.

THINGS IN THE WORLD



5

AGENCY & THE COMPUTATIONAL THEORY OF COGNITION

5.1 Introduction

The object of this chapter is to explore the roles of representation and agency in computation and cognition. What it is that is meant by these terms, in this context, will emerge. Each has several usages and they vary in their pertinence. For example, and this is only to gesture at the subject of computation; we may speak of our mental calculations or of our use of an abacus or of our use of a digital computer or of our use of a differential analyser. All of these are examples of our following the rules of some process with the goal of solving some perceived problem. Some authors also speak generally of certain processes as 'computational' even when there is no problem perceived by an agent as a problem for which these processes are the computation of a solution. I will argue that; although cognitive processes have been described as computational, Normal[1] examples of such processes are not computations. Also, the aggregate mechanism in which these processes are realised (in our case, a brain) is not, Normally, a computer for its owner.

I have already set out what I mean by agency. I will continue to describe agency in terms of a functional characterisation of perception, representation, goal, cognitive process and action; where the cognitive processes are such as to produce action that is *for* the agent in the light of the representations caused by what is perceived. Action that is *for* the agent is intentional action. It is *about* the world and is based on representations of that world. Intrinsic representations within agency may be distinct from the representations inherent in the human

[1] As in 2.8 above, I use the term 'Normal' strictly in the sense stipulated by Millikan, 1984.

practices of computation or communication. Within the processes of agency; perceptions amend and produce representations, representations produce and amend other representations and representations cause action. Whether these processes are apt for description as computational will depend on our characterisation of computation and will be a principle question for this chapter. It may be that an act of human agency depends on a computation and that this computation is carried out as a separate act of agency *in the head* of the agent. It will be argued that; although this computation is carried within the central nervous system of the agent it is not part of the cognitive processes of agency. A computation is an act of agency caused by cognitive processing of intrinsic representations within agency. The result of a successful computation is available for perception by the agent.

5.2 Computation

A question that is sometimes asked is whether the mind is a computer and another related question, is whether the brain is a computer. Each of these questions displays some ontological confusion but apart from this, the questions are not clear. Suppose we re-phrase the underlying question into a form like:

> Has the neural realisation of the mental processes attributed in a propositional theory of mental representation, the same functional characterisation as a computer?

The question is clearer but we are still not including in the question, what the concepts of mind and computation are for that questioner. Computation as I will use the term, is an intentional act. So a computer is something which is used by an agent to perform a computation. We turn therefore to the question: What is a computation? Consider a few examples of computation:

(i) If I am asked to divide a quantity of water by three I may do this by putting a third of the water in an open vessel into each of three other vessels. A practical way of doing this is to mark a graduated scale on to a fourth vessel of uniform cross section and then ladle out equal quantities until I have shared the liquid to a sufficient accuracy. In this simple action I have carried out a computation; I have divided the amount in the vessel by three. If the three vessels are identical then the height of the liquid is a representation of the quotient.

(ii) The flow of current through a thin flat layer of electrolyte between point sources and sinks is approximately given by formulae that are the solutions of linear differential equations of an active electric network of a complexity which is related to the number of sources and sinks. The action of setting up the tank and measuring the currents can be interpreted as either solving the differential equations for particular parameters or finding the roots of a pair of polynomial equations or doing an experiment on an electrolytic tank. Two of these interpretations are computations. In one case the locations of the points represent the eigen-roots of the system and in the other they represent the roots of the polynomials. The third is an experiment; they are all the same physical action but with different interpretations.[2]

(iii) If I am asked to divide a number of sweets between five children and tell them how many each will have, I might do it by dealing out sweets one or two at a time. I am still carrying out a computation but in number rather than quantity. The computation lies in the interpretation of the outcome and the representation, inherent in the interpretation of the natural numbers by the quantities of sweets. Computation is an intentional act and in the example, the intentionality is manifest in the telling.

[2] An early description of this device is in Huggins, 1948.

(iv) As recently as 1964 a calculator was available which could be described as a hand held, mechanical abacus[3]. Multiplication consisted in turning a shaft against an array of cams. Division was a precise analogue of long division using the device as a multiplier. One could even calculate square roots.

(v) We are all familiar with the current hand calculators with which we compute products, ratios, powers and surds by fixed programmes in a small digital machine. This is in contrast with my old hand calculator which was programmed by my actions upon a facilitating mechanism for the same sorts of computation as are pre-programmed into the pocket electronic calculator.

(vi) A modern desk top computer is essentially a Turing machine using the von Neumann architecture but it still consists of many electronic gates wired so that they can engage in processes exactly analogous to those of the old hand calculator and those that we do with pencil and paper. Where the computer differs is that the sequence of sums it does is determined by reference to a sequence of instructions that are also held in digital form and can be entered or amended. The machine is hard programmed in this way to read a variety of programmes, as the user requires. This analogy between logics can be exploited so that each such machine is a realisation of a Turing machine.

In all of the above examples, actions are carried out according to rules and some of these rules concern the interpretation of outcome from input as the solution to a computation. The solution may be in number or in text or in a state of the world. Any process can be interpreted as a computation in some context. In the example of dividing sweets between children, this is not a computation unless the agent carrying out the procedure interprets; the number of sweets and the number of children as representing the input to a computation. And the

[3] This was called the "Curta Calculator" made, I think, in Liechtenstein.

method of distribution as a procedure that will lead to the number in each little hand apt for interpretation as representing the solution. Similarly, with the division of the quantity of water, and all of the examples. In each case we have a convention by which certain input states are interpreted as representing a type of question and a procedure that, if used intentionally, will lead to output states that, within a convention represent a solution. Input and output states represent within conventions and procedures lead from a problem type to a solution type.

Turing, (1937) formalised the notion of computation in terms of a Turing machine. The Turing characterisation of computation is in terms of a class of mechanisms defined by their formal properties of symbol manipulation and memory. This has led to the Church-Turing thesis, Church, (1956), that any problem that can be specified in terms of algorithms can be solved by such a machine.[4] Also, since a Turing machine can be specified in functional terms, any Turing machine can emulate any other Turing machine, Galton, (1990).

Neurological research suggests that many of our mental activities can be modelled in algorithmic terms. Also, the neural processes of cognition consist of sequential neural processes and chemical changes to neurons which could themselves be realisations of processes that may be functionally characterised in algorithmic terms. From these insights the science of artificial intelligence has grown. The contribution of computer science and the mathematics of computability to neurobiology have been substantial, as has the contribution of neurobiology to computer research and its applications. Research into the parallel distributed processing capability of banks of neurons has led to the recognition of a new methodology of computing which has led to the architectural thesis of connectionism in cognitive modelling and a new

[4] The notion of computability within a Turing machine was first applied to the resolution of Hilbert's Entscheidungsproblem but our concern here is with the practice of computation.

methodology of non algorithmic optimisation with practical applications in business and economics.[5]

Computation is a part of the way of life of a community that includes a problem set and the agent to whom the problem is posed. We need to be clear as to what aspect of our characterisation of mind we are construing as analogous to what aspect of computation within the linguistic and computational practices of this community.

In language we speak to one another and we are aware of one another as people with a place in our lives and in whose lives we have a place. The implications of this for our attributions of agency, personhood and responsibility were explored in the previous chapter. Also, we compute and examples of computation include those cited at the beginning of this chapter. Every person or dog that has caught a ball has carried out a procedure that can be described in computational terms but we would not describe their actions in catching the ball as computing. What is essential to a computation, as defined, is that the agent knows the permanence or at least reusability, of the methodology or algorithm, of the solution. She then does not just know how to catch a ball, she knows that her procedure is a solution to problems of that type in general. If we divide sweets between children by dealing them out then our actions can be described in computational terms. Had we had not noted the quantity and made a conscious decision to record the number each got as a solution to that problem then we would not performed a computation. If, when a dog catches a ball, we record every detail of the dogs movements and the trajectory of the ball then we may have a solution of that set of differential equations. Even if we could write them down, if we have not purposefully established all of the initial conditions, we have not intentionally solved the equations. It could be questioned whether,

[5] Although the neural networks are usually simulated in algorithmic format using digital computers. See Sharda, (1994), for an application bibliography.

even if we had the initial conditions and we recorded all of the dogs movements and the trajectory in order to solve those differential equations we would have a solution. We would not know which differential equations we would have solved as the dog is an agent and each throwing of the ball is epistemically contingent for the dog. However the throwing of the ball is not epistemically contingent for us. Monte Carlo methods of solving such equations are distinguished by their experimental nature and if by repeated throwing a die, we map some solution space for a range of such equations then we have an approximation to the solution of a mathematical problem[6]. The distinction still lies in the intentionality of the computing agent. If, after the division of the sweets, a child tells another; "There were fifty sweets and we got ten each", then a solution is available. But we have not carried out a computation unless this is why we did it and we knew that the procedure we were following would produce a solution to problems of this type.

One requirement for computation is that there is a problem and a solution. Another is that the solution be a representation. What the solution represents will depend on the domain of the problem but whatever the domain, whether numeric, qualitative or quantitative, if it can be shared in language then it will be apt for description and can be represented by a proposition. This suggests a first stab at a definition of computation as:

> Some physical actions on representations following rules which, carried out intentionally, will lead to a proposition that satisfies a function.

[6] See "Monte Carlo Methods", G W Brown in: " Modern Mathematics for the Engineer", E F Bechenbach, Mc Graw—Hill, 1956. In Cameron, (1960), the author describes an application of this technique to finding a possible refuelling schedule at Berkeley Nuclear power Station.

A 'methodology of computation' then, is a set of rules for actions that will lead to a proposition that satisfies a function. A computation is an instance of the carrying out of such rules, a solution is the proposition that satisfies the function, the problem is the function that the solution satisfies and the methodology is the computer. This is still not complete since we have not mentioned who is carrying out these actions and why. Rephrased to include this, I suggest as a working definition of a computation:

> Some physical actions of agency on representations, following rules mastered by the agent, which, carried out intentionally will lead to a proposition that satisfies a function and is expected so to do, for and by that agent.[7]

A computational act is an act of agency, carried out to get an answer by a procedure. Although a computation may be a private matter the rules of a computation are not private in principle (if they were we could not include 'will' in the definition above). Since computation involves procedures that follow rules any description of a computational act will include a description of physical processes and these processes realise the functional requirements of this computation. Such processes are computational within the context of this description. In another description these same processes may have no computational significance. This point will come up repetitiously during the following arguments.

The definition of computation, given above, is pitched at a very general level. Although; as Turing, (op cit), showed, there are mathematical problems for which mechanical algorithms cannot be found. And as Gödel, (1931), showed,

[7] Harnad, (1994), defines computation as "implementation-independent, systematically interpretable, symbol manipulation". As has been demonstrated, not all computations are systematically interpretable or require the manipulation of symbols, though some methodologies of computation meet these criteria.

there are logical problems for which algorithms cannot be found, there is a human practice that involves the use of algorithms for the solution of logical and mathematical problems. It is this activity that we call computing. The activity comprises: apparatus, rules, a question which is posed in text or conventional representation and may be described as a function, rule based combination of conventional representations and a solution which, if found, will be in conventional representation and will satisfy the function. None of the representations of a computation will be representations of the agency of the computing agent as the actions of computation are essentially contingent on the goals of the computation. The actions of agency are contingent on states of the world and goals of the agent within a theory of agency. They may be realised in neural states of the agent as in mental calculation or in the states of a digital computer as in a calculator or chess playing computer but computing is an act of agency.

An act of agency is an act on the world and the entire embodiment of the agent is in the world. Any apparatus and text of rules, which are used for computation, comprise a computer in the context of that activity. If I am calculating a budget 'in my head' in order to advise a friend then in the context of that question I am using my brain as a computer. I am not intentionally manipulating my neurons to find that answer any more than I am intentionally manipulating electronic gates when I push buttons on a calculator but in each case I am following rules in asking a question of a device. Represented in my brain are the rules of arithmetic as they are in the structure of the calculator. If I use pencil and paper then the computer comprises; my brain, the representations in my brain of numbers and rules, pencil and paper and the marks on the paper. If I use my old mechanical abacus then, since I intervene for numeric entry and register shift, the computational processes are divided between me, using my brain and hands, and the device. The program or schedule of processes and algorithms, are in my brain but so much of the algorithmic work is done by the device that we are justified in calling it a calculator. If I use a modern pocket electronic calculator then the 'computer' is still distributed between the device and me, but a lot more is in the device. The device now actually contains programs for sequences of algorithms and we might be justified in calling it a computer. If I use my desktop computer then most of the computation is in the device. In particular, the device can be re-programmed for a variety of algorithmic processes and we feel totally justified in calling it a computer. In all of these realisations of that computation there was a function, derived from a question, in text and an answer in representational form, apt for interpretation into text.

Collins English Dictionary defines a computer: "a device, usually electronic, that processes data according to a set of instructions." and distinguishes between digital, analogue and hybrid computers. This definition matches my usage above. I suggest as a working definition:

> A computer is an artefact used by an agent to carry out computations by reference to a program that is within it.

The usage of within is very broadly spatial. The program may be distributed between the memory of the agent, written text and a calculating device and we would then see the computer as distributed over these entities. Programs may be recursive and a recursive algorithm may fail to converge to a proposition. Thus, a computer may operate according to its program but not carry out a computation. It may hang up on an inconsistency or settle into a non-convergent process but writing the program and starting the machine will have been an intentional act of agency.[8]

[8] As will be, switching off and amending the program.

Any computation will be achieved by some stream of physical causal processes that lead from some input states that conventionally represent a function, for the computing agent, to some output states that conventionally represent, for the computing agent, a proposition that satisfies that function. These states and these processes are computational under a description and this description includes the agency of the computing agent. Many processes found in nature could be described as computational in that some final arrangement could be interpreted as the solution of a computation if the input were established intentionally for this purpose. Distinction must be maintained between:

> A process, which is intentionally computational, and a process, which is functionally isomorphic to a process which, if used intentionally, could be a part of a computation.

Many cognitive processes actually produce a representational state that can be interpreted in computational terms. If and only if, the agent interprets such states in computational terms then that agent can be said to have carried out a computation and these states are explicitly and conventionally representational. If such a representational state is used in further representational processing of agency or causes appropriate behaviour then the process is not a computation.

David Marr made extensive use of computer modelling in his work on visual representation. Many of the he visual functions that he explained are called computational because successful computer modelling demonstrates that the processes involved are procedurally isomorphic to processes that could be used in computation and can be explained in these terms. However, this does not imply that we human agent carry out computations with our brains when seeing. If some actions of a university are best explained by taking the intentional stance then we could speak of the agency of the university and the cognitive processing of this agency may include a clerk doing sums. Each of these sums will be an act of the agency of the clerk but it may lead to some act of the university, which is not computational for the university. However, the university may require an answer to some question and similar computations by university staff will be components of a computation by the university. In 3.2, above, I defined agency in terms of action, intentional under a description. Computation is an act of agency and will be seen as a computation within a description that includes that agency.

5.3 Representation by Convention

The distinction between intrinsic representations within agency and conventional representation was considered in chapter 2. In chapter 3, I contended that the function of intrinsic representation within agency is independent of convention, although its content can only be articulated in the doubly conventional representations of language. This chapter is concerned with computation and if, as I contend, computation is an act of agency then the representations of computation are, in some sense, conventional and we should consider this type of representation.

Utterances of propositions represent states of the world within the conventions of language. Maps, pictures and sculptures represent within the conventions of navigation and art. Intrinsic representations represent within the cognitive processes of agency. Representations of convention need not have any similarity to the represented within any given domain. Experts in the blending of beverages and perfumes share a vocabulary within which they can describe smells and flavours. Such a description of the smell of a rose is as much a representation of a rose as a painting or photograph or articulate description but in a different domain of description. An essential component of the use of such representation in communication is that it be within a shared domain of description and that the sharing of this domain will define a linguistic community[9]. A non-linguistic higher vertebrate may respond to a model of a conspecific and this response is recognition but although it may recognise representations it will not recognise these as representations. A dog may recognise the marks left by another dog and such markings are signs, singly conventional in Dretske's, (1988), terminology, for the dog. However, without language, the distinction between sign and object signified cannot be articulated.

We speak ordinarily of believing, hoping or fearing and the content of this belief, hope or fear will be some putative state of affairs. Thus, we speak of these propositional attitudes as 'representing' the state of affairs to which the attitude refers. I have argued in 4.12, above, that a belief is not a physical state of an agent but that the attribution of belief within language is an attribution of representational content to representational states of agency and this attribution is a physical state of a linguistic community. This representation is not conventional in the same way as is that of a sketch or icon as the believer may have no perception of those states that are the subject matter of the belief attributed. However, each attribution of a belief or aspiration is an intentional act, aimed at bringing about a representation of a state of affairs. The subject's attitude to it and the articulate attribution of that attitude is a representation, within the conventions of language, of the attitude attributed and it's content.

In computation we also represent. Fodor, (1975), has said that there is no computation without representation. However, Fodor claims that cognition is computation and a major concern of this chapter is to distinguish between cognitive processes that may be procedurally isomorphic to processes within acts of computation and processes that are component processes in intentional acts of computation. To this end we should consider the role of representation in computation.

[9] Such conventions may be private but they can, in principle, be shared.

I have defined a computation as

> Some physical actions of agency on representations, following rules mastered by the agent, which, carried out intentionally, will lead to a proposition that satisfies a function and is expected so to do, for and by that agent.

Processes of computation may take various forms but all require, at least, the representation of a function and its operands and a representation of a solution. If computation is an intentional act then the form and media of these representations will have been intentionally chosen. In the terms of Cummings, (1989), who takes the example of addition; "To get a physical device to add you have to get it to satisfy a function interpretable as addition. And that means that you have to design it so that getting it to represent a pair of addends, causes it to represent the sum."

It is apparent that these representations of computation are of Dretske's type I. By our convention, a state of the world represents and by our convention we interpret consequential states of the world as representational. In any complex computation, physical causal processes are involved that transform states to states according to rules and so the question arises as to whether these intermediate states represent and if so what is the form of their representation? In a computation, the rules may be followed by the agent who is doing the computation or the rule following may be delegated to a device, as in the case of the digital calculator. In neither case is there any implication of interpretability of intermediate states. If one is interrupted in the middle of a long addition, one may use notes of the current states of the calculation to continue after the interruption but there need be no representations of computation within the partially completed sum with some columns untouched.

However, the symbols used will have their usual conventional interpretation as representations of numbers and the sum can be completed.
The designer of a calculator or a teacher of long addition may interpret each stage in the calculation but they are not then doing computation in the sense of the definition above. If computation is an action of agency that leads to a proposition that satisfies a function then computation does need to traffic in symbols. But some computations do use symbols. Harnad, (1994), defines computation as "implementation-independent, systematically interpretable, symbol manipulation". It would appear that Harnad is here using computation as verb rather than noun. If computation (verb) is to be systematically interpretable then this definition is of a process of computations (noun), each of which has an input that can be interpreted as a function and an output that can be interpreted as a proposition that satisfies that function. Each output state and each input state is a type I representation. In the sense of Dretske, type I, it is a symbol. Many methodologies of computation are based on the manipulation of representations of natural numbers and quantities and these representations are also symbols. This symbolic representation is at the core of the methods of serial computation (verb), ranging from the abacus and long multiplication to serial processing within a digital computer. It is within this usage of 'symbol' that the intermediate states of computation in neural networks are 'sub-symbolic'.

Not all methods of computation employ symbols. Apart from artificial neural networks, including their simulation in von Neumann architecture, some methods of analogue computation may use representations that are continuously interpretable. Other methods such as computation using Monte Carlo methods may not have intermediate states apt for symbolic interpretation.

5.4 Functions of Conventional Representation

Functions of conventional representations within systems that describe our world include; communication between members of a community and the use of representations intentionally processed by an agent within the rules of working out a computation. Communication is the exchange of information with shared concepts and this is achieved by the mutual perception of representations of these concepts. Such representations have meaning within the customs of the community and it is within these customs that these representations are interpreted as expressing these concepts. If a computation is an intentional process, that is, some sequence of intentional manipulation of conventional representations, following rules, which leads to a proposition that satisfies a function, then these representations have meaning within the context of the computation. It is within that context that the input, *intermediate* representations and the solution can be interpreted. Representations used in communication and representations used in computation are real physical states, given meaning by the interpretative practices of the communicating community or the computing agent.

5.5 Conventional Representations that Misrepresent.

Misrepresentation within agency has been discussed in 3.7, above. Our conventional usage of representation also leads naturally to the notion of misrepresentation. Within some conventions, a configuration can be interpreted as a representation with a reference and a sense, in the usage of Dretske, (1988, 70). Since the sense of a representation will have truth conditions, if within the domain of the representation, correct interpretation leads to a false sense for this referent, this configuration, in this domain, misrepresents. The same usage leads equally naturally to the notion of metarepresentations of convention. A configuration in a common domain can represent to users of that domain that a configuration in this or another common domain is a representation. If I were to comment of a map that I believed that a symbol on that map represents a bridge then my statement is a component in a metarepresentation of whatever is represented by that map. If I have got it wrong then my statement is a mismetarepresentation or a misinterpretation depending on whether I am describing map reading or using the map. If I am right in my interpretation but the bridge is not there, then the map is a misrepresentation. If I misread the symbol on the map, perhaps because the light is poor, then intrinsic representations within my agency misrepresent the symbol on the map.

Conventional representations represent within intentional acts of agency. Misrepresentations within agency may combine with misrepresentations within convention. The example of navigation by map gives a general illustration of this combination. If I aim to walk to the other side of a river and I have mastered the rules of map reading; including that one icon at the representation of a river means a bridge and another means a ford, then I will walk to a place at a river represented by the map as having a bridge. My action in walking to that place is due to the conjunction in me of a representation of the terrain derived from my perception of that part of the map and my representation of a mark in the map as representing a bridge. If the map is right and I have read it rightly but due to some neural fault in me I can't make the conjunction then I might give up and walk some where else. I misrepresent the river as not having a bridge and this is a intrinsic misrepresentation within my agency due to my abnormal cognitive processing. If due to bad light, I misread the bridge symbol as a ford and give up, then my misrepresentation of the state of affairs at the river is an intrinsic misrepresentation within my agency due to that abnormal perception. If the map is incorrect and the bridge is marked as a ford then my giving up would be due to an intrinsic misrepresentation within my agency caused by a misrepresentation of convention with correct interpretation. If the map is correct and I misinterpret the bridge symbol as a ford then my wrongly giving up would be due to the misrepresentation within my agency caused by my misinterpretation. My acts of agency would be due to my perception of a ford and my subsequent behaviour would be appropriate to my intrinsic representations of a ford at the river.

The detailed processing issues may be further illustrated by the example of a 'chess playing computer'[10]. The system can be taken to include:

(i) Perception of a move by an autonomous opponent, made on a board available to the perception of that opponent.

(ii) Representation of that move, representation of the new board position, representation of rules and the game and strategies.

(iii) Processes by which representations are appropriately maintained and representations of goals, rules and strategies lead to representation of a selected next move and application of that move to the representation of the new board position.

(iv) Procedures by which the position on the common board is appropriately amended as an action of the 'computer'.

This entity is apt for system description in terms of agency. Its actions are intentional under a description and this description includes; either in function or in realisation, perception, representation and cognitive processes by which the representations are maintained from perception so that action for the goals of the entity is caused by the representations. The principle goal of agency; to win, is realised[11] within the cognitive processes as designed.

[10] This term is in fact an oxymoron since, as I argue below, the human brain or any other mechanism that realises the cognitive processes of agency is not a computer for the agent. However, in common usage, we call this mechanism a chess computer.

[11] 'Realised' in the sense of a possible world picked out by the system described.

If the computer makes a bad move this may be because the strategic principles realised in the cognitive processes (programmed into the computer) are inadequate to the complexity of the situation. The opponent, man or machine, is a better chess player with respect to this game. It may be because, due to a faulty chip, the processing of the implications of a possible move led to a wrong representation of the consequent board position. In this case, that representation is a misrepresentation of that implication; in the wording of our logical theory, the processes are not truth preserving. It may be that the perception of the opponent's move was faulty so that the representations of the current board and consequent implications are misrepresentations and are false.

All of the representations in the computer are representations of agency. They have conceptual content for a theorist in a theory of agency within the system description outlined above, they have no conceptual content for the agent. In this theory we may comment that the state of a memory chip represents that the opponent's king is on g1. We may comment that this misrepresents, since the history of the game to date places that piece on f1. In this theory of this agency, representations have wide content, since this theory is of the agent in the world. In another theory of the operation of this entity, such as the theory described by the program specification for the computer, states may be described in representational terms. These terms may be those of chess but their content would be a narrow content, since they do not represent states or the implications of states, of the world.

Within the system description of agency, misrepresentations have content and this content is in error. The error derives from the function of the representation and this function is to track the world or the implications of the world in a way that is for the goals of the agent within the processes that characterise that agent. However, in the case of the chess-playing computer, some occurrent states will be representations of implications of alternative moves. The logical procedures leading to such representations may involve several layers of conditional terms involving the implications of possible future moves by opponent and computer. Such representations do not represent states of an actual world, though some may be actualised, but, Normally, they do not misrepresent. Their function is to represent states of possible worlds. If the conditional terms by which the computer has been programmed represent true implication then, in the Stalnaker sense, they are true of a possible world and will preserve truth in that possible world.[12] It may be that the strategic assessment of a possible board position leads to an inferior move by the computer and it loses. But if each conditional has correctly described logical consequence no misrepresentation will have occurred. The agency of a chess playing computer is limited in breadth, though deep within that breadth. It has goals and these goals are pursued by action in an epistemically (for the entity) contingent world.

If an agent were to use this computer, with these programs to solve chess problems, then the representations in the computer would not be the type III representations of agency. They would be symbols, representations by convention, explicit in Dennett's terminology. The usage of this equipment would then be as a chess computer, rather than as a computer programmed to play chess and each such usage would be a computational act by the user. If a chip were faulty and misrepresented then this would be a misrepresentation within a computation and in this usage of the computer, would lead to an intrinsic misrepresentation within the agency of the user.

If the same computer were used for other applications then the faulty chip may have no conventional representational implications. It may misrepresent at the symbolic level of numerical representation and lead to a misrepresenting output or at the sub-symbolic level of a neural network and only effect the time to

[12] In Stalnaker's, (1968, 34), terms: "The assertion which the conditional makes is that the consequent is true in the world selected. A conditional is true in the actual world when its

convergence. It may have no functional significance for some applications of the computer.

5.6 Cognition as Computation?

Agency is manifest in action that is intentional under a description and the agency of an entity can be explained functionally in terms of perception, representation, goal, process and action. Perception is a causal process by which changes in the world that are relevant for the agent physically cause change in states of the agent. Such states are representational for the agent in that such states cause actions that are for the agent in the context of the world. They are intrinsic representations, Dretske's type III and 'tacit' in Dennett's usage. They are not conventional representations, available for computation by the agent. This is a functional description and it explains actual behaviour if there is a physical realisation that is apt for description in a system that includes the agent in its world. Such a system description will include descriptions of states of the world and states of the agent. It will also include descriptions of the processes by which states of the world change states of the agent and states of the agent cause actions which change the world or the relation of the agent to the world, in a way that is for the agent. These processes I call the cognitive processes of agency.

Any actual description of a token cognitive event will be in some domain of description. The content of this description will be in the concepts of that domain, possessed by the observer. The world perceived would be described in some domain of description and the content of this description will be in the concepts of that domain. The observer may choose different domains of description and bridging statements will link these. Whatever domain is chosen the content of intrinsic representations within agency will be described in that

consequent is true in the selected world."

domain. This content is realistically attributed. That this content is conceptual does not imply possession of those concepts by the agent. The content is not attributed to the sentient agent but to the representational states that explain its actions. Within the terms of this domain an observer may speak of a world event in specific terms.

For example, if we are describing the behaviour of a bacterium we may speak of a spatial field of density of some chemical which, in our theory, is nourishment for the bacterium and this field has a gradient across the bacterium. We may then explain the behaviour of the bacterium in terms of chemical reactions at the cell surface causing differential flagellate activity so that the bacterium tumbles upstream to its own benefit. In another domain of description we may speak of a nutrient gradient represented for the bacterium by chemical states at its surface and the processes built in to the organism causing it to act by tumbling upstream. We may relate these two domains by saying that the representational state has the content that *nourishment is this way for the bacterium.* This content is necessarily indexical and selfish, since what this content is, is for the agent.

Intrinsic representations within the Normal cognitive processes of an agent behaving Normally in its Normal environment are described in a theory of the behaviour of that agent in that environment. This is a functional theory, invoking the functions of perception, representation, process and action. If this theory is true then any token action in the world can be explained as physically caused by representational states that are physically caused by perceptions of the world. If the theory is true then Normal representational states of an evolved agent will be either those that represent those aspects of the world to which it is adapted or those that represent alternative possible aspects for which it is prepared. The functions of these states will be Normal functions in the sense defined by Millikan, (1984).

Lockery, (1989, p. 152), has described the conditioning of the sea hare, *Aplysia Californica,* in terms of *mental code* and a simple syllogism. His interpretation, with minor adaptation, is as below[13]:

T

Code	Physical State
CS_i predicts - shock	SN_i has facilitated terminals
CS_j is - identical (*correlated*) - to CS_i	SN_i and SN_j are coactivated
-----------------------------	--------------------------------
CS_j predicts - shock (*and closes gill)*	SN_j has facilitated teminals (*and fires motor neuron*)

Where the SN are specific neurons and the CS are symbols for the representational significance of the excitation of each neuron. Normally, stimulation of the siphon or mantle skin of the sea hare causes contraction of the gill and siphon. In the terminology of agency; the agent perceives an event in the world by sensory excitation of nerves at the gill, this event is represented by the neural synapse which causes which causes the action of contraction. Normally, the goal of this action will be apt for interpretation as 'withdrawal from intrusion'. The synaptic stimulation is a real state that represents a real event and causes a real action. The observer may choose the domain of description of representational content but the theory will only be correct if the concepts chosen explain the behaviour within the functional model of agency. The concepts 'shock' and 'predict' have been chosen by Lockery. He could have chosen others and if these were linked by bridging laws the correctness of the theory would not be affected.

Lockery stresses in his analysis that this effect is based on an extrapolation of current experiments. Also, the cognitive architecture of vertebrates may be very different. He suggests that; the physical states as conditioned in Aplysia, may

[13] From Lockery, (1989), with adaptation in italics.

be seen as a real representation of the syllogism. In his words, (p 157): "This suggests that *types* of representational state will correspond with *types* of neurophysiological states, insofar as representations of sensory conditions is concerned. This is the kind of relation that underlies type-type psychophysical identity. However, it is important to remember that the relation to sensory conditions is only part of a representational state's functional role and to identify a representational state with a physical state we must show complete causal role isomorphism."

The gill reflex of the sea hare has Normally, the first order representation of 'something at gill' and this neural state causes the Normal closing reflex. If the behaviour which has been conditioned in these experiments, has Normal function for the sea hare then, in the same domain, we may describe the second order representation as having some content like; 'danger of something at gill'. In animals of greater neural complexity descriptions of cognitive processing in these terms may involve more complex representations linked by logical sentences of extreme complexity. Within such explanations of cognitive processes, representations may be attributed simple propositional or logically compound content. Our understanding of the neurobiology of complex vertebrates is still inadequate to support a modelling of our cognitive processes of agency in neural terms but research suggests that: "*types* of representational states will correspond with *types* of neurophysiological states[14], in so far as representation of sensory conditions is concerned." (Lockery, p 157).

Assuming that this is true we may consider, as a first weak hypothesis, the statement:

[14] This is evidently true at the linguistic level.

> H_1: Any action of agency can be explained by a causal chain of physical processes between physical states each of which can be attributed some order of direct or derived representationality relevant to the explanation.

If such an explanation is true then the representations hypothesised in H_1 are real states since each is realised in a physical state. This is not an explanation in system terms as no counterfactual claims are involved. Each explanation is within a narrative description of a specific action and how it came about.

A stronger hypothesis is:

> H_2: If such explanation is true then the behaviour of any animal, adapted to its environment, can be explained within a system description that includes an environment of states perceived and acted upon and an agent containing physical mechanisms that realise causal chains of such representational states.

H_2 is the hypothesis that the physical realisation of cognition is apt for description in system terms. Evidence from neuro-anatomical research indicates that specific areas of the human brain are associated generally with specific cognitive processes. Other evidence suggests that within and between such areas there is substantial reusability of neural connections and that such connections are highly idiosyncratic due both to innate and acquired variability. Such idiosyncratic realisation may not discredit modular functional models, which explain Normal behaviour of Normal humans, but it may not be appropriate to reify such modules.[15]

[15] See "The Modularity of Mind", Fodor, (1983), for a paradigmatic example of this.

The first hypothesis follows directly from the axioms of this thesis. If it is false then our behaviour is not apt for explanation in terms of physical representational states. If it can be sustained then any cognitive explanation of action will have a semantic content and if the theory is sound, the semantics will be truth apt for the observer. This need not be semantic content for the agent animal since it is these cognitive processes that comprise its agency. If the theory is correct then the content of a physical state that caused the action is semantically equivalent to the result of a computation. It may or may not be the result of a computation since the causal process leading to this physical state may or may not be an intentional act of computation leading to a proposition that satisfies a function for this agent animal. If the action explained by H_1 is not a component action within a computation then this physical state is not the result of a computation.

If H_2 can be sustained then any act of agency can be explained by causal chains of physical states that are predictable, at least ceteris paribus and within the limits of the counterfactual consequences of the system description. Any token causal chain of representational states will be comprised of real physical states with representational content that is apt for interpretation as described in the paragraph above. If this model of the cognitive processes of this animal is correct then the semantic content of any representational state will be semantically equivalent to a computation. If some of these physical states were available for intentional manipulation by an observer, who will be an agent, then the cognitive mechanisms of this animal could be a computer for that observer but this does not make the cognitive mechanisms a computer for the animal. Only if these physical states are available for intentional manipulation by the agent animal itself, and only if the system description of the animal is to be interpreted realistically, not instrumentally, will the entity containing these states be a computer for the animal itself.

If all of our cognitive processes of agency are realised in physical processes that necessarily include our neural processes, and the evidence for this is very strong, then the spectrum of cognitive complexity, from sea hare to primate is associated with a range of complexity of neural tricks, evolved over time, within the lottery of climatic and ecological processes. Any one of these neural tricks can be interpreted as a process, which could play a role in some form of computer. Any of these interpretations can be metaphors that increase our understanding of how the totality of these processes realises the functional requirements of our agency. Any one of them could be used by another agent (so far, only we humans seem to be up to it), as a computer by applying interpreted input stimuli and interpreting the output to solve some computational problem. However, none of this means that the whole or any functionally described module, of the cognitive processes of this organism is a computer or that a Normal token operation of these processes should be described as a computation by the described organism.

As Lockery, p 157, concludes: "Obviously, there are a great many cognitive tasks for which there is no functional parallel among the physiologically tractable invertebrates." In any organisms such representational structures will be part of a complex whole concerned with the perception, representation, cognitive process and action of the whole organism and in the more complex vertebrates there will be many cognitive tasks distinct from associative learning. Actual behaviour of the animal will depend on complex interactions between stored representations and logical modelling in representational terms may prove intractable. However, this should not undermine the demonstration of causal role isomorphism and if this is assumed, then the requirements for my functional characterisation of agency are met at least for pre-linguistic cognition.

However, should we speak of these complex interactions of stored representations as computation? We can start with the simpler but fundamental issue of whether we should speak of the learning of the sea hare as a syllogism and we should try to see this from the point of view of the sea hare. This species of creature has evolved into a form such that, in its Normal environment, it can Normally cope with the variations of that environment. This is a platitude of evolution. For the sea hare, these variations include things which bump it around the gill and the creature has a reflex such that bumps cause the gill to withdraw. Also, our experiments have shown that we can change the character of this reflex to a state change that we can interpret as learning. 'Learning' of this kind may or may not have a Normal place in the Normal life of the sea hare. If it has, then there will be occasions in the lives of some sea hares when some intrusion is blocked by a more sensitive gill withdrawal and the process will include a neural modification which we can call 'learning'. If, on the other hand, it has not then this capability of learning is a coincident fact about the biochemistry of neurons, perhaps a fact which has made the evolution of more complex creatures which do 'learn' possible. Whichever of these is the case, the creature is not carrying out the syllogism during the experiment. The experimenter is carrying out the syllogism, using the neural mechanisms of the sea hare as a computer.

If this 'learning' has a place in the life of the sea hare then we could, in a first description, describe a token event of this kind in terms of a logical syllogism. In a second description, we might also say that the sea hare 'recognised' intrusion and had 'learned' to withdraw its gill in response to intrusion. In a third description of this event, we could list the causal chain of physical events from the first physical contact to the first synaptic stimulation and the way in which subsequent physical contacts caused chemical changes in particular neurons which in their turn caused changes in the overall reflex. The second and third of these descriptions are of the sea hare. The second is in terms of its agency; the creature is described in functional terms of perception, representation, which includes the syllogism, and action, and its representations include those of items which we may describe in our theory as 'irritation' and 'persistence', and action for the goal of survival. The third description is of a physical machine that realises the functional description of the second. Each of these are instances, cuts in possible worlds, from alternative system descriptions of a real sea hare. The first description is an interpretation of a sub-process within the cognitive processes of a sea hare. By describing this sub-process in these terms we are commenting on the fact that by manipulation of the sea hare we could use it as a component in a computation.

Alternatively, we could comment that; if we wished to produce an alternative realisation of the functional description of this sub-process, then any circuit that realised the functions of this syllogism would realise this particular functional component. This latter approach is the platform of AI and the approach is not belittled by our recognition that it does not describe the organism.[16] This sub-process is a fraction of the complexity of a living sea hare and the simplest vertebrates are more complex that the sea hare. If we can demonstrate that functional theories of complex cognitive processes really explain behaviour in modular terms and if we can demonstrate that these functional theories can be interpreted in computational terms then our understanding of these cognitive processes will have increased. Such functional characterisation, together with neurological investigation, may show the actual physical realisation of this computational function to be very similar to a the realisation of a computational device; whether analogue or digital, serial or parallel, serially programmed or connectionist, which we can actually make and use for computation. But this does not imply that the physical embodiment of this process in the living organism is a computer.

A living organism may be apt for description in many systems. From the inside it can be physically described in terms of cellular mechanisms and boundary conditions or functionally described in terms of autopoiesis. From the outside, in a system that includes its environment, it can be described in terms of physical interactions or the functional characterisation of agency. From the outside, a creature with language can be described in a system that includes an environment, which includes other members of a linguistic community to which it belongs. This latter system will include the customs of the community, some of which are to do with language.

Fodor, (1975), said that there is no computation without representation. For Fodor, cognition is computational and the representations in such computation are explicit representations, interpretable within the language of thought. Obviously, boringly obviously, it depends what you mean by 'computation' and I have addressed this question. If you define any process, which could be a part of a computation as a computational process, then it is reasonable to describe a tokening of such a process as a computation. However, the net is too wide. As we have seen, many processes that are observed in nature or are Normally part of non-computational acts could, given an intervention of external agency, be a part of some computation in some context under some interpretation. If as I suggest, we define a computation as an intentional act of agency which includes an interpretation and a methodology then we should define a computer as a machine, with rules for use, picked out by its functional role in a class of computational acts. Such acts will be acts of agency.

A biologist demonstrating a syllogism with the neurons of a sea hare may be using that neural structure as a computer but it is not a computer for the sea hare. In its functional role as a computer for the biologist, particular neural states can be said to represent the terms of the syllogism - *no computation without representation* - but this representation is within the computation of the biologist for the biologist. It is his intentional acts that apply the shocks and observe the behaviour. The solving of the syllogism is only a computation if it

[16] A similar point was made by Putnam, (1988)

is carried out intentionally to get an answer and it may be argued that the biologist is only concerned with how the creature protects its gills. But, if the biologist describes the neural process as a syllogism then he is describing this in computational terms, as illustrated in the table above. If the learning process demonstrated by this has a place in the life of the sea hare then a similar token neural state may represent a prediction for the sea hare. Within one of our possible domains of explanation of the behaviour of the sea hare, we may say that the creature learned that this area is bad for gills. Within another domain, we may say that this learning is realised in particular neural processes. We should only speak of computation or syllogism in this context as a comment on the process analogies between the neural processes of an animal and the functional processes of a computer.

This distinction goes all the way up. In research into human cognition we may find that various computational procedures such as component processes in von Neumann machines have analogues in neural cognitive processes. We may also find that specific cognitive processes may be explained at the neural level in terms of parallel neural networks and such neural processes may find useful application in our computing machines. This synergy between biology and computing has been extremely productive, both in extending our understanding of the neural mechanisms of cognition and extending our competence in applying electronic computer technology to complex problems in diverse fields.

The human brain is not a computer for the human who owns it. If, some alien intelligence were to use one of us to do his sums by intentionally applying stimuli and interpreting the resultant behaviour then that human would be a computer for it. If one of us does a piece of mental arithmetic for another then some components of her brain are used as a computer by the person who specifies the sum and interprets the answer. This distinction cuts both ways. The Von Neumann machine we have on our desks may be used as a computer. It may also be used for word processing or as a database or as a chess opponent.

It is a tool, we call it a computer because, historically, that has been its principal use.

5.7 Conclusions

A computation is an intentional act of agency; it comprises a sequence of processing of representations, at least some of which are conventional because their interpretation is an intentional act, inherent within the conventions of the computation. Cognition, as in the cognitive processes of agency, is not an intentional act. The intentionality of agency depends on the processing of intrinsic representations. Cognition is not computation even though, the realisation of functionally described modules of neural architecture may be procedurally isomorphic to modules, apt for use in computation.

What we commonly call a computer is an artefact with the function of computation. A digital computer is a current realisation, using the von Neumann architecture, of a Turing machine. Thus any cognitive process, apt for description in system terms, can be simulated, to a given degree of accuracy, on a digital computer. The human brain or that of any complex mammal may be apt for description in system terms. This is hypothesis 2 of section 5, above. If this is so then, in principle given sufficient computer power and time and knowledge of Normal neural processes, a brain could be simulated on a digital computer. However, this does not make the brain a computer. If, some agent were to use these computational isomorphisms to solve some class of problems by input to and output from the neural mechanisms of a human's brain then that brain and the input-output transducers would be a computer for that agent. But it would not be a computer for the owner of the brain. Even if that agent were the owner of the brain, that brain is not Normally a computer for its owner. If the representations within the brain are component representations within a computation then they are explicit representations capable of interpretation by the computing agent in the context of the computation. In such usage, the representations are not intrinsic representations, either innate or caused by perceptions of agency, and within the cognitive processes of agency, causing action for the agent owner of the brain.

A human agent may carry out a computation using explicit representations of number, quantity or proposition and some of these representations may be within her brain. It is difficult to imagine any computation in which some explicit representation of quantity or procedure is not within a human brain. But this does not make the brain a computer or our Normal cognitive processes of agency computations. The distinction is between the processing of representations within acts of agency, some of which may be computations and the processing of representations within the cognitive processes of agency, some of which processes may be procedurally isomorphic to some processes of computation.

6

DETERMINISM & AGENCY

6.1 The Problem of Materialism

Some materialists face an apparent dilemma. These are those materialists who are concerned to retain their own intuitions of free will, to attribute responsibility appropriately to others and accept it appropriately for them. Their materialism constrains them to accept the responsibility for our own destiny, which is involved in the acceptance that our physical constitution owes no debt to any immaterial agency. They believe that the physical world is characterised by physical laws, which constrain the possible outcomes of physical events. They are concerned that this belief implies that these physical laws determine the outcomes with which they are concerned so that these laws negate our responsibility for the outcomes of our actions.

My concern in what follows is to attempt to solve this dilemma. This solution, as I will argue, comes at a price. If a man's responsibility for his actions is bound up with his freedom to choose to act otherwise, then we must distinguish between two aspects of the same material entity. I will argue that we can and must distinguish between the agent which acts and is causally responsible for its own actions and the person who is that agent and who is held responsible within a linguistic community for those acts. The person, who holds himself responsible for his actions within a conscious narrative of the self, is quantitatively the same material entity as the agent which acts and generates that conscious narrative. The person held responsible for himself and others is the same material entity as the agent which acts in ways that change its future agency and that of other agents within the linguistic community.

This distinction between the agent and the person is not a Cartesian dualism. It is a choice between two modes of description of a single material reality. We must also distinguish between actual metaphysical freedom and I will hold with Hobbs that such freedom can only be construed as an absence of physical constraint, and the freedom of a person to choose to develop or degenerate and to 'change her mind'. This, latter, 'freedom of the will' is a social construct with correctness conditions of attribution within the customs of the community, whereas actual metaphysical freedom is a matter of fact apt for ascription of metaphysical truth.

I have argued previously that agency is essentially regulatory and thus actions of agency can be explained within the theories of cybernetics. Within this account, each action of agency can be explained as a unique response to an unpredictable event. We human agents live normally within groups, characteristics of which include language and the attribution of personal responsibility. The customs of attribution of personhood and responsibility are social constructs that have evolved within our linguistic communities. I shall argue that the physical responsibility of actions of agency is a fact of agent cause. I shall also argue that this is compatible with the attribution of personhood between human agents who are members of linguistic communities and that a component of the attribution of personhood is the attribution of personal responsibility.[1]

Actions of agency are determined and this determinism is not incompatible with a freedom of choice which is expressed in the autonomy of the agent to chose its response to new events according to its own states, within its own regulatory disposition. Actions of the person are actions of agency but the personal

[1] Since I am a realist about the realisations of physical systems this makes me a social realist. People are not fictions constructed by our narratives. Their reality lies in the reality of everyday practices of recognition, blame and praise; by real humans about real humans.

descriptions of on-going agency include the attribution of personhood to the agent recognised by another agent. Within such continuous sequences of descriptions responsibility is attributed to the person for actions of agency which includes the agent changing over time. Each person within this community of attribution is a self-conscious agent. This consciousness includes the memories of events and the recognition of self and others by which the attribution of personhood and personal responsibility to self and others is maintained. This recognition of self and other and the custom of attribution of responsibility psychologically entail the feelings of personal responsibility for actions of self and feelings of blame or praise for actions of others. As people we are taught to choose and we learn to explain our actions of agency as actions of choice. These explanations lead to profound intuitions of freedom of choice and the existence of these intuitions is compatible with the determination of each of our actions by our states at the time of the action and the incident events to which we are reacting. From this I shall argue that our intuitions of freedom of the will, within our community of personhood, are compatible with the determination of our actions by our material states and perceived events which includes the actions of ourselves and others.

6.2 The Determinist Debate

Determinism is the view that everything that happens has a cause. Cause is a notion of explanation, we seek for a cause to explain an event. If this is all that there is to 'cause' then determinism is the view that every event can be explained in causal terms. Indeterminists deny this and their denial of this is the statement that there are some events whose antecedents do not make them necessary. The basic intuitions of indeterminism are that if we are free to choose our actions then they must be underdetermined and that determinism makes nonsense of moral responsibility. These are cogent considerations and I will be concerned to answer them. Compatibilism holds that such causal necessitation does not reduce our freedom to choose since, although these choices are links in the causal chain, they are themselves freely made. I would count my argument as for compatibilism were it not that my use of freedom is not the one that compatibilists are concerned to reconcile with determinism.

Galen Strawson, (!986, p. 31) sets out the distinctions:

> Incompatibilism holds that the falsity of determinism is a necessary condition of freedom. In itself it involves no view about whether determinism is true or false, or about whether or not we are free. 'Libertarianism', by contrast, is the name of a positive incompatibilist theory of freedom, one that purports to show that we are free and so assumes (or argues) that determinism is false.

I will be concerned to challenge the libertarian thesis since I will argue that an ontological commitment to the reification of the concepts of 'freedom of the will' and 'freedom of choice' is incompatible with the regulatory determination of actions of agency. On his first page, Strawson, articulates the notion of freedom that concerns the libertarian and which I am concerned to oppose:

> Freedom is now defined in terms of true responsibility, true responsibility in terms of desert and desert in terms of freedom.

I will argue that desert is a matter of attribution by a community for actions of a person where personhood is also attributed by a community to the on-going agent. Real, matter of fact responsibility, for the actions of the agent can only be attributed to the real, matter of fact agent at the instant, who acts.

Strawson concludes his account with three reasons why;

"... [We] are not really free and truly responsible agents at all, even if we cannot help believing that we are.

> (1) There is a clear and fundamental sense in which no being can be truly self-determining in respect of its character and motivation in such a way as to be truly responsible for how it is in respect of character and motivation.
>
> (2) When we act at a given time, the way we act is, in some quite straightforward sense, a function of the way we then are, in respect of character and motivation. We act as we act *because of* how we are then in terms of character and motivation.
>
> (3) It follows that there is a fundamental sense in which we cannot be truly responsible for our actions. For we cannot be truly responsible for the way we are, and we act as we act because of the way we are."[2]

Strawson grants a special epistemic status to the narrative of the conscious person and argues that insistence that our conscious experience of freedom is a real freedom in his strong sense would entail a distinction between human and theoretical truth. He concludes that the price is too high.

I agree that this price is unreasonable and should not be paid. If we define freedom in simple terms of an absence of physical constraint and describe attributions of personhood and personal responsibility as social constructs within the life of a human linguistic community then actual responsibility remains defined as the *responsibility* of agent cause for its consequences and no contradictions arise. Feelings of personal freedom and responsibility are phenomena, explicable within our theories of the neurological basis of cognition and the customs of attribution of mind and personhood that have evolved within our communities. The price we pay is the distinction between the agent at the instant who acts and the on-going person whose personhood is continually reattributed by the recognitions and analyses of the conscious self and others in the community.

6.3 Cause and Determination

The community is aware that its environment can be described in causal terms that include the laws of science. In an earlier age this community attributed choice and responsibility to all components of its environment and this has led to a dual usage of the term 'cause'. If a stone falls by erosion, blocks a stream which floods a village and kills a man then the falling of the stone or the blocking of the stream or the flooding of the village can be said to have caused the death. If a man kills another from malice or greed then that man caused that death, not the swinging of the arm or the cutting of the sword but the act of the man. These two examples use two distinct notions of cause. Following our usage of, 2.9, above, we may distinguish these as physical cause and agent cause.

The modern physics of quantum theory denies that our physical world can be fully explained in terms of constant conjunction since, at the level of the very small, events may be alternatively described by theories that are mutually inconsistent. Thus, a particle may be ascribed a position or a velocity but not both and can be described as a wave motion or particle but not both. Also, nuclear events occur with probabilities that are independent of any known incident cause. For the purpose of this enquiry, I will stay with the causal implications of classical physics since this is the backcloth to our normal attributions of freedom and responsibility. Also, the distinction between the determinate descriptions of classical physics and the indeterminate descriptions

[2] Ibid, pp. 311-2

of quantum mechanics need not effect our attributions of responsibility. If I place an enemy for five minutes in a cell containing some element which emits a lethal particle according to a Poisson distribution with an average time between emissions of five minutes then he has a fair chance of survival. But should he die, his death will be as clearly my responsibility as if I had shot him. If one of the reasons that caused me to act in some way included a random event within my brain then my act of agency was as determined by my states as if each had contributed to a determinate causal chain. We may speak with pride of serendipitous reasons for our actions but not of risking a fortune on the turn of a card. The former opens up new choices and the latter constrains our agency but both introduce a random element into our choice of action.

To say that an act is determined is not to say that it is programmed. In its usage in human affairs 'programme' implies a programmer: an agent who intentionally designed and implemented that programme. Alternatively, we may say of a stream of events that they are so patterned that they appear programmed since we can predict future events in the stream. In this usage 'programmed' appears synonymous with 'determined'. However, as we have seen, the determination of an act of agency is not of this kind since the act is determined by the current state of the agent and the current state of the world and these are independent, prior to an act of agency.

That an event is determined does not imply that it is predictable. We must distinguish between predictability for an observer and predictability for an agent. A perfect observer, Laplace's all knowing intelligence, may be able to predict the actions of every agent. There are cogent arguments from cybernetics and quantum mechanics to the effect that this could not be so and that, even if it were, nothing could be done with the prediction but this is not to the point. An event that requires regulatory choice from an agent is necessarily unpredictable for that agent since this is why it requires regulatory choice. The programmed act of an agent may have been determined by previous choice of that agent but as I have argued previously, it is still a regulatory act. If this choice included formal randomisation of the outcome, the randomness of the outcome is determined by the choice of the agent, although the actual quantitative outcome is not. Thus, within this usage of 'determination' a random event may be appropriately described as determined. Since each such choice is determined, each act is determined.

Determinism is often opposed to the arbitrary or the random. The thought here is that if a stream of events cannot be predicted then some outcomes must be arbitrary and the stream of events is then essentially random. The definition of randomness is a vexed and difficult issue that I have not space to pursue. What is at stake is nearer to the opposed notions of correlation and orthogonality. The intuition is that responsible choice must be in some way independent of the outer world of events and that this independence will be manifest in orthogonality between event and action. Some philosophers have looked for this independence in the combinatorial processes of decision. Dennett[3] follows a hunch of Russell in suggesting that a choice is: " ... [A]n intelligent selection from what may be a partially arbitrary or chaotic or random *production"* Some aspects of our mental processes may well be random in this way but this is speculation and we are looking for an ontological criterion, some reason why the very idea of responsible choice implies an orthogonality between event and action. Dennett was looking for the right thing but he was looking in the wrong place. The orthogonality between event and response is necessary from the nature of agency as a regulatory process.

Any unconstrained act of an agent is an act chosen by that agent in response to some event. That choice by that agent is determined by that event since it is the response of that agent as constituted at that time to that event. Also, that choice

[3] "On Giving Libertarians What They Say They Want" in Dennett, (1981).

by that agent is determined by the constitution of that agent at that time. Thus the response of that agent to that event is determined. Reflex, autonomic and integrated acts of an agent are determined by the inherited and developed constitution of that agent, which is itself determined by previous causal chains in the ontogeny and inheritance of that agent. 'Determinism' is an observer concept, belonging in an open system description of the world containing the linguistic community of agents in which the terms of freedom and responsibility are used. 'Freedom' and 'responsibility' are agent concepts belonging in the closed system description of the linguistic community in which constraint is an option and responsibility is attributed. Any agent member can describe his community in open system terms and recognise its determinism. This is totally compatible with his attributions of freedom and responsibility to himself and others, including agent members of other species or even his own artefacts of regulation in his closed system descriptions of his community from his point of view[4]. This external notion of responsibility is of an epistemic responsibility attributed by members of the community in virtue of their beliefs about that agent in contrast with the metaphysical notion of responsibility as necessarily relating an agent to its acts of choice.

In our explanations of behaviour we distinguish between predictable and unpredictable behaviour and we link predictability with determinism and unpredictability with autonomy and the freedom of the agent to choose. I contend that we should distinguish not between predictability and unpredictability of behaviour in the light of events but between the predictability and unpredictability of events for the agent. When real choices occur we act according to our own nature. We are responsible for these choices because they are our choices but each choice can be explained by reference to the entity we are at the time of the choice. We are not responsible for being that

[4] This is the solution to the dilemma posed by Nagel, (1986) in the words; "We can act only from inside the world but when we see ourselves from outside, the autonomy we experience

entity. To follow that logic leads immediately to regress as it conflates the attribution, by an articulate community, of responsibility to one of its members with the responsibility for the way that member is. A mother may rear a child in ways that inculcate feelings of self-reliance and personal responsibility in the person whom that child becomes. That person will be aware of her responsibility for her actions and will be held responsible by the community. That mother is at least partially responsible for that person. Another mother may bring up a child to a lack of feelings of personal responsibility. The community will also hold the person whom that child becomes responsible for her actions. She may not be aware of her responsibility for her actions but they will still be her actions, each determined by events in the world and her current agency.

6.4 Regulation

I have previously characterised agency in terms of regulation. Also, I have argued that the thick notion of agency that we attribute to ourselves can be characterised in terms of complex, integrated, learning and reflexive regulatory processes. It is these processes that are realised in the complex mechanisms of our brains and it is these mechanisms that realise the functional system of perception, combinations of representation, goals and action that describes the regulatory theory of agency.

The complex agency that we attribute to ourselves includes our conscious and subconscious processes of decision, our reflexive articulation of these to one another and ourselves in language and gesture and our co-operation and competition with one another. The complexity of the physical processes involved are currently beyond our descriptive powers but it is an axiom of this thesis that all realisations are physical Another axiom of this thesis, argued for

from inside appears as an illusion, and we who are looking from outside cannot act at all."

in 3.9, above, is that this thick characterisation of agency is still essentially regulatory. This is a thick and complex realisation of regulation. It is within this realisation of regulation that each action of a normal human being in a normal human community can be described as a component of a regulatory response and can be described as determined by the regulatory processes that are realised in the human agent.

We are animals and animals react to change perceived. But we are very complex animals and our reactions are to complex patterns of perceptions of states of affairs; past and present, sensed and learned. I shall use the term 'event' as an abbreviation for this complex pattern of perception but every such 'event' is assumed to include some phenomenon of change which is relevant to us in that it requires some action or chosen inaction by us. An event is a real change in a real state of affairs. An event is identified by its location in a causal nexus that may include its perception by an agent. An event thus described, is a particular.[5] A person may choose to examine a state of affairs that may not be changing but this act of observation will have been caused. There will have been some particular event in which a part of its causal nexus was the perception that led, via the cognitive processes of agency, to the choice of this act of observation. An act of observation may also be an act of reflex agency as a sudden noise may draw the attention.

Within this theory, the cybernetic theory of regulation, especially that regulation is essentially counter-entropic, can be applied to the explanation of our behaviour in the world. Regulation is counter-entropic because it maintains order within the entity against disorder in the world. Order and disorder are relative to the agent. Each action of an agent is a response to events in the world, perceived by the agent, that are contingent for the agent. Such events may or may not be metaphysically contingent. Laplace may have been right and all events are predictable for an all knowing intelligence but they are epistemically contingent[6] for the agent. The decision processes of the agent have evolved to produce the action that is best for the agent by the criteria of these processes. It is because of the harmony between 'best for the agent' in the processes and 'best for the agent' in the world that this sort of entity in this sort of community in this sort of world has survived. Because of this harmony the regulatory agent can cope with events that are predictable in type, in that the processes can cope but are not predictable in token and it is this sort of unpredictability that regulation is about. For the agent, each token event could have been otherwise in the system that picks out possible worlds of states of affairs, relevant for this agent.

Any regulator is in the business of maintaining or achieving some goal state against variation in an environment. It is of the nature of regulation that this variation should be unpredictable for the regulator. This is what regulation is about. If variation is predictable then for any agent designed or evolved, the behaviour of the response is programmed. From the humble temperature controlled chamber to the guided missile and from the prokaryotic cell to us: we are all in the business of regulation. As an example of an artefact of regulation without consciousness, consider a guided missile.

Firstly, the missile is equipped with propulsion and steering capability and with sensors which respond to such features of its relation to its environment as its orientation, wind pressure, obstacles ahead and features of the terrain below. Secondly, it is equipped with physical representations of these, features in its world, representations of its prime goal of landing on a destination and representations of secondary goals such as keeping going and avoiding objects ahead, some of which may be intentionally aimed at it. Thirdly, it is equipped

[5] This is exactly the characterisation of events given by Davidson, (1969) and (1970).

[6] The notion of 'epistemic contingency' was introduced in 3.9, above.

with devices by which representations are combined to derive actions that are appropriate to the situation in the world for the represented goals that have been programmed into the missile. All of this design is based on predictable types of event which are relevant for this missile with its goals. It is designed as a regulator because actual, token events of this type are essentially unpredictable for it. If a gust of wind shifts the orientation then the missile will correct. This gust of wind was not predicted but from the instant of the gust and its sensing, the response of the missile was determined. In the terms of our discussion we may say that the missile was responsible for that correction and that it chose to correct in that way. There might have been a learning capability in the electronics so that the designer could not have predicted the response but it would still have been determined for that missile for that event. There may be clutter sent by an enemy to confuse the controls. The missile will still be responsible for any response that it makes, including one which leads to its destruction. If an enemy succeeds in affecting the combinatorial equipment in the missile so that its reactions led to self-destruction then this action would still be the choice of that missile but it would not be of the same agency. In the terms of our previous discussion, this is not now a Normal agent. It, the modified agent, is still responsible for what it does; though our judgements of that responsibility have changed. If, in the design of this missile, token events of wind and debris had been predictable or ignored then the missile would have been programmed and its behaviour would not be apt for explanation in terms of agency. Had the enemy, by some transmission, converted the regulatory arrangements to a programme then, the missile would cease to be an autonomous regulator and could not be a responsible cause. Its actions would be those of the agency of the enemy.

As with the missile, so it is with us. We have goals; some of these are delegated to autonomic functions like breathing and sweating and others are central to the self to whom integrated agency and the intentional attitudes are attributed. Of these latter, central goals, some are inherited and some are acquired. Many of our actions are part of a regulatory response but are programmed within that act; as when we ride a bike, drive a car or even chatter amicably. The distinction between programme and choice can be a matter of degree. We may stop to check a map, consider, choose and drive on and all of this within our decision to attend that conference. We are all aware that interwoven with the stream of expected events, for which we are programmed, is another stream of the unexpected, things that require decision. It is to our response to these unpredictable events that the notions of choice and responsibility apply. Organisms are programmed as well as regulatory. Each of us is programmed from conception for our cellular development and growth and this programme is coded in our DNA. A part of this programme is the engineering of our various regulatory capabilities, which range from simple reflexes to our ability to learn complex strategies for choice in our social world.

Regulation is concerned with response to the unpredictable. If an organism is provided with a simple response to an anticipated event then that response is programmed by that provision. This is the regulation of reflex rather than the regulation of choice. If an organism is prepared to cope with the unexpected then it is prepared to choose. There is no hard line between the response of choice and the response of reflex. As we vertebrate organisms have evolved in neural complexity our regulatory arrangements have evolved into a tangle of feed forward and predictive arrangements that defy description in terms of functional circuitry, but the simple truths remain. When we are responding in a programmed way to predicted events we are not making choices. When we are making choices we are dealing with the unpredictable and these choices will be for the goals of the agent making them. The goals of the social agent often extend beyond the selfish needs of the solitary agent. We are social animals and our group behaviour has evolved to include sacrifice for the group or the family. The social and inherited conditioning of an adult human has led to a concept of goal that is complex but it is for the current goals of a unique adult human that its choices will be made. Such choices will be its choices made in response to those events, for its goals. They are acts of agency, interpreted as choices by a theorist since they could have been otherwise.

6.5 Actions of Agency

Before we consider our responsibility for our actions we should examine what it is to be an action of an agent. If we consider the agent as the entity in the skin then an action of that entity is any change in the entity which is caused by the internal processes of the entity and which causes change in the *world of the agent*. This is an inclusive definition, the entity to which we attribute agency is in the world and it is of the world. Thus, the world of the human agent includes the neural structures within which the processes of human agency are realised. An act of human agency may be a thought, such as the running through of a mnemonic, by which the act of the agent of the present may change the agent of the future without any visible change to other members of the community who attribute personhood to that agent.

Some of our actions are the result of regulatory processes that are autonomic in that they are not normally under our conscious control. We breathe and we sweat; we can hold our breath to survive briefly under water but we cannot restrain our perspiration. It would seem then that some of our autonomic processes are within our conscious control and some are not. In speaking of conscious control I am not concerned, with the distinction between our conscious and sub-conscious mental processes. I am concerned with the distinction between those of our actions that are autonomic within us as integrated agents, as the tropisms of a leaf or root are autonomic within a plant and those which are actions of the integrated agent. Since I can hold my breath the question of my responsibility for my breathing or not breathing can be sensibly raised. It may be that we physically hold our breath within a dream. Should this happen then it would be the act of the agent though not within the narrative of the conscious self. If it is a physiological fact that I cannot start or stop perspiring then any agency attributed to this regulatory process is a thin autonomic agency and attribution of responsibility is merely a question of causal explanation. Other actions are reflexes such as a knee jerk or a blink or a sneeze. Again, within this class of actions, we can distinguish those actions that we can control or inhibit from those which are autonomic. A knee jerk can be inhibited by a separate constraint but is generally not within our control, I don't know if I can always control a blink but I do know that, within variable limits, I can control a sneeze. It would seem that among our reflex actions, some are actions of autonomic agency but some may be sensibly described as actions of the integral agent and are therefore, apt for attribution of responsibility to the integral agent.

Other actions which may be described as components of autonomic processes are the component actions of activities like walking, swimming, cycling, or playing a well rehearsed piece on a musical instrument. Each of these activities may be initiated by an action of the agent but component actions are so habituated that they may be sensibly described as component actions within an autonomic process. Again, as with our inherited autonomic processes, the distinction can be made case by case and we can distinguish; within those of our actions which are components of a practised programme, between those that have become autonomic and those that are, at least potentially, acts of the agent. For these latter actions, the process can be stopped or restarted and these are acts of the agent. How or when we make the distinction is not important, what is important is that the distinction can be made.[7]

Some of our actions may be component actions within plans that we have initiated. Within such a plan we expect that an action will lead to some consequent event which we will perceive and if this event is 'according to plan'

[7] This is close to the argument in Ryle, (1949, II.3)

then our action will be 'as planned'. A first intuition is that such actions are not regulatory since they are not responses of the agent to events contingent for the agent. However, just as an act of sneezing is not autonomic since it can be inhibited, so our actions within a plan may not autonomic since events may not be according to plan. Although an event within a plan may be according to plan, it could have been otherwise and its occurrence is epistemically contingent for the agent. If the event, as perceived by the agent, is recognised by the learned and evolved cognitive processes of agency as requiring further combinatorial processing of representations then, in a Normal agent, this processing will occur and this may lead to the action of a change of plan. Thus, even within a plan proceeding according to plan, actions of an agent should be described as regulatory.

The thick notion of agency with respect to plans can be contrasted with the thin agency of inborn reflex action and the programmed actions of cycling or rehearsed performance. The essence of a programme is that each step necessitates the next and since, at least some, of our programmed behaviour may be interrupted by an act of conscious agency the distinction between plan and programme is vague. But starting and stopping programmes and starting and stopping plans are all regulatory at some level or range of levels. I conclude that all actions of human agency are regulatory in that they are caused by those complex combinatorial processes and those representational states and goals that functionally describe this agent. How these functional states and processes are actually realised in we humans is a matter of on-going research and debate and is not the subject of this thesis. Except, in so far as I rely on the contention that this realisation is within a physical stuff, the behaviour of which is, in principle, capable of description by the laws of our physical sciences.

If an 'agent acts' within the sense of agency and action that I have identified with regulation then this action is this agents response to some event in the world. That there is a response is caused by this event but this agent causes this response. It is that response which was causally implied by this agent's current set of representational states and this agent's current structure as this agent's response to this event. In 1971, Davidson posed a dilemma. He wrote: "Either the causing by an agent of a primitive action is an event discrete from the primitive action in which case we have problems about acts of the will or worse, or it is not a discrete event in which case there seems no difference between saying someone caused a primitive action and saying he was the agent." (1980, p. 52)

Davidson rejected the first horn of this dilemma since the causes of the primitive act do not include the agent. Doing is synonymous with agent cause although intentional action is physically caused through cognitive processes that combine representational states. The second horn does not impale us since an agent does not cause an act of agency. Agent cause is synonymous with primitive action. As Davidson remarks: "an agent causes what his actions cause". Primitive intentional action will generally result in events beyond the intention of the act.

Our normal form of speech is to say that an agent acts or that an action was the action of an agent. An entity which is a token of the natural kind 'human being' is a causal nexus that can be described in different ways and one of these is the system description of agency. If a clerk within a university were to write to a prospective student on university headed paper, we might say that the university wrote to the student. We would not say that the university physically caused the letter to be sent, though we would say that consequences of the sending of the letter were caused by the action of the university. The clerk caused the letter to be sent within the causal nexus of roles and procedures that are described within a system description of that university.

All actions of agency are regulatory. Thus, all actions of agency which are apt for the attribution of choice can be described as actions chosen by the agent for the agent. These actions may be actions of initiation, actions in reaction to the unexpected or the default actions that are within an on-going plan. I stipulated earlier that action is change of the agent that causes change in the world. But this change in the world depends on a world in which such change may occur. The agent may choose to act but the act may be constrained. The agent may not be free and we need to consider what it is for an agent to be free.

6.6 System Descriptions of the Temporal Agent

The basic functional system of agency comprises perception, representation, goals and action. In we humans and in all organisms, this system is realised entirely in physical components and all processes are physical processes. This is the naturalistic stance. It is an axiom of this thesis. An action of an agent is an action of that agent at the time of that action and is determined by the states of that agent at the time of that action. If those processes by which the action is determined are internally constrained in then any such constraint is a physical state of the agent at that time and is therefore an attribute of that agent. If we construe freedom as an absence of constraint and we attribute agency to the corporate entity that acts then, we cannot sensibly speak of freedom of an agent in terms other than the lack of Hobbsian physical constraint on action.[8]

If we question whether, under some characterisation of agency, an agent can be described as free to choose then we must distinguish between the freedom of the agent and the freedom of the choice. We may postulate some state of the agent, say the last representation prior to action as his choice and the other states of the agent as having come to that choice. If we somehow affect these other states so as to restrict their power to affect this state of choice then we

[8] Peter van Inwagen, (1989), argues that, in the classical philosophical usage of 'free will' our wills are at most very rarely free and that this does not imply that they are ever free at all. My arguments of the preceding paragraph conclude that statements about of *will* and its *freedom* are phenomena of practices that have evolved within a linguistic community.

have changed the agent. This change may be more or less permanent. It may be possible to remove from an agent the power to decide. This may or may not be reversible but at the time of inaction this agent is an agent that does not act because the combinatorial processes by which action is normally chosen do not converge or in the terms of another theory, 'it can't make up its mind'. The community may identify this agent by continuity of external features but the agent who acted was the agent existent at the time of the action or inaction. At any time there is only one actual, physical agent but this agent is in the environment, which includes other members of the linguistic community. Any system description of agency will be an abbreviation, since no finite description could include all of the alternative events and their effects on the states of the agent that will determine his actions. We may choose from various alternative descriptions of an agent, not physically constrained.

(i) We may describe the agent as instantaneously constant in its physical states so that the counterfactuality of the occurrent system will relate to some range of possible events. Outcomes will be possible actions that are the responses of this agent at this time to these events. In this system, the actions of the agent in response to unforeseen events are determined by the physical states of the agent. The agent is free to act since it is not constrained. Its actions are the actions of this agent and it is this agent which causes events by action and it is this agent causation that is responsible for the consequences of such action.

Within a token system of this type a state of affairs perceived by the entity will be described as leading to an action. However, an observer may say that the agent will 'choose' an action and that what the agent does is the agent's choice. For the observer, within this system description of the agent in an environment, the agent could act alternatively since the agent is not constrained and the agent could have been otherwise. Some of the actions of the agent at the instant may be such as to change the agent in the future. Each such action will be apt for explanation in terms of perceptions, representations and goals at that time. The current goals of agency may or may not be consistent with the future health or welfare of the agent in the future but it is the agent at the instant that acts.

(ii) We may describe the agent over some period of time and characterise the agent by his physical states at the start. The counterfactuality of the system will relate to the changes of the agent that will result from events in the environment and from his actions that are responses to events in the environment. Also his actions will change the environment and thus influence the subsequent events to which he will respond. This description will include the conscious self of the agent, which may also observe and describe those aspects of agency available to consciousness in both narrative and system terms. Each action of this on-going agent can be described in the terms of the first description, each action of the agent described in the first description will change the agent in the future and will be described in the narrative descriptions of the on-going agent. Thus, the on-going agent is determined by the actions of the agent at the instant and the agent at that instant is free, although its character and dispositions which determine its decisions are, at least in part, determined by past actions attributable to the on-going agent. The agent becomes but what it is that becomes is the on-going agent described. It is the agent at the instant that acts and we should not confuse the determination of its acts by its current character and dispositions with the constraint of an agency external to the evolving agent. To use 'freedom of the agent' as an absence of some past determination is to confuse 'freedom' which is a lack of physical constraint with the absence of some determination of the agent by its past.

We humans have consciousness. How this works is not known, although Dennett, (1991) has given us a plausible explanation. We share with one another, in language, our experience that when consciously faced with alternatives we are conscious of our 'choosing'. In our conscious experience we are aware that we could act alternatively. Our descriptions to others and ourselves of such experiences may be narrative or system. We may end a narrative with a phrase like ' ...and this is what I chose to do'. Also, we may consciously describe ourselves in system terms and it is in such descriptions that we attribute to ourselves the power to choose.

(iii) We may describe the linguistic community, which includes the agent. In such a description we will include other agent members of the community whose actions affect the agent and those members of the community whose actions describe the agent in both narrative and system terms. To say an agent is describing herself or another in narrative and system terms may not be to attribute any articulate complexity of description. We all commonly make statements like; 'I wonder what he will do about X since I know that he likes Y' or 'I wonder what X will do to Y'. We may also ponder in terms like: 'what shall I do about X given that I like Y', without giving much thought to the counterfactuals involved and distinguish these from narrative descriptions of what somebody did. Within a system we may describe how the community, which includes the agent, will, either separately or collectively, attribute responsibility to the on-going agent that they describe. Within the customs of the community and its concepts, denoted by terms in the language in which the on-going agent is described, the responsibility attributed by the community is made normative. An agent may be described as responsible, in that consequence appears to have been considered before action or irresponsible if it does not so appear. It is within this system that we can describe how the customs of the community may include punitive or rewarding behaviour toward agents and the criteria by which the community attributes personhood to its members. It is within this system that we can describe the community questioning its own attributions of responsibility to people and debating whether determinism, freedom and responsibility are compatible.

Within the linguistic community, we attribute responsibility to agents in terms of their choice, which we attribute synonymously with the attribution that they could have acted otherwise. This attribution of the power to choose, that he could have acted or could act otherwise is an attribution of choice to the on-going agent as identified by the community. This usage in our communities is not unambiguous. If we are aware of a rationale for the agent, which we expect him to find compelling, we may say that he could not have acted otherwise. This inconsistency does not mean that the agent has no power to choose, we are drawing attention to our deduction that this rationale makes his action more predictable for us; it has reduced the contingency in our description of this on-going agent in this environment.

6.7 The Person in the Community

We may describe a community as comprised of autonomous agents, animals like us, who are bonded by the social customs of the community, which include a spoken and written language. This system also includes an environment, the events of which, members can partially control and partially predict. Each of these people is an animal with an inherited repertoire of biological drives such as fear, lust and anger and these are denoted by terms in the language. Content that is associated with the customs of the community may be denoted by such terms as desire, belief, responsibility, choice and freedom and these become concepts possessed by members of the community. Other content is of features of the world for the community and this are described by propositions using concepts. Also, each of these people is an animal with a repertoire of goals. Some of these are inherited, such as staying alive and finding a mate, some are acquired through interactions within the community. Some are articulated in propositions using concepts denoted by terms in the language.

One practice of such a community is the attribution of belief. Such attribution is associated with the exposure of a member to perceptions of states of affairs in the world or exposure to communication about such states of affairs, which he may be expected to take as reliable indications of the truth of the proposition denoting that state of affairs. Another practice of the community is that of occasionally restraining members from action by force or edict and the community recognises that accidents of the environment may also restrain action. Thus, they may say of a member not so restrained that he is free to act although, in this attribution they are not distinguishing between the freedom of the agent and the implications of a perceived edict for an agent's choice of action.

Just as the community attribute belief to members because of their observed perceptions and freedom because of their lack of physical restraint, so they attribute responsibility to members on account of their observed actions. This attribution is based on the appreciation that the response of a free agent to an event is his unique response to that event. Within the community, an agent member can be described in terms of a unique set of goals and perceptions, which lead to the attribution of a unique set of desires and beliefs. Not all of these goals and perceptions are consciously available to the agent or other agent members. Some are inherited and some are forgotten. Those not articulated are private to the individual. Each agent can be attributed any of a denumerable infinity (denumerable as they are logically derived from a finite base) of desires and beliefs that are logical consequences of those which are apt for attribution because of goals and perceptions that are acquired or inherited.

A human community is comprised of people and it is as persons, that members of a community attribute responsibility to one another. The concept of a person is distinct from the concept of an agent. Locke says that "person is a forensic term, appropriating actions and their merit; and so belongs only to intelligent agents, capable of a law, and happiness, and misery. This personality extends itself to what is past, only by consciousness-whereby it becomes concerned and accountable."[9] This is close to my distinction between the on-going person to

[9] An Essay Concerning Human Understanding, Book 1, XXVII, 26

whom responsibility is attributed and the actual responsibility of the agent at the instant. The Lockean account was opposed by Bernard Williams, (1973), on the grounds that false memories can be engendered. For Williams, personhood is supervenient on physical continuity. Parfit, (1984), questioned whether bodily continuity is adequate since a baby is initially of the same flesh as its mother and a person can be physically reduced without loss of personhood. Parfit concluded that personal identity is unimportant and that what is important is the normative nature of personal continuity. I will argue that personhood is a matter of attribution and that normativeness is within the practices of the community.

Dennett, (1976) gave six conditions for personhood. In order of dependence, these are:

(i) that persons are rational beings,

(ii) that persons are beings to which states of consciousness are attributed or to which psychological or mental or intentional predicates, are ascribed,

(iii) that attributions of personhood depend in some way on the stance adopted to it,

(iv) that attributions of personhood are reciprocal, people attribute personhood to others,

(v) that persons are capable of verbal communication,

(vi) that persons are self conscious in a way that they share.

The first three of Dennett's conditions almost amount to the forgoing characterisation of agency. Dennett's insistence on the intentional stance is famous but it does not contribute to this analysis. If, as Dennett argues and I agree, an intentional system will Normally behave in a way which is rational in the context of its intentional predicates, then rationality is a component of intentionality. If an entity is apt for description as an intentional system then that description will be the act of an observer. It is this act of description that Dennett describes as 'taking the intentional stance' but the entity is prior to its description. We are talking here about real entities, which may be described in various ways. An entity will be apt for description in a particular system whether or not such a description is articulated. The agency of an agent is prior to its membership of a community or its description as an agent by members of that or any other, community. Using the methodology of system description we can rephrase Dennett's first three conditions as follows.

> Necessary conditions for the attribution of personhood to an entity by other members of a linguistic community include, that the members of the community recognise the agency of that entity. This condition implies that the entity be apt for system description as an agent and that this description be made and used by members of the community.[10]

This condition, together with Dennett's remaining conditions, amount to a definition of a linguistic community in terms of its criteria for membership and could be rephrased:

> A human linguistic community is a social group of human beings who are agents with self-consciousness. Customs of the group include participation in a common language, which includes articulation of self-consciousness and recognition of other members

[10] Though not in the technical usage of Chapter 2.

> by their agency and articulate self-consciousness. The term used reflexively within the group to name the social kind picked out by this custom is 'person'.

Described in this way, necessary and sufficient conditions for personhood are contingent on the customs of the linguistic community, which in their turn, will depend on the physical constitution of the members. Another custom of human linguistic communities is the attribution of 'responsibility'. Responsibility is attributed for the consequences of actions of members. Thus responsibility for the consequence of an action is attributed to a person. The self-identity of the person and the identity of the person for the community are within the system descriptions of the agent by the agent and of the agent by other members of the community. Each of these is of a continuous on-going entity described in factual and counterfactual terms. But actions of agency are actions of that agent at that time and will be determined by the states of that agent at that time. Such actions of agency may include actions, which change the agent in the future. The system description of the agent, which exists in principle though not available to the community in practice, will include in its counterfactual terms the possibility of actions which change the agent in the future by internal deliberation or external consultation. The person for the community or the person for the self-conscious agent is an on-going entity whom they recognise and their recognition will be of a sort of running average of the changing agent. Because of these distinctions 'responsibility' has at least three senses.

> (i) The responsibility of the agent at the instant is the basic fact of agent cause. Just as a stone that falls into a stream is responsible for the flooding of a village so an agent who releases a brake is responsible for deaths caused by a runaway vehicle. Agent cause is distinct from external physical cause in that it is determined by the regulatory actions of agency.

> (ii) Responsibility attributed to a person by other people is for those events caused by actions of the agent recognised as a person. However, responsibility is attributed by people to a person and this person is recognised as an on-going agent responsible not only for his actions of agency that effect others but also for those actions of agency that effect his future self.
>
> (iii) As each agent attributes responsibility to others so each agent attributes responsibility to himself. Attribution is itself a conscious act of agency which is determined by incident events and the current states of the agent. Thus, although members of the community may differ in their attribution of responsibility to a person and may or may not reach a consensus within the institutions of the community, such attribution is distinct from the responsibility consciously attributed by an agent to the person whom he recognises as himself.

For each human agent, the self of consciousness is the awareness, more or less articulate, of the current instant and this awareness will include the currently consciously remembered past. The continuity of this self for the self-conscious agent lies in the continuity of memory. This self is continuously changing but in a continuous process of change in which each self, as it is, is a change from the previous self, as it was. The self of consciousness is more or less aware of its responsibility for its actions and these may be actual actions of the agent or actions imagined. The current self experienced in consciousness is determined by current and previous states of the body, including the neural system, some of which are immediately caused by events in the world. This is the conscious self of Dennett, (1991). Dennett offers an explanation of what is going on within us when we produce, for others, or ourselves a description of what we thought, felt, heard or did. In his words,

> "ur tales are spun, but for the most part we don't spin them; they spin us. Our human consciousness and our narrative selfhood, is their product, not their source. These strings or streams of narrative issue forth *as if* from a single source ... their effect on any audience is to encourage them to (try to) posit a unified agent whose words they are, about whom they are; in short to posit a centre of narrative gravity. (ibid, p. 418)

Just as the self of consciousness is continuously changing so the person, identified by the community and to whom this self is attributed, is changing. This, also, is a continuous process in which each state is a development from the previous state. This process of change is recognised by the community and the identity of the person for the community is maintained by the continuity of the person in the conscious memories of the member agents and the text of their records. Similarly, the identity of the self for the self is maintained by its continuity of the self in the memories of agent and the text of his records. Although we may separately describe a self and a person, these are two descriptions of one entity, which may also be described in terms of agency.[11]

6.8 Attributions of Freedom

An animal is free if it is unconstrained. We are free in so far as our actions by which we manifest our choices are unconstrained. If I have chosen but my choice has been made in the light of my knowledge of statutory laws then my choice was not constrained. My knowledge of the law and its consequences are among my representational states at the time of the action. If I have chosen but my actions are constrained, say by tying my hands or removing paper and

[11] McCall, (1990, 7), makes this point clearly: " ... there exists only one ontological entity, a biological social and self conscious being, this entity is thought of and conceived in different ways. The terms person, self and human being, when referring to the concepts or ways of

pencil from my cell then I am not free to act. Such lack of freedom is an aspect of the situation and not an aspect of the agent. The choice of the agent, to which we attribute responsibility, is free in so far as it is unconstrained. This choice is determined by the events to which it is a response and the representational states at that time of the agent who chose. The distinction between choice and action would be described in counterfactual terms such as: he would have acted thus had his hands not been tied. An action may be made predictable by a constraint but the choice, what the agent would have done, is unpredictable in so far as it is a reaction to an event, which was unpredictable for that agent. Although that choice was determined by that event and the current states of that agent.

I shall hold axiomatically that an agent is not free to act if it is physically constrained and that this is all that there is to freedom. Physical constraint may be intentional, as when one agent restrains another or it may be accidental as when a man is trapped by a fallen tree. The act of constraint will be the act of an agent and the constraint will be caused by that agent and may be held to be the responsibility of that agent. An accidental constraint will be an accidental cause and the notion of responsibility does not arise. Without such constraint an agent is free. This is Hobbe's notion of liberty, set out most clearly by Locke, who wrote:

> [T]hat liberty is not an idea belonging to volition or preferring; but to the person having the power of doing or forbearing to do, according as the mind shall choose or direct.[12]

conceiving an individual have different, distinct meanings." In the terminology of this thesis they are component concepts in different system descriptions that include the same entity.

[12] Book 1, XXI, 10

Locke went on to argue that "liberty, which is but a power, belongs only to *agents*, and cannot be an attribution or modification of the will, which is also but a power." This may be regarded as a very weak notion of freedom but any extension of this notion is incompatible with my account of agency. Given this account, the extension of the notion of freedom beyond that of physical restraint on the power of the agent to act on the world is incoherent.

6.9 Attributions of Choice

Agency is concerned with perception and action. Simple realisations of agency such as; unicellular organisms, simple artefacts of control and the components of plants, act in a way which can be reliably predicted by an observer who understands the mechanisms of agency and can measure and interpret variations of the environment as it is for the agent. Such actions of agency are by the agent and are for the agent. They are determined by the events in the environment and the states of the agent. They may be predictable for an observer in virtue of the transparency of the mechanisms of agency. We would not normally speak here of the agent choosing to act in this way since we know that; although for the agent, the event and response are unique events of agency; for the observer, the event and response are programmed by the mechanisms involved. The observer can choose the action by her choice of stimulus. However, each Normal action of agency is an act of choice of the agent since the events perceived are epistemically contingent for the agent.

In more complex realisations of agency such as the higher vertebrates, the octopus or complex artefacts of regulation such as a learning robot, the mechanisms of agency are no longer transparent. Actions of agency are still determined by the states of the agent and events in the environment but the observer cannot now choose actions of agency. If this were an artefact we would normally say that choice has been delegated to the agent. If this were an organism we would normally say that the organism has the power to choose.

Since outcomes are essentially unpredictable for the observer the observer must say of each action that; for him, the observer, the agent could have acted otherwise. Choice, on this interpretation, is a modal notion and is linked to the modal 'can'. Taylor, (1966, 40-56), denies that the 'can' of human agency is *merely* a matter of *epistemic contingency.*[13] He concludes that: "what is meant by 'can' in: 'I can move my finger' is that it is something 'within my power'."

Taylor defines epistemic contingency in terms of the knowledge of the agent. In this context, I am describing the agent in system terms and the contingency that I describe is epistemic for the agent, described in that system. The idea that something is 'within my power' is an idea articulated within the stream of consciousness of a conscious agent and belongs in another system description to which I shall return. Choice is attributed within a system description of agency in which the agent can not be manipulated by the observer and is implied by the autonomy of the agent.

Most complex animals can be described as making choices. When in discourse we comment on such choosing we can without any anthropomorphism attribute the responsibility for that action to that animal and thus attribute the responsibility for that choice to that animal. This is not to suggest that each animal is self-consciously aware of having made that choice. Our attribution of choice to ourselves is within our description of ourselves as agents and it is within that description that we are aware of our power to act. If, for the reasons previously given, we attribute agency to other creatures then our attribution of choice to them is within the functional description of agency. If as agents, we act, then as agents, we choose and those choices for which we wish to consider the attribution of responsibility are those which are in response to the unpredictable in our lives. Other complex creatures, such as dogs, cats and our anthropoid cousins, display individual personalities and it is with reason that we

[13] Authors italics.

hold them responsible for their actions but we cannot tell whether they hold us responsible for ours. Each of these creatures, like us, displays a range of reflex behaviours, which can be distinguished from the behaviour of choice, which is manifest in its reaction to complex challenges of the world. A cat that decided to go out on a wet night rather than soil the house and Buridan's ass had a choice to make. Where we differ from other animals is in our social articulation of our notions of free will and responsibility and this is based on our articulation of our sense of self to others and ourselves within the customs of our linguistic communities.

Three senses of 'choice' emerge from our three system descriptions of the temporal agent. These appear similar in content but they are in distinct domains and should not be conflated.

(i) An observer, considering an open system description of an agent, is concerned with a metaphysical notion of counterfactuality associated with the agent system. When he says 'the agent chose' as a synonym for 'the agent could have acted otherwise' he is describing in system terms the variety of outcomes that are contingent on the current states of the agent. We could say that he is describing the possible worlds accessible from the actual world prior to the action, in the absence of detailed knowledge of the states of the agent.

(ii) An observer considering a system description of a conscious linguistic agent will describe the agent subject as describing his conscious experience in the vocabulary that he has learned within the linguistic community. During this learning, he will have been offered choices say, between sweets and treats. He will have learned that his decisions may have consequences. His narrative account of conscious choice will be in concepts learned by observation of his and other agents, action and consequence that he has learned to call 'an act of choice'. This is a pre-scientific account but it must be in some way like this that we learn to choose and master the concept of choice.

(iii) The community usage of 'choice' will be close to that of the conscious agent since they are in the same language game. The attribution of choice within a linguistic community is a custom of the community. In a system description of a linguistic community we would include the practice of attributing choice to an agent as, in the description of an agent, we attribute choice to an agent. In the description of an agent we are describing that agent as it is now. The *can* in the counterfactual account derives from the contingency associated with this agent, here and now. In the description of the community we will also include the customs of the community, which include the attribution of the on-going identity of personhood to each agent member. For the community, 'what a person *can do'* will be taken over time. The person who has chosen to become depraved could have acted previously so as to become otherwise and in the eyes of the community, is the person who executed a depraved act. According to the customs of the community, this person should have not executed this depraved act but this person chose to act in this way and could have chosen otherwise.

6.10 Attributions of Responsibility

The responsibility attributed to an agent is a relation between an event and an agent. The attribution will be by an agent about an event and an agent. I will take it as a matter of definition that an event is the responsibility of an agent if and only if that event was physically caused by an intentional action of an agent. If an action is the action of an agent then it is the action of that agent at that time. This is a metaphysical notion of responsibility and it derives from the epistemic notion of contingency which I have described above, It is a matter of causal explanation of a regulatory response, no different in kind from the autonomic action of perspiration[14] though enormously more complex in degree. It is distinct from the legal practice of holding a person responsible for those consequences of an action that he or she should reasonably have foreseen. Within this notion of agent responsibility, an agent is causally responsible for the results of unforeseeable actions of its offspring but not intentionally responsible. An act of agency is by definition intentional. Whether specific consequences of the primitive action[15] lie within that intention is determinable by reference to the theory of agency that explains that action.

This strict restriction on the temporality and autonomy of agency has important implications. (i) If an agent is aware of edicts which penalise certain actions but he is physically free to act then his action, whatever it is, as an un-programmed action of choice in response to an event, is his action of choice and he is responsible for that action. (ii) If an agent is radically but temporarily changed by ingestion of some substance, such as alcohol or a truth drug, then it is this, changed agent who has acted and that choice of that action is the responsibility of that agent at that time. The responsibility for the ingestion of the substance may be of that agent if he imbibed the alcohol freely or of another if the truth drug was administered by force. Subsequently, the recovered agent may be judged by the community, which includes himself, as to his desert[16]. Responsibility for the act cannot logically be separated from the agent who chose the act. The acts of administering a drug or encouraging the consumption of alcohol are the acts of agents who are responsible for their choice of these acts. (iii) If an agent is changed permanently by physical or psychological trauma, as by brain damage or brain washing then it is still this, changed, agent

[14] The odour of perspiration is the responsibility of the thin autonomic agency of the body's cooling system. The use or not of a deodorant is the responsibility of the thick and complex agency to which personhood is attributed.

[15] This is Davidson's, (1971), usage of the term.

[16] The custom of the community is to attribute responsibility to the on-going agent and judge desert according to the contribution to consequence or the state of the on-going agent, by accident or the actions of other agents.

who is responsible for his acts. The community may have sympathy for the agent, anger with the agent who caused the trauma and mourn the loss of the agent gone. However again, responsibility for an act cannot be logically separated from the agent who chose the act. I am not, in this thesis, considering the theory of justice or desert, which I see as separate issues from freedom and responsibility. (iv) If an agent is hypnotised and instructed by the hypnotist to carry out acts without the intention of the agent then these acts are the acts of the agent hypnotist. The system that describes these acts must include the agent hypnotist, the means of communication and the passive agent whose body is being used as well as the environment in which the acts are performed. (v) An agent can choose to change his future self. The actions resulting from that choice are the responsibility of that agent at that time. The future agent will be responsible for his actions as and when he acts. (vi) An agent can act to change another as when a mother encourages a child to face the consequences of its own actions with the intention of producing an adult aware of its responsibility for its own actions. The mother is responsible for the consequences of her actions as the future adult will be responsible for his. An adult, not so trained, may not be aware of its responsibility for the consequences of its actions but it is still responsible as that agent at that time which acted with that intention.

People attribute personal responsibility to people. The act of attributing responsibility is an act of agency and that agent at the instant that acted is responsible for the consequences in the world of that attribution. Normally, personal responsibility is attributed to a person who is recognised by the attributer as that agent who is that person.[17] Attributions of responsibility may have consequential implications through the community; A attributes

[17] In the aberrant case of multiple personality an observer may attribute responsibility between separate people whom he distinguishes within one body. At any time there is one agent who acts and metaphysical responsibility rests with that agent. This metaphysical responsibility can be distinguished from the responsibility, epistemic for an observer, attributed by that observer to a person identified by that observer.

responsibility to B for the consequences of his attributing responsibility to C for the consequences of C's action. A, B and C are each people who have each acted and each of these actions have been actions of agency with consequences in the world for which these agents are responsible. When we speak of a person as 'personally responsible' we are drawing attention to the identity between the person and the agent in the context of an event which is the consequence of an action either in prediction or in actuality. In all of these events of attribution of personal responsibility the attributions are acts of agency and each attribution is within a system description of that human being as a person who is an on-going agent.

6.11 Spontaneity and Responsibility

Some writers hold that the spontaneity of human choice stems from the rationality of human nature. Kant wrote, and McDowell, (1994, p. 4) has quoted:

> If the *receptivity* of our mind, its power of receiving representations in so far as it is in any wise affected, is to be entitled sensibility, then the mind power of producing representations from itself, the *spontaneity* of knowledge should be called the understanding. Our nature is so constituted that our *intuition* can never be other than sensible; that is, it contains only the mode in which we are affected by objects. The faculty, on the other hand, which enables us to think the object of sensible intuition is the understanding. To neither of these powers may preference be given over the other. Without sensibility no object would be given to us, without understanding no object would be thought. Thoughts without content are empty, intuitions without concepts are blind.

For Kant and for McDowell, it is from this spontaneity of knowledge that rational decision stems. In McDowell's words:

> When Kant describes the understanding as a faculty of spontaneity, that reflects his view of the relation between reason and freedom: rational necessitation is not just compatible with freedom but constitutive of it. In a slogan, the space of reasons is the realm of freedom.

McDowell's concern is to show that spontaneity is a human faculty, a second nature which is developed with language by the normal human conditioning to rational behaviour. This he calls Bildung (ibid, p. 84). His argument goes something like this. We know that we are free or at least, this is our fundamental intuition. We know that we are rational and we ascribe the effectiveness of our rationality as responsibility for our actions. If a rational agent is responsible for her actions then these actions stem from her, from her understanding and her rationality. They have not been caused by external agents within some causal, law-like structure. They arise spontaneously, from her. Hence the hypothesis that her exposure to the processes of Bildung have enabled this spontaneity. McDowell, in company with Davidson, (1982), but for different reasons, rejects the suggestion that other animals beside man may posses any of these characteristics. Davidson rejects the notion that other animals may have beliefs since without language they have no concepts. For McDowell, second nature evolves out of first nature, which we share with other animals, but only man develops this second nature. In Part IV of the Afterword to his six lectures McDowell is at pains to deny just Kantian spontaneity to dumb animals. He freely admits that cats and dogs are not automata but states that the freedom they lack is the freedom that consists in potentially reflective consciousness to putative norms of reasoning.

McDowell is making two distinctions and each of these extends the usage of an everyday word into a complex conceptual holism, which is some way from its ordinary usage. Freedom is now not just a lack of constraint on action but a freedom that consists in potentially reflective consciousness to putative norms of reasoning. Spontaneity is now; not just an original response, though predetermined by current events and current agency, but a response that depends on understanding within a conceptual structure. Each of these extended definitions depends not just on language, but also on the metalanguage of reference to text.

I agree with McDowell that obedience to putative norms of reasoning by an agent requires understanding of these norms and this requires reference to the text in which such norms are held. However, a child who has recognised that turning knobs is wrong is no more and no less, responsible for so doing than a cat which has recognised that jumping on beds is wrong. McDowell would say, and I would agree that a person who can knowingly articulate their responsibility for turning knobs has entered the space of reasons. But, the child and the cat are still free in that each, may or may not, act against their conditioning. None of us finds any inconsistency in blaming the child or the cat for the resultant mess, though our behaviour may be influenced by the ages of the child and the cat.

What emerges from this is that Kantian freedom and spontaneity are social constructs. In early language, we learn what is or is not permitted. Before or without, language we act and observe the consequences of our actions. As fully practising members of a linguistic community we can speak of responsibility and share with others our articulations of our responsibility for our actions. This is McDowell's second nature but he speaks of acquiring a second nature as: " ... having ones eyes opened to reasons at large." (ibid, p. 84) In this usage, it is by the recognition of our responsibility that we are responsible and by the recognition of our rationality that we are rational and this would seem at risk of leading to an infinite regress. This regress may be avoided by the suggestion that our rationality is acquired step by step with our acquiring the ability to recognise and articulate it. This is an empirical question and research may answer it. My account, above, is an alternative approach by which we can restore simplicity to our notions of freedom and spontaneity and retain a baldly naturalistic account of us in our world.

We are rational animals in so far as our behaviour is rational. Each action of any animal is its response to some situation. This situation may be more or less complex. It may be a dog's response to a cat climbing a tree, it may be a child's response to a car arriving at the time Daddy comes home or it may be an adult person's response to a letter from the bank. The dog may bark at the tree, the child may run to whoever gets out of the car, the adult may decide to sell the car or pack his bags and run away. Each of these responses is that of a rational animal. The action may be more or less rational; the dog may be barking up the wrong tree, the child may run up to a salesman, the adult may even have misread the bank statement. We, in the linguistic community, will assign responsibility to the adult because he is a fully-fledged member of the linguistic community who has had training in money. The community has judged that person as responsible and that person knows that this is so. These are the practices of this community. It is this and many other social practices that define this community. The dog had spontaneity and freedom, there were pressures on it to chase that cat and bark but there was no strict, deterministic law in virtue of which it could be predicted. The child did not have to run to the car: it did because it chose to. We can analyse this singular event and explain this running but it was the child's choice; not one the child might choose to explain, except as; "I thought it was Daddy".

6.12 Conclusions

If determinism is the thesis that every event has a cause then; within the domain of classical physics and given the distinction between the causes of events, which are epistemically contingent for an agent and the responsible cause which is an action of agency, the truth of determinism is self-evident. If determinism is the thesis that; given full knowledge of a situation then from the laws of science, again within classical physics, the outcome is predictable then again I agree, since the response of an agent to an event is given by the combination of this agent and this event. Although, in a material world the observer will be a regulatory agent and his world is epistemically unpredictable for him. Although its outcome is predictable, the event is unpredictable. If determinism is the thesis that all events are predictable then I hold that this is false, since an agent is a regulator and the regulatory response of an agent is to events that are epistemically unpredictable for this agent.

Within the above definition of determinism if an agent is free to act then the actions of that agent are free since this is all that is meant by freedom. If an agent is mentally changed but left free to act then the actions of this agent are still free but it is a different agent since it has been changed. Strawson, (1986), in his opening paragraph, defines 'free' in "... the ordinary strong sense of the word. According to which to be a free agent is to be capable of being *truly responsible* for one's actions". I think that Strawson's usage is less natural and more problematic than my own but each usage is coherent. My use of freedom is in the tradition of Hobbes, Locke and Hume. This usage is resilient to arguments of agent change since, such changes modify the agent but leave him free unless constrained.

If incompatibilism is the thesis that determinism and freedom are incompatible then in the terms of the above definitions, I am a compatibilist. If libertarianism is opposed to compatibilism then I hold that libertarianism is false but the central tenet of libertarianism that we are free to choose is true within my usage of 'freedom' and my usage of 'choice'.

Whether responsibility depends on rationality depends on the community notions of responsibility and normality. A Normal agent will behave rationally in that its actions are a rational response for it to events that are relevant for it in the world. An agent may be abnormal in that its capacity for rational response is defective. How the community judges the responsibility of agents is contingent on the customs of that community. In my terms, if an agent chooses, then it must choose as a regulatory agent, since to be an agent is to be regulatory. That agent is responsible for its actions, however irrational they may be by the standards of Normal rationality for that type of agent

My thesis is on the metaphysics of agency, based on the use of system and so my use of choice and responsibility is that of the system description of the agent who acts. Had my thesis been on the metaphysics of moral philosophy, based on the use of system then my use of choice and responsibility would be that of the community and would have an equal claim to truth. This is not to deny the truth of our profound intuition of personal responsibility for our lives. Over the relatively short times over which we consider their actions, agents are normally unchanged with respect to their decision making properties and, by and large; the agent to whom the community attributes personal continuity and responsibility, and the agent responsible are sufficiently alike for our intuitions to be reliable. Were they not, these customs and concepts would not have been viable and would not have evolved within our communities.

Many writers have felt that to accept a deterministic account of our actions is to rob us of the dignity which is inherent in our freedom for our own actions and our ability to rationally choose. Also, that to accept this is to deny the right of members of a community to hold themselves and others responsible for the consequences of their actions. What follows must be an account of my own intuitions but I believe that the evidence from evolutionary biology and neurology that is in so far supports them.

I find dignity in the appreciation that each of us, as we are now, will be responsible for any actions that we are about to make and for all of their consequences. These actions may result in consequences that will change each future self and for each of us; the agent of the here and now, is responsible for these. None of us is without responsibility for what he or she is today. Many actions of many people in the past and many accidents of history have contributed in the historical process that has led to the current person attributed in the here and now. This process must be determinate as it has determined 'what' each person is, although the numerical identity of each person is given by their continuity of structure and components, and naming and recognition is within the customs of the community. What one is determines what one does. In Walt Disney's immortal words "I am what I am and that's what I am." (I leave it to the reader to complete the quotation).

To say 'the actions of an agent are determined by what the agent is' is not to say that the actions are rational. I have learned rational techniques for predicting the likely consequences of my own and others actions and sometimes I use these. However, I have learned that my actual actions are chosen within a wider rationale. In this sense they are also rational in that each action could in principle be predicted from a complete knowledge of my states, including my perceptions of recent events. Each of us has been shaped by our own histories. In this sense what we are is determined and in this sense each of our actions is determined. However, to be an autonomous agent is to maintain an internal autonomy against external events that are epistemically contingent.

Each of our actions is determined by our agency but only a selection of the states of this agency is available to the narrative recall of consciousness. I believe that the distinction between the conscious self and the acting agent is at the root of the profound intuitions that many writers have expressed regarding spontaneity and personal responsibility. In conscious decision making one is aware of the conflict of difficult choice and one will engage in lateral thinking to develop a novel solution. In poetry and in music, solutions and ideas 'come into consciousness' but within the limits of the 'entry' metaphor, each has come from somewhere. Each of us is a unique combination of character and disposition with a unique history of exposure to events including ideas. Each of our choices is a unique event, uniquely determined by our complex history but and this is what I think is found difficult, it is still our choice made by us.

We are all free in so far as we are not physically constrained but each of us is constrained by our location, by our physical abilities, mental capabilities and by the combination of character and dispositions that make each of us what we are now. This combination of constraints is what one is in the here and now. To say that it determines one's actions in response to events, which require choice does not reduce the freedom to choose. Strawson summarised the apparent paradox of objective freedom in the first page of the second chapter of his book:

> "Surely we cannot be free agents, in the ordinary, strong, true–responsibility-entailing sense, if determinism is true and our actions are ultimately wholly determined by "causes anterior to [our] personal existence"[18]? And surely we can no more be free if determinism is false and it is, ultimately, either wholly or partly a matter of chance[19] or random outcome that we and our actions are as they are?"

[18] The quoted clause is from H. Sidgwick, The Methods of Ethics, p. 66.

[19] The current constitution of each of us is due to the accidents of our experience and inheritance, but this is not the 'chance' that Strawson is concerned to oppose.

This apparent dilemma is a consequence of his definition of freedom. If we reject the assumption that freedom must lie in the acceptance of true responsibility, where true responsibility is defined in terms of 'desert' and accept the contention that our actions are "wholly determined by causes anterior to [our] personal existence." Then we can accept the first horn of the dilemma since it is clearly not a matter of chance that our actions and we are as they are. Our actions are determined by the events that have shaped us and these are anterior to our 'personal existence' where the community attributes personhood to the ongoing agent. Strawson has denied himself this solution by his insistence that true responsibility must entail desert, giving freedom, responsibility and desert equal ontological status. In my alternative thesis, agent responsibility is determinate and is independent of desert, personal responsibility is attributed within the customs of the community and is bound up with the attribution of desert.

With this sense of 'freedom' we can only act as free agents if we are not constrained and as agents, we act as the agents that we are. We can only be the agents that we are if what we are, is determined prior to our action. Thus our actions are wholly determined by the events to which they respond and by the agents that we are and each of these are causes anterior to our *personal* existence. Secondly, it is partly a matter of chance that our actions are as they are since the fact that events require choice by us implies that they are random for us and the way in which we are derives from the accidents of history that have shaped us. Neither of these statements implies that we are wrong to feel free to choose, since our conscious feeling of freedom is a component of conscious choice. Or that we are wrong to attribute responsibility since our concepts of normative responsibility are woven into us as members of our community designated as 'persons'.

My theory of agency is not anthropocentric. My theory of human action is of a complex agent for which Normal action is that chosen in virtue of the representations and combinatorial processes existent in that animal at the time of that action. Humans are very complex agents and they live in communities with language but it is as agents that they act. Because of their complex integrated agency and language, they can articulate the conscious processes of their choice and they call it thought. Because of this, they feel different from other less complex natural agents without language and their theories of conscious action have emphasised this difference. I do not feel belittled by the appreciation that the decisions which I make are made for me by my states that I call my reasons just as my cat goes out at night for her reasons. My dignity for me lies in my autonomy and I hope that for the cat it is the same.

7

IMPLICATIONS

If we believe in a world solely comprised of some physical stuff, causally closed and causally symmetric in unidirectional time then it follows that we, who are in this world, are of the same stuff. This is the thesis of naturalism. If we believe that the phenomena of this world, including the phenomena of our perception, are constrained by universal physical laws that describe our expectations from similar events then this thesis implies that the way we are is constrained by these same laws. Not to believe in these implications is to believe that we are in some way set aside from the world that we are in. If we also believe that this world is all that there is, then we can only explain how we have become what we are, in terms of a causal history of accidental events. To ask why we are what we are, is not a sensible question unless we believe in some purpose transcendental to this world. This is not to say that there is no such purpose; the challenge faced by the thesis of naturalism is to assume nothing that is not explicable within a framework of physical cause.

We observe our world and ourselves within it and as observers we can comprehend what it is to be an observer. A theory that explains the physical context of observation is the theory of agency. We cannot prove this theory, we can only test it by observing phenomena in the world, experimenting with artefacts and considering its plausibility. Does it fit within our framework of physical law? Does it account for the phenomena of biology and the other life sciences? If we accept this theory then we have an austere account of man in nature. Accidents of physics led to the formation of this universe, which has included this planet circling this sun. Accidents of physics, describable within the science of chemistry, led to the phenomena of life: apt for system description in terms of agency, reproduction and autopoiesis. Accidents of genetic evolution, describable within the science of biochemistry, have led to the complexities of ecology and cognitive process that we see in our current taxonomy of life. Further accidents of genetic and social evolution have led to the development of human linguistic communities and the practices of communication, observation and description. Within these practices we speak of purpose but these are the purposes of our daily lives. Because of the ubiquity of purpose in our descriptions of our daily lives our intuition is to look for purpose in our aetiology. As naturalists, we test this intuition and fail to justify it.

The signs that denote things in our world are given conventional meaning for us by our practices of use to refer to such things. By our agency we classify our world for us. Any thing that we name, we name because it has natural meaning for us through the classifications of our agency. The conventional meaning of the symbol for us is the relation between our perceptions-of-agency of the symbol and our perceptions-of-agency of the thing. As agents we do not name, we act. Our Normal cognitive processes of agency cause us to act in ways that are for our goals in the light of our representations. There is then a relation between the content of our representations and the content of the world represented such that the natural meaning of the world that is represented is the natural meaning of the representation and this relation can be thought of as truth. But this thought is in language: as agents we do not think, we act. Conventional truth and conventional meaning are concepts in a metalanguage, only available after semantic ascent. Natural truth and natural meaning are concepts within a theory of agency but this theory can only be articulated in a metalanguage that includes the denotation of propositions. It is in this metalanguage that truth and meaning can be defined.

If the conventional meaning of a symbol perceived corresponds with the natural meaning of a thing perceived then we speak of the symbol as true. By our theories of our world we generalise and distinguish between realist and anti-realist conceptions of truth. To be a realist about a concept is to hold that application of this concept tracks a state of the world classified for us by our agency or derived from such classifications by theories that we hold true. Our intrinsic representations within agency are real states, realised in the neural states that are physically caused by our perceptions of real physical states in the world. That an attribution of belief is a state of a linguistic community does not make it any less real. Its realisation is within the conventional articulations of the community and these articulations are actions that have been caused by perception of agents in the world by agents in the community. Such attributions will be reliable in so far as the perceptions of the community and their theory of one another's goals are reliable. That such practices of attribution and the attribution conditions for belief have survived implies that they are appropriate for the survival of the community.

We are in the world and we are of the world. We are also observers of the world that includes ourselves. The accidents of history that have led to our existence as active, sentient and articulate observers have enabled each of us to recognise the responsibility of our own unique agency. Through the self and articulately shared, consciousness of our linguistic communities, we accept responsibility for the personal development of others and ourselves. That we have occurred by accident and that our choices are to do with the contingent in our lives does not imply that we lack choice or personal responsibility for our futures. That a psychopath has been created by the actions of others is a fact of history that places responsibility on those others who, as agents, were responsible for their actions. But this does not change the current pathology or change the current physical and personal responsibility of that agent for his or her actions. We cannot be solely responsible for what we are; we cannot be responsible for the contingent events that make us choose; but, who we are is what we are and we are responsible for what we do

The range of complexity of realisation between the austere agency of a bacterium, plant organ or thermostatic control and the rich learning agency of a normal living human is vast. Also, the advent of socialisation and language has led to complex emergent phenomena such as the attribution of personhood and belief and our political and economic institutions. My task has been to show that these phenomena, though complex beyond theoretical prediction, are not mysterious.

The agent at the instant, continually changing in its realisation of intrinsic representations and cognitive process and the person recognised as an ongoing entity within the social grouping of agents, is the same physical entity. The same intrinsic representations within agency that are causally maintained by perception, of world states, speech and text, and that cause action by each human agent are components in the holisms of perception that are recognised by each person of themselves and others and cause attributions of responsibility and belief.

Although so far, none of our artefacts compare with us in the rich complexity of our integrated agency, the principle difference between them and us is of function: not 'function of' but 'function for'. We organic agents owe the rich variety of our agency to our evolutionary history. Although we can speak of our organic structure as realising the functional requirements of perception, representation, cognitive processes and action, our function is for ourselves. The primary goal of any life form is set by its autopoiesis. However, since it survives in a contingent world a basic goal within the theory of its agency is to survive. Other goals of complex organisms emerge from the evolution of complexity of classification of the world for them. Our goals have evolved biologically and socially and they have been shaped by our own on-going agency but they are ours. Our artefacts, however complex and to whatever extent their agency or autopoiesis is delegated, owe their goals to us and their initial function is for us. It may be that any such artefact could cease to have function for us and through an artificially created autopoiesis, have function for itself but its function would still stem from our agency.

APPENDIX A

AUTOPIESIS, ALLOPOIESIS, AETIOLOGY AND AGENCY: CRITERIA FOR LIFE

A1.1 The Concept of Autopoiesis

Humberto Maturana and Francisco Varela created the concept of autopoiesis in the early 1970s to identify those system characteristics that distinguish the living from the non-living. Since then, autopoiesis and its applications has become a discipline in its own right as various writers, including its creators, have considered the application of the concept to problem areas as diverse as law, family therapy and social organisation.[1] In developing this concept, Maturana and Varela brought together four fundamental observations about the nature of living systems:

> Firstly, their autonomy. Although any living thing may depend on another for its survival, each living thing is a separate entity, bounded in space and self-defined.
>
> Secondly, life is a physical phenomenon. Every token phenomenon can be explained in physical causal terms.
>
> Thirdly, the phenomena of life are explained within descriptions made by observers who are themselves living entities, external to the phenomena that they describe. Observers can perceive an entity and its environment and enter into relations of perception with each. The physical components of an entity can only relate to other components.

[1] A summary of autopoiesis and its applications can be found in Mingers, (1995)

> Fourthly, since they are rooted in physical cause, the phenomena of life are, in the terminology of 2.6, above, accidental. Explanations of life should have no recourse to ideas of teleological[2] function or purpose. The observable phenomena of living systems result purely from the causal interactions of causally contiguous components. Any explanation of these interactions in terms of function is by an observer who can interact with the component and the whole.

In these terms they concluded that; what makes an entity living is that it is organised in such a way that all its components and processes jointly produce those self-same components and processes, thus establishing an autonomous self-producing entity. This autonomous self-production of autonomy is what is meant by *autopoiesis*. The structure of living entities is such that they produce and maintain themselves. Entities that are active but do not produce themselves are *allopoietic*, that is, 'other producing'. Cars, robots and factories are allopoietic. They change the world and they produce things but what is produced is other than them. Autopoietic entities are self-producing, allopoietic entities are other producing. Cars or factories, however automated or regulatory, may produce services such as transportation or things that are used elsewhere but do not produce themselves. In the terminology of this thesis, an entity can have agency without being autopoietic.

A1.2 Criteria for Autopoiesis

Maturana, (1980), defined an autopoietic system as:

> A dynamic system that is defined as a composite unity as a network of production of components that,

[2] Maturana and Varela speak of teleonomy to emphasise the instrumental nature of purpose attributed by an observer.

a) through their interactions recursively regenerate the network of productions that produced them, and
b) realise this network as a unity in the space in which they exist by constituting and specifying its boundaries as surfaces of cleavage from the background through their preferential interactions within the network.

Varela et al., (1974), gave a six point key for the identification of an autopoietic system which can be summarised as follows:

(i) Has the entity an identifiable boundary?
(ii) Can the entity be described as constituted by components?
(iii) Is the entity apt for description as a mechanism or structure of mechanisms?
(iv) Do the components that constitute the boundaries do so in virtue of their own mechanisms?
(v) Are the components of the boundary produced by the mechanisms of the entity either of itself or by transformation of substance imported through the boundary?
(vi) If questions (i) through (v) are answered positively and all other components of the entity are either produced by the interaction of components as in (v) or are permanent constitutive components that partake in the production of other components then you have an autopoietic entity in the space in which its components exist.

The first three criteria require that the entity be apt for system description as a physical mechanism. The remaining three questions set out the central notion of autopoiesis; the notion of an autonomous entity that, in virtue of its own organisation, maintains not only its own components and boundary but also maintains that organisation. In Maturana's terminology, a living entity is structurally determined and organisationally closed. Structural changes maintain the organisation that defines the identity of the living entity. In the terminology of this thesis, living things are apt for description, within a closed system, as autopoietic in that they are characterised by a circular organisation of processes that continually produce and replace the components and boundary necessary to that organisation

A1.3 The Place of the Observer

> "Everything said is said by an observer. ... The observer is a living system and an understanding of cognition as a biological phenomenon must account for the observer and his role in it."

It was in these terms that Maturana commenced his essay on the "Biology of Cognition"[3] that set the stage for his co-development of the concept of autopoiesis. Life and its manifestations do not depend on their being observed. But the description of life and its explanation does depend on the evolution of living things that, through the complexity of their cognitive processes and the consensual interactions of language can describe and explain those processes that constitute their own living organisation.

A1.4 Autonomy and Function

In their description of autopoiesis, Maturana and Varela eschew the notions of function and purpose.

> Autopoiesis specifies certain necessary condition and if these arise then an autopoietic unity is established ... There is no need for functionalist explanations or teleonomic ideas such as purpose in the explanation of living things. (Mingers, 1993, p. 38)

[3] Part 1 of (Maturana and Varela, 1980)

> [I]f living systems are physical autopoietic machines, teleonomy becomes only an artifice of their description, which does not reveal any feature of their organisation, but which reveals the consistency of their operation within the domain of observation. Living systems as physical autopoietic machines are purposeless. (Maturana and Varela, 1980, p. 86)

Maturana and Varela distinguish not between 'function' and 'realisation' but between 'organisation' and 'structure'.

> [Organisation] refers to the relations between components that define and specify an entity as apt for description as a system of a particular class and determine its properties as an entity.
>
> [Structure] refers to the actual components, their relations and how these determine the space taken by the entity but the structure does not determine the properties of the entity.

This appears close to the distinction between function and realisation given in 2.9 above, wherein some capacity of an entity is explained by reference to some component of that entity, the presence of which enables that capacity, and the function of that component can be realised in alternative physical structures. But this appears to oppose the very definition of autopoiesis. There are two distinct notions of function in play here. The teleonomic notion of *function* that Maturana and Varela are concerned to oppose is of function *for*, where *for* implies some use of the entity in some application and *function* of this kind would negate the autonomy of autopoiesis. A more modest use of function as enabling the essentially self-serving capacity of autopoiesis does not threaten the autonomy of autopoiesis and this is the argument that I will pursue.

A1.5 The Capacity of Autopoiesis

Interpreting the above within the methodology of this thesis: the capacity of autopoiesis can be described within a containing system that has no explanatory need for inclusion of an environment and in this sense, the system is closed. The system will include the boundary of the entity and those components so structured that they comprise a mechanism with processes that maintain the boundary, the components and the self-maintaining structure. This does not mean that the entity is isolated from its environment. All of these processes will dissipate energy and the entity is apt for functional description under another system description, open in the Bertalanffy sense, which includes mechanisms with the functions of import of energy and excretion of waste.

Also, living things are plastic in that their structure can grow or change in ways appropriate to their ongoing autopoiesis. Maturana described those processes by which the realisation of autopoiesis in an entity is either enabled or modified by environmental perturbations as *structural coupling* and the dynamic outcome of such coupling as *ontogenetic structural drift.*

This mode of description of an entity is essentially self-centred. An entity is either autopoietic or it is not and with respect to this question, the only interaction an autopoietic entity has with its environment are those that trigger its birth, ontogenetic change or death. Within this description of autopoiesis, an organism does not act, its autopoiesis is functionally autonomous and it is blind to the environment due to which it survives, grows or dies.

Agency, as I have defined it, is concerned with action in the world and an entity is characterised as having agency in so far as it manifests the *capacity* for action in the world that is *for* it. Autopoiesis as defined by Maturana and Varela is concerned with self maintenance in the world and an entity is characterised as being autopoietic in so far as it manifests the *capacity* to maintain its own organisation by its own structure. From this analogy, it seems reasonable to apply the methodology of functional explanation to the mechanisms of autopoiesis.

From Varela's six point check list an entity is autopoietic if:

> (i) The entity has physical boundaries and these boundaries have the function of maintaining organisational closure whilst facilitating structural coupling. (ii), (ii) and (iv) There are components of the entity and these have function within a system description of the entity as mechanism. Any physical causal explanation of such a mechanism will explain how specific physical interactions realise functional components of the mechanism. (v) & (vi) If the components of the boundaries of the entity and other components are produced by interaction between components of the entity and transformations of imported substance then we have the functions of production, import and transformation.

In these terms, we have a physical boundary with a function and this function is realised by specific causal relations between components of the entity. We have a functional organisation that enables the capacity of autopoiesis. This is realised in the structure of components that comprise mechanisms and these mechanisms realise the functions of importation, transformation, production and excretion that enable the capacity of autopoiesis.

A1.6 The Complementarity of Autopoiesis and Agency

Autopoiesis describes a living entity from the inside. The function of self-maintenance may be partially realised by homeostatic mechanisms and these satisfy the criteria for agency but such functional components are not autopoietic if they are maintained within the autonomous organism. Maturana's concept of structural coupling is of an autopoietic entity achieving a dynamic balance with its environment, but the real physical world is continually changing relative to each entity within it. For autopoiesis to persist in a dynamic world any structural coupling must continually adapt to the changes of that world. Autopoiesis does not require agency but an autopoietic entity will only survive in so far as the dynamic adaptation of its structural coupling can maintain those variables essential to its autopoiesis within their essential limits against contingent events in the environment, and this requires agency.

Autopoiesis is a necessary criterion for life and agency is a necessary criterion for the ongoing maintenance of autopoiesis in an entropic world. The explanations of agency place an entity in an environment and require functions that relate to the furtherance of the goals of its agency within the environment. The explanations of autopoiesis place an environment around an entity and require functions that relate to the self-maintenance of the entity.

Autopoiesis requires *production* and an organisation that gives purpose[4] to this production. Also, that this purpose be the maintenance of that organisation. Production requires the *import* of material, the *metabolism* of that material and the *export* of waste

Agency requires *action* that changes states of the world, *perception* of states of the world, states of the agent that *represent* states of the world. Also, an *organisation* such that perception of the world by the agent leads to action that changes the world in ways that are for the goals of the agent, in particular, its *goal* of remaining autopoietic.

Agency does not logically imply autopoiesis since the function of acting for a given purpose against the contingencies of a changing world is independent of

[4] Note that this is not the teleological notion of final purpose.

the function of self-maintenance. Autopoiesis does not logically imply agency since its functional specification as production to maintain the organisation that produces, does not mention the environment of risk and resource.

However, any physical entity will exist contingently in the context of an unpredictably changing world and the ongoing maintenance of such existence by the processes of autopoiesis must depend on agency. In so far as states of the world that are relevant to autopoiesis are predictable, in the sense that the organisation of the entity can be described as if it were programmed for them, then autopoiesis does not require agency. In so far as the world is unpredictable for the entity then continued successful autopoiesis requires agency

A1.7 Teleology and Aetiology

If we are concerned to explain the behaviour of an entity in the world then we may, in the terms of Dennett, (1987) choose a stance for this explanation. We may choose the intentional stance and recognise that the entity is apt for description in terms of agency. We may choose the autonomous survival stance and recognise that the entity is apt for description in terms of autopoiesis. We may choose the historical stance and recognise that the aetiology of the entity is apt for description in terms of evolution. If the entity passes our test for the autonomy of autopoiesis but fails on agency, although it survives in a changing world, then we must conclude that the resources for its autopoiesis are not contingent on its relationship to the world. If the entity passes our test for agency but is not autopoietic then its purposes do not derive from the purpose of maintaining its structure. Either; its goals are inconsistent with its ongoing autopoiesis and it will not survive against the contingencies of the environment, or it is an artefact and its agency depends on other agency that requires explanation. If the entity passes our tests for agency and autopoiesis but has not naturally evolved then; either its occurrence is an unexplained accident or its purposes are real in the sense contended by Richard Taylor. In this, latter case, its agency and autopoiesis are realised within artefacts of other agency that require explanation.

Real states of representation and real processes of perception, cognition and action can be explained in terms of their function or purpose but the nearest that we can get to real purpose without agency is the self maintaining purpose of autopoiesis. Because of the ubiquity of attribution of purpose in our daily lives we look for purpose within our selves but our purposes derive from our history. Our best current explanation of this derivation is our theory of evolution as a sequence of accidents by which, living things and their reproduction effect and are effected by, one another and their environment. I conclude that necessary and sufficient conditions for the attribution of natural life to an entity are

(i) that its autonomy be apt for explanation within the closed system description of autopoiesis and

(ii) that its behaviour be apt for explanation within the open system description of agency and

(iii) that its aetiology be apt for description in terms of the accidents of evolution.

If to be autonomous is not to depend on other agency then autonomous life in a physical domain is natural life since it has occurred naturally.

APPENDIX B

REQUISITE VARIETY

In his treatment of cybernetics, Ashby, (1956), develops a methodology of system description in which states of mechanisms are described in tabular terms. He defines a determinate machine as; "... that which behaves in the same way as a closed single valued transformation." (p. 24). "When a real machine and a transformation are so related, the transformation is the **canonical representation** of the machine and the machine is said to **embody** the transformation." (p. 29). Variety is defined with respect to a set and generally, the set will be the range of identifiable states of some machine. Any given machine will, by its mechanisms, constrain[1] the range of values possible for its states. With respect to a location within a machine the variety of its possible states is the number of distinguishable elements within the set of values that describe these states. In these terms regulation can be described as a transformation that reduces the flow of variety of some variables, relevant to an entity, from an environment within which disturbances to these variables arise. Looking ahead; in the diagram on page 251, D is the environment and E is the entity, T is the mechanism by which the disturbances reach the entity and R is a regulatory[2] mechanism that effects T so as to reduce the flow of variety from D to E.

What Ashby proves, within a metaphor of combinatorial games, is that; only variety in R can block the variety from D. In a slogan; only variety can destroy variety. In the terminology of communication theory, R's capacity as a regulator

[1] Within this terminology, the laws of physics are a constraint on the states possible for a machine.

[2] Regulation is not just error control. Hiding when stalking prey or the use of camouflage, are examples of regulation.

cannot exceed R's capacity as channel of communication. Or, rephrased into the metaphor of combinatorial games;

> In a formal system that describes a class of events and outcomes available to an instigator D and a respondent R, where R is controlling against variation due to D, the variety of outcome in R cannot be less than the variety of D divided by the variety of R.

Ashby demonstrated this law by a series of simple decision tables.[3] These can be read either as input tables for an elementary game or as an illustration of a principle of regulation. As a game, if the convention of the game is that an item in a table is either good or bad for R and R must choose a response to each move of D then R will establish some strategy, expressible as a decision table, to maximise his benefit from the game.

As a first simple example, consider Table 1, below. If r is attempting to score an *a* then, if D plays 1 then R will play β, if D plays 2 then R plays α and if D plays 3 then R plays γ. If R plays according to the strategy:

1	2	3
β	α	γ

He can always force the outcome a. In fact R has complete control.

Table 1

		R		
		α	β	γ
	1	*B*	*a*	*c*
D	**2**	*A*	*c*	*b*
	3	*C*	*b*	*a*

[3] Examples are adapted from Ashby, (1956, 202-206).

Other situations may be less favourable. For example, in table 2, if R wants a or d then he can always chose α, if R wants a and d does not play 5 then R need only play γ. But, if R wants b and D plays 2 or 3 then R loses. These simple example begin to show that, in any strategic situation with pay off for a respondent, the numerical relationship between the variety of incident states and the variety of responses available will be significant.

Table 2

		R			
		α	β	γ	δ
	1	*b*	*d*	*a*	*a*
	2	*a*	*d*	*a*	*d*
D	**3**	*d*	*a*	*a*	*a*
	4	*d*	*b*	*a*	*b*
	5	*d*	*a*	*b*	*d*

However, tables that require a real response from R are those with no repetitions within columns. In such tables, any change by D requires a response from R. For example, in Table 3, since no column contains repetition, any change by D requires a change by R.

Table 3

		R		
		α	β	γ
	1	f	f	k
	2	k	e	f
	3	m	k	a
	4	b	b	b
D	**5**	c	q	e
	6	h	h	m
	7	j	d	d
	8	a	p	j
	9	l	n	h

R must react 9 times. Since at each decision he has to do something, to choose the letter again is just as much a decision as to change and if letters are repeated in a row he still has to choose. Suppose he chooses the strategy shown, with its outcome in transformation in Table 4

Table 4

1	2	3	4	5	6	7	8	9
α	ß	γ	ß	α	ß	γ	α	γ
f	*e*	*a*	*b*	*c*	*h*	*d*	*a*	*h*

However we interpret R's moves, they represent three unique alternatives. They have a variety of 3. Similarly, D's moves have a variety of 9. The outcomes in the example have a variety of 7. Our interest here is in how small the outcome variety can be and we can show that; for any table, constructed as Table 3, the outcome variety cannot be less than

$$\frac{\text{D's variety}}{\text{R's variety}}$$

This can be readily demonstrated from Table 3. Suppose that at each move in response to a move by D, R chooses the column, which will minimise the increase in variety. Since this move will introduce a new letter he must change column, which may introduce another letter. At best he will have to return to the first column after three moves which must increase the variety and by the time D's repertoire is exhausted, at least two (in this case) changes will have been made. In the example of Table 5, below, it can be seen that R has an obvious strategy which will produce an output in which only *a*, *d*, and *g* occur.

Table 5

		R		
		α	β	γ
	1	a	b	h
	2	b	a	i
	3	c	b	a
	4	d	c	b
D	5	e	d	c
	6	f	e	d
	7	g	f	e
	8	h	g	f
	9	i	h	g

If the elements of Table 3 are interpreted as possible states of the world, which include possible states of an agent, then choices by D can be interpreted as actual states of the world and choices by R can be seen as possible responses by the agent. It must be remembered that the agent is in the world and is of the world. The interpretation of this result in this context is that unpredictability of outcome for an agent against a world which varies in ways that are unpredictable by the agent can only be reduced, for the agent, in so far as the responses available to the agent as regulator are of a variety which, at least, matches the variety of challenges in the environment for the agent. Only variety in R can force down the variety in D. In so far as the challenges posed by the environment are unpredictable, they can be seen as information for the agent and the same result can be derived in information theoretic terms[4].

[4] Rothstein, (1951), equates physical information with negative entropy. But the entropy need not be a matter of thermodynamic states. Any occurrent contingent state can be informative and this information may be measured by the relative probabilities of its alternatives.

If D, R and E are three variable, such that each can be interpreted as an information source; that is, data regarding hitherto unknown states of the world, possibly causally independent. Then, if all lie within some stable probability distribution, the entropies (in Shannon's usage of the term), can be stated: H(D,R,E) is the entropy of the vector with the three as components, H(E|D) is the uncertainty in E when D is known, H(R|E,D) is the uncertainty in R when E and D are known and so on. The previous criterion, that no element shall occur twice in a column, is interpreted as the criterion that the entropy of the outcome 'E' is not to be less than that of D, that is,

$$H(E|R) \geq H(D|R)$$

From this it can be readily shown, from the multiplication axiom for general events,

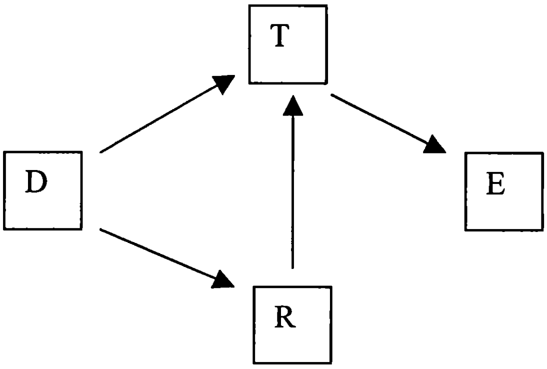
$$P(A,B) = p(A)p(B|A) = p(B)p(A|B),$$

that

$$H(E) \geq H(D) + H(R|D) - H(R).$$

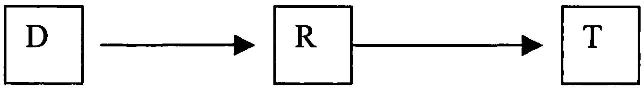
The implication of this is that the entropy of the outcome has a certain minimum and that this is least when H(R|D) is zero, that is, when R is a determinate function of D and then, H(E) is H(D) - H(R). The minimum value of the outcome entropy can only be reduced below that of the disturbance by an equal increase in that of the regulator

These results can be interpreted in terms of communication.



In the above diagram, D can be interpreted as a source of disturbance or events that are epistemically contingent, for the entity R. T can be interpreted as the state of the world common to D and R and E as the outcome of states of the entity. In terms of variety, if R does nothing then variety in D can pass straight through to E. It may not, since T may be such as to block change (an organism may not need to seek shelter if it is behind a natural barrier) but generally, suppression of variety transmission to E is blocked by variety in R.

Alternatively, R can be seen as a transmitter



And an alternative formulation of the Law of Requisite Variety is that *R's capability, as a regulator cannot exceed R's capacity as a channel of communication.* This is equivalent to Shannon's theorem that the amount of noise that can be removed from a correction channel is limited to the amount of information that can be carried by that channel.

All organisms are in the business of regulation. This may be the homeostasis of the control of internal variables or heterotelic action in the environment fot the goals of agency. Each organism that survives by its regulation does so in virtue of a variety of types of representational state that at least matches the variety of states of the world that may threaten the organism

This theorem applies in time as well as space, organisms may evolve to utilise auto correlation and cross correlation in their predicting. Selection pressures may favour more rapid response to changing circumstance. They may favour integrated processing of several changing phenomena. One result of Ashby's gloss on Shannon's theorem is that rate of change and variety of what changes are two aspects of the same criterion. Variety in both in time and space offers an ecological niche and organisms in that niche will survive that possess the regulatory variety to match it.

GLOSSARY

Accident: If for some given event no causal predecessor can be found by a theorist, then the incidence of such an event is effectively random for that theorist; It is an *accident*.

Action: The causal process by which representational states of agency intentionally change states of the world, including the agent, for the goals of the agent.

Agency: The capacity of an entity in virtue of which the entity acts to maintain or attempt to achieve its goals in a dynamic and unpredictable environment.

Allometry: The differential growth rate of parts of the body between members of a group.

Allopoietic Machine: A machine which has as its product something different from itself, as a car or a university.

Autonomy: Independent of others and having function for itself; surviving as an organism independently of other organisms or parts.

Autopoietic Machine: A machine that has as its product; its boundary, its structure and the components necessary for its own continuation.

Cause: Two successive, directly or ancestrally, physically contiguous states of affairs are apt for description as causally linked iff true descriptions of each are available that are within the same domain of physical description. And within this domain, they and their intermediate causal ancestors, are connected by physical laws such that the second is counterfactually dependent on the first.

Closed system: A bounded entity is organisationally closed if its organisation does not causally depend on its environment. It is materially closed if there is no import or export of substance across the boundary

Cognition: Cognitive processes are those causal processes by which perception changes representation, representations change representations and representations cause actions so as to enable the agency of an entity

Concept: (i) Within the usage of the philosophy of language; possessing a concept is knowing what it is for something to be its semantic value. Concepts are abstracta, realised in the practices of language. Representations with conceptual content are realised within sentences of language spoken by people to people about things in the world.

(ii) Within the usage of ethology; concepts may be attributed to a organism if their attribution enables the explanation of complex abilities to generalise over variable stimuli, to rapidly produce appropriate responses to the common features underlying these stimuli and to modify behaviour when it is discovered that perceptual stimuli are unreliable guides to underlying features.

Content: The content of a representation is the state of affairs that it represents. Content can be conceptual, as in the representations of language or non-conceptual as in intrinsic representations within agency. The non-conceptual content of intrinsic representations may be described within the concepts of a theory.

Contiguity: Two states of affairs may be described as physically contiguous if they are in the same domain of physical description and spatiotemporally located such that one may imply the other within a physical law.

Cybernetics: The field of cybernetics is the field of control theory and those aspects of communication theory that are relevant to control.

Directive Correlation: The correlation between events in the world and the actions of an intentional agent in a goal oriented response to such events. To be distinguished from the statistical correlation of events which may or may not have a causal explanation.

Entropy: (i) The tendency to disorder in closed physical systems. (ii) A statistical measure of the disorder in a closed system, given in terms of the disjunctive information latent within the system.

Function: To ascribe a function to something is to ascribe a capacity to it that is singled out by its role in an explanatory analysis of some system. When a capacity of an entity within a system is appropriately explained by analysing it into a number of other capacities, the analysing capacities also emerge as functions. A function is a relation in that it is of an item and for an item, although an item may have function for itself (see Autonomy)

Goal: The goals of agency are a cut in the set of possible worlds picked out by a system description of an agent in the world. In this system an agent may be described as pursuing its goals against events in the world that are epistemically contingent for it.

Heterotelis: Regulatory mechanisms between an entity and its environment by which the entity pursues its goals in the environment.

Homeostasis: Regulatory mechanisms within an organism by which it maintains its autopoiesis.

Information: (i) A disjunctive usage connected with the probabilities of alternative states of affairs that may be used as a measure of the variety of states of the world that are epistemically contingent for an agent,

(ii) A conjunctive usage connected with the combination of propositions by a linguistic agent as premises of inference.

Intentionality: A state of an entity is intentional if it is about aspects of the world that are relevant for the entity. A description of causal change is a description of intentional action if and only if the change can be explained as intended by an agent and caused by an intended change of that agent.

Machine: A bounded physical entity with capacities that are functionally enabled by mechanisms.

Meaning: A concept within a metalanguage that pertains to the relation between the conventional denotation of a term within the object language and the state of affairs denoted.

Mechanism, (i) Noun: a nexus of physical cause, seen as a holism in an explanatory context.
Mechanism, (ii) Principle: the view that every biological event is a combination of physical causation.

Metalanguage: A language in which terms denote items of an object language.

Observer: Anything that is described is described by an observer. The observer is an agent, in the world and of the world that contains the agency that the observer describes. Although, within language, the observer can recursively describe what it is to be an observer describing the agency of an observer, there is no transcendental view point from which such description can be made.

Open System: A bounded entity is organisationally open if its organisation causally depends on its environment. It is materially open if there is import or export of substance across the boundary. Organisms and artefacts of regulation may be organisationally closed and materially open. The dynamics of such systems may include plasticity in which material transactions lead to organisational change in the longer term. Entities in such systems may be organisationally closed with respect to short-term dynamics but open in the longer term.

Perception: The causal processes by which states of the world change representational states of agency.

Representation. One item represents another when, in some context or by some convention and with some function, that item stands in for another. Representations may be classified into three types.
Type I. Symbols, in which the power of representation derives from conventions of definition and conventions of interpretation. The function of symbolic representation is for the members of the community whose conventions define its use and interpretation.
Type II. Icons and natural signs, in which the power of representation derives from their natural or artificial resemblance or correlation, but their interpretation is by convention. The function of such representation is for the interpreter.
Type III. Intrinsic representation in which the representational power derives from processes of the entity for which the representation has cognitive function. Such entities may be biological or artificial though if artificial, representations within the entity will have function for the entity and for its maker.

System: A mode of description, based on a theory, comprised of statements and conditionals, that is concerned with the possible outcomes implied by physical entities and contingent events in an actual world.

Teleology: A mode of explanation in which current actions of agency are for the current goals of agency.

Text: Any token linguistic transaction that exists subsequent to its initial production and to which reference can be made. Text may be realised in the marks of textual media or the memories of speakers..

Theory: Causal relations cited between features of a system description of an actual physical world constitute a theory. This theory is true if the system is a true description of an actual world. System theories may be at the level of physical causal mechanisms or they may be at the level of system capacities and functional components that enable those capacities.

Truth: A concept within a metalanguage that pertains to the relation between the conventional use of an assertive sentence within the object language and the state of affairs asserted. The statement, P is T iff p, where 'P' is the sentence and p is the state of affairs, implies two relations: one between a true usage of a sentence and a state of affairs and the other between a state of affairs and a putative true usage of the sentence. To express these relations we require a meta-metalanguage.

Variety: the number of states in an environment required to be distinguished by a regulator.

BIBLIOGRAPHY

Abercrombie M, Hickman C.J. & Johnson M.L. (1980), The Penguin Dictionary of Biology, Penguin Books.

Akins K. (1993), What is it like to be boring and myopic? in: "Dennett and his Critics", ed Dahlbom B., Blackwell, Oxford, 1993.

Alberts B., Bray D., Lewis J., Raff M., Roberts K., & Watson J.D., (1989), The Biology of the Cell, Garland Pub., New York.

Allen C.& Hauser M., (1991), Concept attribution in nonhuman animals, in: "Readings in Animal Cognition", ed. M. Bekoff and D. Jamieson, Bradford Books, MIT, Cambridge, 1996.

Anscombe G.E.M. (1979), Under a Description, Nous, XIII, 219.

Anscombe G.E.M. (1957), Intention, Blackwell, Oxford.

Aristotle (-340), The Generation of Animals, Trans. The Loeb Classical Library, Heinemann, 1953.

Ashby, W.R. (1952), Design for a Brain, Wiley, New York.

Ashby, W.R. (1956), An Introduction to Cybernetics, Chapman and Hall, London.

Asimov I. (1982), The Complete Robot, Harper Collins, London, 1993.

Baker L.R. (1987), Saving Belief, Princeton Univ. Press, New Jersey.

Baker L.R. (1995), Explaining Attitudes, Cambridge University Press.

Bar-Hillel Y. (1964), Language and Information, Addison Wesley.

Bateson G. (1972), Steps to an Ecology of Mind, Paladin, New York.

Bateson G. (1979), Mind and Nature, Wildwood House.

Beaken M. (1996), The Making of Language, Edinburgh University Press.

Bechtel W. & Abrahamson A. (1991), Connectionism and the Mind, Basil Blackwell, Oxford.

Bermudez J.L. ((1995), Nonconceptual content: From Perceptual experience to Subpersonal Computational States, Mind and Language, 10, 334-369.

Bell J.S. (1987), On the Einstein-Podolsky-Rosen Paradox, in: "Speakable and Unspeakable in Quantum mechanics", Cambridge University Press.

Bertalanffy L. von (1950), The theory of open systems in: physics and biology, in: "Systems Thinking", ed. F.E. Emery, Penguin Modern Management Readings 1969.

Bertalanffy L. von (1968), General System Theory, Braziller, New York.

Block N. (1986), Advertisement for a semantics for psychology, in: "Mental Representation", ed. P. Stich & T.A. Warfield, Blackwell, Cambridge, Mass., 1994.

Boulding K.E. (1956), General Systems Theory - The skeleton of science, Management Science, 2, 197-208.

Brentano F. (1874), Psychology from an Empirical Standpoint, trans, Rancurello et al., Routledge & Kegan Paul, London, 1973.

Brillouin L. (1949), "Life, thermodynamics and cybernetics", in: Maxwell's Demon, ed. H.S. Leff and A.F. Rex, Adam Hilger, Bristol, 1990.

Broad C.D. (1929), The Mind and its Place in Nature, Kegan Paul, Trench, Trubner.

Brown G.W. (1956), Monte Carlo methods, in: "Modern Mathematics for the engineer", McGraw Hill, 1956.

Cameron W.A. (1960), Unpublished report to CEGB on the use of random numbers in scheduling the refuelling machines at Berkeley Nuclear Power Station.

Cameron W.A. (1995), Systems for philosophy, in "Critical Issues in Systems Theory and Practice", ed: K. Ellis, A. Gregory, B.R. Mears Young & G. Ragsdell, Plenum Press, London, 1995.

Cannon W. (1932), Wisdom of the Body, Norton, New York.

Carruthers P. & Smith P.K., (1996), Theories of Theories of Mind, Cambridge University Press, Cambridge.

Cartwright N. (1983), How the Laws of Physics Lie, Clarendon Press, Oxford.

Checkland P. (1981), Systems Thinking, Systems Practice, John Wiley & Sons, Chichester.

Church A. (1956), Introduction to Mathematical Logic, Princeton University Press.

Churchland P.M. (1981), Eliminative materialism and the propositional attitudes, in: "Mind and Cognition" ed. W.G. Lycan, Blackwell, Cambridge, 1990.

Corning P.A. (1995), Synergy and Self-Organisation in the Evolution of Complex Systems, Systems Research, 12, 2, 89-122.

Crane T. (1992), The nonconceptual content of experience, in: "The Contents of Experience", ed. T Crane, Cambridge University Press, Cambridge.

Cummins R. (1975), Functional analysis, Jnl. of Phil., 72, 741-764.

Cummins R. (1989), Interpretational semantics, in: "Mental Representation", ed. P. Stich & T.A. Warfield, Blackwell, Cambridge, Mass., 1994.

Cussins A. (1990), The connectionist construction of concepts, in: "The Philosophy of Artificial Intelligence", ed M Boden, O.U.P. Oxford, 1990.

Cussins A. (1992), Content, Embodiment and Objectivity; The Theory of Cognitive Trails, Mind, 101, 651-688.

Davidson D. (1963), Actions, reasons, and causes, in: "Essays on Actions & Events", D. Davidson, 1980, Clarendon Press, Oxford.

Davidson D. (1969), The individuation of events, in: "Essays on Actions & Events", D. Davidson, 1980, Clarendon Press, Oxford.

Davidson D. (1970), Mental events, in: "Essays on Actions & Events", D. Davidson, 1980, Clarendon Press, Oxford

Davidson D. (1971), Agency, in: "Essays on Actions & Events", D. Davidson, 1980, Clarendon Press, Oxford.

Davidson D. (1982), Rational Animals, Dialectica, 36, 318-327

Dennett D.C. (1976), Conditions of personhood, in: Dennett (1981),

Dennett D.C. (1981), Brainstorms, Bradford Books, Cambridge.

Dennett D.C. (1983), Intentional systems in cognitive ethology: the Panglossian paradigm defended, in: "The Intentional Stance", Bradford Books, MIT Press.

Dennett D.C. (1982/3), Styles of mental representation, in: "The Intentional Stance", Bradford Books, MIT Press, 1989.

Dennett D.C. (1986), Content and Consciousness, Routledge and Kegan Paul.

Dennett D.C. (1987), The Intentional Stance, Bradford Books, MIT Press.

Dennett D.C. (1991), Consciousness Explained, The Penguin Press, England.

Dretske F. (1981), Knowledge and the Flow of Information, Basil Blackwell, Oxford.

Dretske F. (1986), Misrepresentation, in: "Mental Representation", ed. P. Stich & T.A. Warfield, Blackwell, Cambridge, Mass., 1994.

Dretske F. (1988), Explaining Behaviour, Bradford Books, MIT, Cambridge.

Dretske F. (1993), Conscious Experience, Mind, 102, 263-283.

Dunbar R. (1996), Grooming, Gossip and the Evolution of Language, Faber and Faber.

Eden C. (1989), Using cognitive mapping for strategic options development and analysis, in: "Rational Analysis for a Problematic World", ed. J. Rosenhead, J. Wiley, 1989.

Edelman G. (1992), Bright Air, Brilliant Fire, Penguin Books, London.

Evans G. (1982), The Varieties of Reference, Clarendon Press, Oxford.

Falk A.E. (1981), Purpose, Feedback and Evolution, Philosophy of Science, vol. 48, pp. 198-217.

Festinger L. (1954), A Theory of Cognitive Dissonance, Row, Peterson, Evanston, Illinois.

Field H. (1972), Tarski's theory of truth, in: "Reference, Truth and Reality", ed., M. Platts, Routledge & Kegan Paul.

Fish S. (1980), Is There a Text in This Class, Harvard University Press, Cambridge.

Fodor J.A. (1975), The Language of Thought, Crowell, New York.

Fodor J.A. (1980), Methodological solipsism considered as a research strategy in cognitive psychology, The Behavioural and Brain Sciences, 3, 63-109.

Fodor J.A. (1983), The Modularity of Mind, Bradford Books, MIT Press.

Fodor J. A. (1985), Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-mecum, Mind, 94, 76-100.

Fodor J.A. (1987), Psychosemantics, Bradford Books, MIT, Cambridge.

Fodor J.A. (1990), A theory of content, II: the theory, in: "Mental Representation", ed. P. Stich & T.A. Warfield, Blackwell, Cambridge, Mass., 1994.

Fodor J.A. and Pylyshin Z., Connectionism and Cognitive Architecture: a Critical Analysis, Cognition, 28, 139-96.

Foester H. von (1982), Observing Systems, Intersystems Publications, Seaside, CA.

Forrester J.W. (1961), Industrial Dynamics, Cambridge Mass., The MIT Press.

Frisch E. von (1964), Biology, Harper & Row, New York.

Fuenmayor R. (1993), Systems science, addressing global Issues - The death rattle of a dying era?, in: "Systems Science, Addressing Global Issues", eds. F. A. Stowell et al., Plenum Press, London, 1993.

Gallie W.B. (1957), Uncertainty as a philosophical problem. In: "Uncertainty and Business Decisions", eds. C.F. Carter, G.P. Meredith & G.L.S. Shackle, Liverpool University Press, 1957.

Galton A. (1990), The Church-Turing Thesis: its nature and status, AISB Quarterly, Autumn 1990, No74, 9-19.

Gödel K. (1931), On Formally Undecidable Propositions, trans. Basic Books, New York, 1962.

Goldman A. I. (1970), A Theory of Human Action, Prentice-hall, New Jersey.

Grice H.P. (1957), Meaning, in: "Philosophical Logic", ed., P.F. Strawson, Oxford Readings in Philosophy, 1967.

Griffin D.R. (1981), The Question of Animal Awareness, Kaufman, Los Altos, California.

Halliday M.A.K. (1994), The construction of knowledge and value in the grammar of scientific discourse, with reference to Charles Darwin's “The Origin of Species”, in: "Advances in Written Text Analysis", ed. M. Coulthard, Routledge, New York, 1994.

Harnad S. (1994), Computation is Just Interpretable Symbol Manipulation; Cognition Isn't, Minds and Machines, 4, 379-390.

Harnad S. (1996), The origin of words: a psychophysical hypothesis, in: "Communicating Meaning: Evolution and Development of Language", eds. B. Velichkovsky & D. Rumbaugh, Erlbaum, New Jersey.

Hart J.W. (1990), Plant Tropisms and Other Growth Movements, Unwin Hyman, London.

Hill T.E. (1971), The Concept of Meaning, Humanities Press.

Hofstadter D.R. (1979), Gödel, Escher, Bach, an Eternal golden Braid, Harvester Press.

Holbach Baron, (1817), The System of Nature, trans. Mde Mirabeau, Sherwood, Neely and Jones.

Horgan T, (1993), From Supervenience to Superdupervenience: Meeting the Demands of a Rational World, Mind, 102, 555-586.

Huggins W.H. (1948), A note on frequency transformations for use with an electrolytic tank. Proc. I.R.E., 36, 421.

Humphreys P. (1989), The Chances of Explanation, Princeton University Press, New Jersey.

Husserl E. (1931), Cartesian meditations, trans. Cairns D. and Nijhoff M., The Hague, 1977.

Jackson M.C. (1993), Systems practitioners: facilitators of the bringing – forth of social systems, in: "Systems Science, Addressing Global Issues", (eds) F. A. Stowell et al., Plenum Press, London, 1993.

Jantsch E. (1980), The Self Organising Universe, Pergamon Press, Oxford.

Kamis G. (1991), Self Modifying Systems in Biology and Cognitive Science, Pergamon, Oxford.

Kellner H. (1989), Language and Historical Representation, Univ. of Wisconsin.

Kent W. (1978), Data and Reality, North Holland.

Kim J. (1973), Causes and counterfactuals, in: "Causation", eds. E. Sosa & M.Tooley, Oxford University Press, 1993.

Kim J. (1970), Causation, nomic subsumption and the concept of event, in: “Supervenience and Mind”, J. Kim, Cambridge University Press, 1993.

Kim J (1993), Supervenience and Mind, Cambridge University Press.

Klir G.J. (1969), An Approach to General Systems Theory, Van Nostrand Reinhold, New York.

Koehler (1938), The Place of Values in the World of Fact, Liverright.

Koestler A. (1967), The Ghost in the Machine, Hutchinson.

Kripke S. 1963, Semantical Analysis of Modal Logic, I, Zeitstschrift Für mathematische Logik und Grundlagen der Mathematik, 9, 67-96.

Lande A. (1958), The Case for Indeterminism, in: "Determinism and Freedom", ed., S. Hook, New York University Press, 1958.

Laszlo E. (1972a), Introduction to Systems Philosophy, Gordon and Beech, New York.

Laszlo E. (1972b), The Systems View of the World, Basil Blackwell, Oxford.

Lawrence E. (1989), A Guide to Modern Biology, Longman, New York.

LePore E., and McLaughlin B. L. (1985), Actions, reasons, causes, and intentions, in: "Actions and Events", eds. E. LePore & B.L. McLaughlin, Blackwell, Oxford.

Lewis D. (1973), Counterfactuals, Blackwell, Oxford.

Lewis D. (1986), On the Plurality of Worlds, Blackwell,Oxford.

Lewis, D. (1993), Causation, in: "Causation", eds. E.Sosa & M.Tooley, Oxford University Press, 1993.

Locke J. (1690), An essay concerning human understanding, in "Classics of Western Philosophy", ed. Cahn S. M., Hackett, Indianapolis.

Lockery S. (1989), Representation, functionalism and simple living systems, in: "Goals, No-Goals, and Own Goals", eds. A. Montefiore & D. Noble, Unwin Hyman, London.

Lyons W. (1990), Intentionality and Modern Philosophical Psychology, 1, Philosophical Psychology, 3, 247-269.

Lyons W. (1991), Intentionality and Modern Philosophical Psychology, 2, Philosophical Psychology, 4, 83.

Lyons W. (1992), Intentionality and Modern Philosophical Psychology, 3, Philosophical Psychology, 5, 309.

Lyons W. (1995), Approaches to Intentionality, Clarendon Press, Oxford.

Mackie J.L. (1965), Causes and Conditions, American Philosophical Quarterly, 2, 245-264.

Mackie J.L. (1974), The Cement of the Universe, Clarendon Press, Oxford.

Malcolm N. (1968), The Conceivability of Mechanism, Phil. Rev., 77, 1968, 45-72

Manier E. (1971), Functionalism and the Negative Feedback model in Biology, Boston Studies in the Philosophy of Science, 8, 225-240.

Marr D. (1982), Vision, Freeman, San Fransisco.

Maturana H.R. & Varela F.J. (1973), Autopoiesis: The organisation of the living, in: "Autopoiesis and Cognition", 63-134, D. Reidel, Dortrecht, 1980.

Maturana H.R. (1980), Man and society, in: "Autopoietic Systems in the Social Sciences", ed. F. Benseler, P. Hejl and W. Kock, Campus Verlag, Frankfurt, pp. 11-31.

Maturana H.R. (1981), Autopoiesis, in: "Autopoiesis: A Theory of Living Organisations", ed: Zeleny M., Elsevier-North Holland, New York, 21-33.

Maturana H.R. (1987), The biological foundations of self consciousness and the physical domain of existence, in: Caianiello E. ed., "Physics of Cognitive Processes", World Scientific, Singapore, 324-379.

Mayr O. (1970), The Origins of Feedback Control, The MIT Press, Cambridge.

McCall C. (1990), Concepts of Person, Avebury, Aldershot, UK.

McDowell J. (1994), Mind and World, Harvard, Cambridge.

McFarland D.J. (1971), Feedback mechanisms in animal behaviour, Academic Press, London.

McFarland D.J. (1989), Goals, no-goals and own goals, in "Goals, No-Goals, and Own Goals", A. Montefiore & D. Noble, Unwin Hyman, London.

McGinn, C. (1989), Mental Content, Basil Blackwell, Oxford.

Midgley, G. (1994), Five Sketches of Post-modernism: Implications for Systems Thinking and Operational Research, Systemist, 16, 187.

Millikan R.G. (1984), Language, Thought, and Other Biological Categories, MIT Press, Cambridge.

Millikan R.G. (1993), White Queen Psychology and Other Essays for Alice, The MIT Press, Cambridge.

Mingers J. (1993), An Examination of Information and Meaning, Sytemist, Vol 15 Feb 1993, 17.

Mingers J. (1995), Self-Producing Systems, Plenum Press, London.

Monod J. (1971), Chance and Necessity, Alfred A Knopf, New York.

Nagel E. (1961), The Structure of Science, Harcourt, Brace & World, New York.

Nagel E. (1979), Teleology Revisited and Other Essays, Columbia, New York, 1979.

Nagel T. (1974), What is it like to be a bat?, Philosophical Review, 83, 435.

Nagel T. (1986), The View from Nowhere, Oxford University Press, New York.

Neander K. (1995), Misrepresenting and Malfunctioning, Philosophical Studies, 79, 109-141.

Newton Sir Isaac (1731), A Treatise of the System of the World, Dawsons, London, 1969.

Olson D.R. (1988), On the origins of beliefs and other intentional states in children, in: “Developing Theories of Mind” ed. J.W. Astington, P.L. Harris and D.R. Olson, Cambridge University Press, New York.

Oparin A.I. (1956), The Origin of Life on the Earth, trans. Ann Synge, Oliver and Boyd, 1957.

Parfit D. (1984), Reasons and Persons, Oxford University Press.

Peacocke C. (1989), Perceptual content, in: "Themes from Kaplan" ed J. Almog, J. Perry & H. Wettstein. Oxford University Press, New York.

Peacocke C. (1992), A Study of Concepts, Bradford Books, Cambridge.

Peacocke C. (1994), Nonconceptual Content: Kinds, Rationales and Relations, Mind and Language, 9, 419-29.

Perner J. (1993), Understanding the Representational Mind, Bradford Books, MIT, Cambridge.

Perry J. (1979), The Problem of the Essential Indexical, Nous, 13, 5.

Plato (-330), The Epinomis, Trans. The Loeb Classical Library, Heinemann, 1953.

Powers, W.T., (1973), Behavior: The Control of Perception, Aldine, Chicago.

Putnam H. (1960), Minds and machines, in "Dimensions of Mind: a symposium", ed: S.J.Hook, New York University Press, New York.

Putnam H. (1973), Meaning and reference, The Journal of Philosophy, 70, 699-711.

Putnam H. (1973b), Philosophy and Our mental life, in "Modern Philosophy of Mind" ed W. Lyons, 1995, Everyman, London.

Putnam H. (1988), Much Ado about Not Very Much, Daedalus, 117, 269-281.

Putnam H. (1992), Why functionalism failed, in "Inference, Explanation and Other Philosophical Frustrations" ed: J. Earman, University of California Press, Berkeley, 1992.

Quine W.V. (1960), Word and Object, MIT Press, Cambridge.

Ramsey W, !997), Do Connectionist Representations earn their Explanatory Keep?, Mind, 12, 1, 34-66.

Rosen R. (1991), Life Itself, Columbia University Press, New York.

Rosenblatt F. (1962), The Principles of Neurodynamics, Spartan, New York.

Rosenblueth A. and Wiener N. (1950), Purposeful and non purposeful behaviour, Philosophy of Science, 17, 318-326.

Rosenblueth A., Wiener N. and Bigelow J. (1943), Behaviour, Purpose and Teleology, Philosophy of Science, 10, 18-24.

Rothstein J. (1951), "Information, Measurement and Quantum Mechanics", in: Maxwell's Demon, ed. Leff H.S. and Rex A.F., Adam Hilger, Bristol, 1990.

Rumbaugh S.S. and Lewin R, Kanzi, (1994), The Ape at the Brink of the Human Mind, Doubleday, London.

Russell B. (1919), Descriptions, Ch XVI of "Introduction To Mathematical Philosophy", George Allen and Unwin, London.

Ryle G. (1949), The Concept of Mind, Hutchinson.

Ryle G. (1979), On Thinking, Basil Blackwell, Oxford.

Sartre J.P. (1943), Being and Nothingness, trans. H. E. Barnes, Arrowsmith, Bristol, 1958.

Sayre K.M. (1976), Cybernetics and the Philosophy of Mind, Routledge & Kegan Paul, London.

Schrödinger, (1944), What is Life, Cambridge University Press, 1967.

Seyfarth R. and Cheney D. (1992), Inside the mind of a monkey, in: "Readings in Animal Cognition", ed. M. Bekoff and D. Jamieson, Bradford Books, MIT, Cambridge, 1996.

Shannon C.E. (1948), A Mathematical Theory of Communication, Bell System Technical Journal, 30, 379-423 and 623-656.

Shannon C.E. and Weaver W. (1949), The Mathematical Theory of Information, University of Illinois Press.

Sharda R. (1994), Neural networks for the MS/OR analyst: an applications bibliography, Interfaces, 24, 2, 116-130.

Sidgwick H. (1930), The Methods of Ethics, Macmillan, London.

Smuts J.C. (1926), Holism and Evolution, Macmillan, London.

Sommerhoff G. (1950), Analytical Biology, Oxford University Press, Lomdon.

Sommerhoff G. (1969), The abstract character of living systems, in: "Systems Thinking", ed. F.E.Emery, Penguin Books, Mddx.

Spaul M.W.J. (1993), Critical systems thinking, post modernism and the philosophy of Richard Rorty, in: "Systems Science, Addressing Global Issues", (eds) F. A. Stowell et al., Plenum Press, London, 1993.

Stewart E. and Cohen J., 1997, Figments of Reality, Cambridge University Press, Cambridge.

Stalnaker R (1968), A theory of conditionals, in: "Conditionals" ed. F, Jackson, 1991, Oxford University Press, Oxford.

Stich S.P. and Warfield T.A. ed. (1994), Mental Representations, A Reader, Blackkwell, Cambridge.

Stowell F.A., West D. & Howell J.G. (1993), Systems Science, Adressing Global Issues., Plenum Press, London.

Strawson G. (1986), Freedom and Belief, Clarendon Press, Oxford.

Tacket A. and White L. (1993), After OR: an agenda for postmoderism and poststructuralism in OR, J. Opl. Res. Soc., 44, 867-882.

Tarski A. (1944), The Semantic Conception of Truth and the Foundations of Semantics, Philosophy and Phenomenological Research, 4, 351.

Taylor R. (1950), Comments on a Mechanistic Conception of Purposefulness, Philosophy of Science, 17, 310-317.

Taylor R. (1966), Action and Purpose, Prentice-Hall, New Jersey.

Turing A. (1937), On Computable Numbers with an Application to the Entscheidungs Problem, Proc. London Math. Soc., x/ii, 1937, pp. 230-265.

Valberg J.J. (1992), The puzzle of experience, in: The Contents of Experience, ed. T. Crane, Cambridge Univ. Press, Cambridge.s

van Gelder T. (1995), What might cognition be, if not computation., Journal of Philosophy, 15, 345-381.

van Inwagen P. (1975), The Incompatibility of Free Will and Determinism, Philosophical Studies, 27, 185-99.

van Inwagen P. (1983), An Essay on Free Will, Clarendon Press, Oxford.

Varela F.J., Maturana H.R. and Uribe R. (1974), Autopoiesis: The Organisation of Living Systems, Its Characterisation and a Model, BioSystems, 5, 187-196.

Varela F.J. (1979), Principles of Biological Autonomy, North Holland, New York.

Walsh D.M. and Ariew A. (1996), A Taxonomy of Functions", Canadian Journal of Philosophy, 26.4, 493-514.

Westphal L.C. (1995), Source Book of Control Systems in Engineering, Chapman & Hall, London.

Whitehead A.N. (1969), Process and Reality, Free Press, New York.

Wiener N. (1961), Cybernetics, Cambridge, Mass

Wilkes K.V. (1978), Physicalism, Routledge & Kegan Paul, London.

Williams B. (1973), Problems of the Self, Cambridge University Press,

Williams T. et al. (1995), The Effect of Design Changes and Delays on Project Costs, Journal of the Operational Research Society, 46, 809-818.

Wilson E.O.(1992), The Diversity of Life, Penguin Books, 1994.

Wimsatt W. C. (1971), Some Problems with the Concept of 'Feedback', Boston Studies in the Philosophy of Science, 8, 241-256.

Wittgenstein L., Philosophical Investigations, trans. G.E.M. Anscombe, (1945), Blackwell, Oxford, 1989.

Wolstenholme E.F. (1993), The Changing Role of System Dynamics, in: "Systems Science, Addressing Global Issues", (eds) F. A. Stowell et al., Plenum Press, London, 1993.

Wright L. (1973), Functions, Phil. Rev., 82, 139-166.

GLOSSARY

Accident: If for some given event no causal predecessor can be found by a theorist, then the incidence of such an event is effectively random for that theorist; It is an *accident.*

Action: The causal process by which representational states of agency intentionally change states of the world, including the agent, for the goals of the agent.

Agency: The capacity of an entity in virtue of which the entity acts to maintain or attempt to achieve its goals in a dynamic and unpredictable environment.

Allometry: The differential growth rate of parts of the body between members of a group.

Allopoietic Machine: A machine which has as its product something different from itself, as a car or a university.

Autonomy: Independent of others and having function for itself; surviving as an organism independently of other organisms or parts.

Autopoietic Machine: A machine that has as its product; its boundary, its structure and the components necessary for its own continuation.

Cause: Two successive, directly or ancestrally, physically contiguous states of affairs are apt for description as causally linked iff true descriptions of each are available that are within the same domain of physical description and they and their intermediate causal ancestors are connected by physical laws within that domain such that the second is counterfactually dependent on the first.

Closed system: A bounded entity is organisationally closed if its organisation does not causally depend on its environment. It is materially closed if there is no import or export of substance across the boundary

Cognition: Cognitive processes are those causal processes by which perception changes representation, representations change representations and representations cause actions so as to enable the agency of an entity

Concept: (i) Within the usage of the philosophy of language; possessing a concept is knowing what it is for something to be its semantic value. Concepts are abstracta, realised in the practices of language. Representations with conceptual content are realised within sentences of language spoken by people to people about things in the world.

(ii) Within the usage of ethology; concepts may be attributed to a organism if their attribution enables the explanation of complex abilities to generalise over variable stimuli, to rapidly produce appropriate responses to the common features underlying these stimuli and to modify behaviour when it is discovered that perceptual stimuli are unreliable guides to underlying features.

Content: The content of a representation is the state of affairs that it represents. Content can be conceptual, as in the representations of language or non-conceptual as in intrinsic representations within agency. The non-conceptual content of intrinsic representations may be described within the concepts of a theory.

Contiguity: Two states of affairs may be described as physically contiguous if they are in the same domain of physical description and spatiotemporally located such that one may imply the other within a physical law.

Cybernetics: The field of cybernetics is the field of control theory and those aspects of communication theory that are relevant to control.

Directive Correlation: The correlation between events in the world and the actions of an intentional agent in a goal oriented response to such events. To be distinguished from the statistical correlation of events which may or may not have a causal explanation.

Entropy: (i) The tendency to disorder in closed physical systems. (ii) A statistical measure of the disorder in a closed system in terms of the disjunctive information latent within the system.

Function: To ascribe a function to something is to ascribe a capacity to it which is singled out by its role in an explanatory analysis of some system. When a capacity of an entity within a system is appropriately explained by analysing it into a number of other capacities, the analysing capacities also emerge as functions. A function is a relation in that it is of an item and for an item, although an item may have function for itself (see Autonomous)

Goal: The goals of agency are a cut in the set of possible worlds picked out by a system description of an agent in the world. In this system an agent may be described as pursuing its goals against events in the world that are epistemically contingent for it.

Heterotelis: Regulatory mechanisms between an entity and its environment by which the entity pursues its goals.

Homeostasis: Regulatory mechanisms within an organism by which it maintains its autopoiesis.

Information: (i) A disjunctive usage connected with the probabilities of alternative states of affairs that may be used as a measure of the variety of states of the world that are epistemically contingent for an agent,

(ii) A conjunctive usage connected with the combination of propositions by a linguistic agent as premises of inference.

Intentionality: A state of an entity is intentional if it is about aspects of the world that are relevant for the entity. A description of causal change is a description of intentional action if and only if the change can be explained as intended by an agent and caused by an intended change of that agent.

Machine: A bounded physical entity with capacities that are functionally enabled by mechanisms.

Meaning: A concept within a metalanguage that pertains to the relation between the conventional denotation of a term within the object language and the state of affairs denoted.

Mechanism, (i) Noun: a nexus of physical cause, seen as a holism in an explanatory context.
Mechanism, (ii) Principle: the view that every biological event is a combination of physical causation.

Metalanguage: A language in which terms denote items of an object language.

Observer: Anything that is described is described by an observer. The observer is an agent, in the world and of the world that contains the agency that the observer describes. Although, within language, the observer can recursively describe what it is to be an observer describing the agency of an observer, there is no transcendental view point from which such description can be made.

Open System: A bounded entity is organisationally open if its organisation causally depends on its environment. It is materially open if there is import or export of substance across the boundary. Organisms and artefacts of regulation may be organisationally closed and materially open. The dynamics of such systems may include plasticity in which material transactions lead to organisational change in the longer term. Entities in such systems may be organisationally closed with respect to short-term dynamics but open in the longer term.

Perception: The causal processes by which states of the world change representational states of agency.

Representation. One item represents another when, in some context or by some convention and with some function, that item stands in for another. Representations may be classified into three types.
Type I. Symbols, in which the power of representation derives from conventions of definition and conventions of interpretation. The function of symbolic representation is for the members of the community whose conventions define its use and interpretation.
Type II. Icons and natural signs in which the power of representation derives from their natural or artificial resemblance or correlation but their interpretation is by convention. The function of such representation is for the interpreter.
Type III. Intrinsic representation in which the representational power derives from processes of the entity for which the representation has cognitive function. Such entities may be biological or artificial though if artificial, representations within the entity will have function for the entity and for its maker.

System: A mode of description, based on a theory, comprised of statements and conditionals, that is concerned with the possible outcomes implied by physical entities and contingent events in an actual world.

Teleology: A mode of explanation in which current actions of agency are for the current goals of agency.

Text: Any token linguistic transaction that exists subsequent to its initial production and to which reference can be made. Text may be realised in the marks of textual media or the memories of speakers..

Theory: Causal relations cited between features of a system description of an actual physical world constitute a theory. This theory is true if the system is a true description of an actual world. System theories may be at the level of physical causal mechanisms or they may be at the level of system capacities and functional components that enable those capacities.

Truth: A concept within a metalanguage that pertains to the relation between the conventional use of an assertive sentence within the object language and the state of affairs asserted. The statement, P is T iff p, where ‘P’ is the sentence and p is the state of affairs, implies two relations: one between a true usage of a sentence and a state of affairs and the other between a state of affairs and a putative true usage of the sentence. To express these relations we would require a meta-metalanguage.

Variety: the number of states in an environment required to be distinguished by a regulator.