City Research Online

City St George's, University of London

Citation: Whobrey, D. (1999). Aspects of Qualitative Consciousness: A Computer Science Perspective. (Unpublished Doctoral thesis, City, University of London)

Aspects
of
Qualitative Consciousness:
A Computer Science Perspective

Darren J.R. Whobrey

A Thesis Submitted in Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy

Department of Computer Science
City University
Northampton Square, London

May 1999
(Revised November 1999)

Table of Contents

List of Tables

List of Figures

Acknowledgments

The author wishes to thank Allen Long and James Fetzer for many useful discussions and suggestions on improving the logical structure of the thesis. Igor Aleksander for helpful suggestions. Peter Osmon, for allowing the author the freedom to explore the field of Science. Cassandra Millhouse, for considerable motivation and understanding. Special thanks to my family for their untiring support and patience. Finally, thanks to friends for their humour.

This research was supported in part by a SERC grant.

Contact Address:

Flat 4, 27 Willow Road

Hampstead, London NW3 1TL

Tele: 0171-435-0382.

Email: darren@blueberry.netkonect.co.uk

What we seek we shall find;
What we flee from flees from us.
Emerson, "Fate," The Conduct of Life, 1860.

Declaration

The author grants power of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to the author. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Statement of Contribution

This disclaimer is to state that the research reported in this thesis is primarily the work of the author and was undertaken as part of the doctoral research.

Abstract

Aspects of Qualitative Consciousness:
A Computer Science Perspective

The domain of artificial intelligence (AI) has been characterised by John Searle [Sear84] by distinguishing between *weak* AI, according to which computers are useful tools for studying mind, and *strong* AI, according to which an equivalence is made between mind and programs such that computers executing programs actually possess minds. This dissertation explores a third alternative, namely: the prospects and promise of *mild* AI, according to which a suitable computer is capable of possessing species of mentality that may differ from or be weaker than ordinary human mentality, but qualify as "mentality" nonetheless. The purpose of this dissertation is to explore the prospects and promise of mild AI.

The approach adopted explores whether mind can be replicated, as opposed to merely simulated, in digital machines. This requires a definition of mind in order to judge success. James Fetzer [Fetz90] has suggested minds can be defined as sign using systems in the sense of Charles Peirce's semiotic (theory of signs) and, on this basis, argues convincingly against strong AI. Determining if his negative conclusion applies to mild AI requires rejoining Fetzer's analysis of the analogical argument for strong AI and redressing his laws of human beings and digital machines. This is tackled by focusing on the nature and form of the operational relationship between the physical machine and mind, and suggesting some operational requirements for a minimal semiotic system independently of any underlying physical implementation. This involves four steps.

Firstly, as a formal foundation, a characterisation of systems is developed in terms of the causal structure and ontological levels in the system, where an ontological level is individuated by the laws that are in effect. This is in contrast to levels of organisation, such as levels of software abstraction. This exploration suggests the necessity – as a matter of natural law – for a mediating level between the physical machine and mind that is or, at least, appears to be necessary for producing forms of mentality. The lawful structure that appears to be required within this level and between levels is examined with respect to the prospects for implementing a semiotic system.

Secondly, how a system can operate in terms of semiotic processes based on a network of instantiated dispositions is explored. These are modelled as the temporal counterparts of state-transitions and stationary-representations, which are termed causal-flows and temporal-representations, respectively. They highlight the varying interactive structure of temporal patterns of causal activity in time. For the purposes of replicating mind, preserving the causal-flow structure of mental processes arises as an important requirement.

Thirdly, the system structure sufficient for generating consciousness is explored – a necessary condition for a cognitive semiotic system. This suggests a requirement relating to the causal accessibility of the contents of consciousness. This structuring is driven by the system's need to signify reality by categorising these aspects as operational entities upon which decisions can be made. Consciousness arises through the manner in which the signified reality is generated. This makes mind and consciousness the result of a co-ordinated occurrent system wide activity.

Fourthly, in a mathematical sense, brains and computers can be classified as types of numeric and symbolic systems, respectively. These systems are compared and conditions formulated under which they may give rise to equivalent ontological levels. Peirce's triadic sign relation is analysed in terms of ontological levels and the results used to clarify the nature of the ground relation in machine forms of mentality.

According to the theorems developed, the introduction of a dispositional mediating level might effectively enable a suitable computer to replicate species of mentality. An important factor in determining whether a computer is suitable for this purpose is its performance capacity and thus some estimates are calculated in this respect. It is shown how these requirements, along with a number of others, can help in the development of semiotic systems and variants, such as the iconic state machine of Igor Aleksander [Alek96].

Chapter 1

Introduction

1 Overview

The domain of artificial intelligence (AI) has been characterised by John Searle [Sear84] by distinguishing between *weak* AI, according to which computers are useful tools for studying mind, and *strong* AI, according to which an equivalence is made between mind and programs such that computers executing programs actually possess mind. This dissertation explores a third alternative, namely: the prospects and promise of mild AI, according to which a suitable computer is capable of possessing species of mentality that may differ from and be weaker than ordinary human mentality, but qualify as "mentality" nonetheless. The purpose of this dissertation is to explore the prospects and promise of mild AI.

The approach adopted will be to explore whether mind can be replicated, as opposed to merely simulated, in digital machines. James Fetzer [Fetz90, p17] defines *replication* as "effecting the right functions by means of the very same – or similar – processes," in contrast to *simulation* as "effecting the right functions from inputs to outputs" – this is discussed further in chapter six.

Determining whether mind can be replicated in digital machines requires a definition of mind in order to judge success. Fetzer has suggested minds can be defined as sign using systems in the sense of Charles Peirce's semiotic (theory of signs), but argues convincingly against strong AI. Consequently, to determine whether this applies to mild AI requires rejoining Fetzer's analysis of the analogical argument for strong AI and redressing his laws of human beings and digital machines.

To refine the applicability of Fetzer's argument the operational requirements for a minimal semiotic system are examined independently of any underlying physical implementation. This involves two steps. Firstly, exploring how a system can operate in terms of semiotic processes underpinned by a network of instantiated dispositions. This suggests an important requirement based on the temporal causal-structure of processes. Secondly, exploring the system structure sufficient for generating consciousness – a necessary condition for a cognitive semiotic system. This suggests a necessary requirement relating to the causal accessibility of the contents of consciousness. It is shown how these requirements, along with a number of others, can help in the development of semiotic systems and variants, such as the iconic state machine of Igor Aleksander [Alek96].

As a formal foundation, a characterisation of systems is developed in terms of the causal structure and ontological levels in the system, where an ontological level is individuated by the laws that are in effect. This is in contrast to levels of organisation, such as levels of software abstraction – see Herbert Simon [Sim96]. In a mathematical sense, brains and computers can be classified as types of numeric and symbolic systems, respectively. Consequently, these two types of system are compared and conditions formulated under which they might give rise to equivalent ontological levels.

The analysis of Peirce's semiotic in terms of dispositions and ontological levels, carried out for the purposes of determining the operational requirements for implementing a semiotic system, reveals that a mediating dispositional level is necessary between mind and brain. From this, and in conjunction with the ontological and causal characterisation of systems, the case looks promising for mild AI and for the possibility that a suitably programmed digital machine could replicate species of mentality. However, it is also found that this is currently infeasible due to the physical performance limitations of digital machines.

1.1 Logical Development

A review is presented of the main definitions of AI. The motivation for the thesis is stated as the desire to explore whether mild AI can succeed. A literature review shows that an answer to this is problematical. The strategy taken here to address this is presented and centres on adopting a definition of mind as a semiotic system and exploring how such a system can be replicated on a suitable digital machine. The conventions and methodology adhered to are subsequently reviewed. Finally, a summary of the logical contribution of each chapter is presented.

2 Defining AI

This section presents an analysis of the main definitions of AI currently in use in relation to determining whether or not the goals of mild AI are achievable.

2.1 Introduction to Systems

The concept of a system is used extensively in what follows so a brief description is now presented, a more in-depth analysis is presented in chapter six.

A system is said to be some bounded thing that has internal structure. A formal system is an uninterpreted symbolic system characterised by a set of descriptive sentences in a language, and a set of rules that govern how it evolves. Here, a symbol is defined to be that which can be used to designate an arbitrary expression. Formal systems are mere abstractions and are related to their real counterparts via an interpretation. A dynamical system is a system that changes state in time. A digital computer is a dynamic symbolic machine for automating the manipulation of formal systems.

2.2 Origin of AI

A conference organised by John McCarthy in 1956, called "The Dartmouth Summer Research Project on Artificial Intelligence", is acknowledged as one of the foundational events in the history of AI and from which the field got its name. Marvin Minsky attended the conference, and later had this to say [Mins68]: "AI is the science of making machines do things that would require intelligence if done by men." Several years before the conference Alan Turing [Tur50] published a paper exploring machine intelligence and whether machines could think. Turing suggested the issue could be settled without having to solve the mystery of consciousness. This was in response to a talk by a colleague, Geoffrey Jefferson, who proposed that not until a machine has "feelings" (i.e. consciousness) could machines be equated with brains. John von Neumann [Neu56] was also having doubts as to the applicability of the equivalence between digital machines and brains, and suggested that the brain could be operating in an analogue manner as well. Unfortunately, these and similar thoughts from Frank Rosenblatt [Rse62] were soon swept aside by the more popular symbolic approach as championed by Allen Newell and Herbert Simon [Nwl61], and Marvin Minsky and Seymour Papert [Mins69]. Thus, even in the early days there was debate concerning the prospects of mild AI.

2.3 Definitional Dimensions of AI

Stuart Russell and Peter Norvig [Russ95] classify definitions of AI along four dimensions:

1) Human versus ideal performance. Humans perform tasks in a certain way and brains use particular mechanisms. These may not be the most ideal, optimal or only way of achieving the end result. AI research at the human end of the dimension seeks to explore approaches that have a semblance to human thinking and behaviour. Research at the ideal end is driven by the desire to find (normative) solutions that need not adhere to the human way of doing things.
2) Thought versus behaviour. This ranges from focusing on internal mental processes and reasoning to emphasis on observable action and behaviour.
3) Theoretical versus practical. Whether the interest is in theoretical results or practical applications.
4) Conscious or not. Does the system accommodate consciousness in some capacity?

Based on the first two dimensions AI can be organised into four main categories: systems that either think or act, either like humans or rationally. The following subsections describe these categories in more detail.

2.4 AI and Intelligence: Systems That Act Rationally Like Humans

Binet, the inventor of intelligence tests, characterised intelligence as involving such abilities as reasoning, imagination, insight, judgement and adaptability. The concept of intelligence has proved difficult to clarify, but its core features are thought to be abstraction, learning and dealing with novelty – see Arthur Reber [Rebe86]. Bearing this characterisation in mind, the Encyclopaedia Britannica gives two definitions of AI. The first corresponds to systems that act like humans: the capability of a machine to imitate intelligent human behaviour. The second corresponds to systems that act rationally: a branch of computer science dealing with the simulation of intelligent behaviour in computers. In this case rationality is taken to mean logical reasoning to justifiable conclusions. Russell and Norvig use rationality as a characterisation for an ideal concept of intelligence, which amounts to studying intelligence or thinking in somewhat formal terms irrespective of the biological mechanism.

Allen Newell and Simon [Nwl76] give a more descriptive definition for general intelligent action such that “in any real situation behaviour appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity.” They go on to equate intelligence with the “ability to extract and use information about the structure of the problem space, so as to enable a problem solution to be generated as quickly and directly as possible.” Later, Newell

[Nwl90] defines intelligence as the degree to which a system approximates a knowledge-level system. Perfect intelligence is defined as the ability to fully utilise all the knowledge a system has at its disposal in order to achieve a goal.

This action-focused category of AI is not primarily concerned with mirroring the internal mental processes of humans, nor is it directly concerned with consciousness. However, three unsettled issues that affect the success of this category are: the lack of an acceptable definition for intelligence, whether action can be modelled formally, and whether intelligence depends on consciousness. All too often, once a machine is devised for performing a task that once was said to require intelligence, either the task is relegated or the machine is said to lack understanding, and the definition of intelligence is revised accordingly. It could be argued that for a system to imitate human behaviour requires it having a mind that is as fully featured as humans. In which case, the category merges with the "systems that think like humans" category.

2.5 GOFAI and Symbolic AI: Systems That Think Rationally

John Haugeland [Haug85] coined the phrase "Good old-fashioned AI" (GOFAI) for what is now called classical or symbolic AI. This is the view that intelligence arises from an ability to reason, where reasoning involves the symbolic manipulation of a set of facts and rules. This category of AI is interested solely in developing systems that can think rationally as opposed to modelling human thought processes. At first one would expect this approach to do well given the current sophistication of logical formalisms. However, as Haugeland [Haug97] points out there are still immense technical problems concerning knowledge representation, management and learning. Even "new-fangled" AI approaches, such as connectionism, face these problems. Haugeland goes on to suggest that AI needs to consider "the whole 'phenomenology' of an inner life," in particular building systems that understand and care about truth and falsity. Yet again the criterion for AI to succeed in this category has extended into the following category.

2.6 Strong and Weak AI: Systems That Think Like Humans

The classical view of AI treated the mind as a computer – see Haugeland [Haug97]. This was encapsulated by "The physical symbol system hypothesis" proposed by Newell and Simon [Nwl76]: "A physical symbol system has the necessary and sufficient means for general intelligent action." The 'necessary' stipulation implied that human minds were also symbol systems. This led Searle [Sear84, p28] to distinguish “strong AI” as the view “that the mind is to the brain, as the program is to the computer hardware,” and that the mind is a computer program. Hence, a computer executing the right kind of program would possess a mind. Searle (p31) suggests this view implies “that mental processes and program processes are identical.” On the other hand, “weak AI” is “the view that the computer is a useful tool in doing simulations of the mind” – see Searle [Sear97, p9]. Fetzer, [Fetz90, p61] gives an alternative conception as: “The ‘strong’ thesis that AI concerns how we do think. … The ‘weak’ thesis that AI concerns how we ought to think.”

2.7 Foundation of AI

Hubert Dreyfus [Drey97, p156] presents an analysis of four assumptions underlying the computational view of AI. These are a biological assumption that at some level the brain processes information; a psychological assumption that the mind operates on information according to rules; an epistemological assumption that all knowledge can be formalised; finally, an ontological assumption that information about the world essential to intelligent behaviour can be assimilated as a set of independent facts.

Classical AI takes a representational stance that leads predominantly to declarative, logic based formalisms – see Kim Sterelny [Ster90] and Fetzer [Fetz90, p269]. For non-representational approaches see Rodney Brooks [Bro97] and Timothy van Gelder [Geld95].

More recent approaches to AI have adopted other underlying assumptions. For example, (extreme computational) connectionism is based on the processing of distributed representations – see William Bechtel and Adele Abrahamsen [Bech91]. At the other extreme, the dynamical systems approach is based on the hypothesis: “natural cognitive systems are dynamical systems, and are best understood from the perspective of dynamics” – see van Gelder [Geld95].

Perhaps the most fundamental assumption is that the thesis for material chauvinism is false. That is, strong and mild AI support the belief that the basis of mind is not limited to biological mechanisms. In particular, some form of property dualism is not responsible for producing consciousness. This assumption is very important for it hints that it may well be possible to replicate mind.

These approaches propose different architectures are necessary for reproducing mind. However, it may be that each has its place in a theory of mind depending on the level of description that is of interest. To determine the prospects of mild AI the relationship between these approaches needs to be clarified. For example, could a simulation of the neural network specified by the dynamical system approach replicate mind?

3 Motivation

The thesis is motivated by a desire to answer the question, Could machines have minds? To tackle this question the thesis extends the work of Peirce and Fetzer on semiotics and dispositions, by providing a contribution toward a viable operational explanation for the mind/body relation, and, necessarily, the generation of consciousness. It supports these claims by elaborating how the key causal mechanisms that appear to be involved in semiotic systems might be implemented. The thesis examines and focuses solely on the ontic nature of the relationship between the mind/body, independently of any particular sensory modality – vision is used merely as an illustrative example.

Implementing semiotic systems in terms of neural network components trained on-line, was found to be practically awkward and too low levelled. To practically implement semiotic systems it is suggested that a programming language formulated to model both lawful relations and dispositional properties of the kinds that seem to be involved in mind/body relations can most advantageously specify the causal connections of systems of this kind. A preliminary stage in the development of this language emerges from the use of group-theory and its operators to reflect the general structure of these relations.

3.1 The Prospect of Mild AI Minds

According to a literature review the current prospects look bleak for mild AI and machines having minds. However, the issue is far from settled as the following discussion of the literature shows.

There have been a number of books written on the limits of AI with respect to the prospects of machines replicating mind, some of the more prominent ones being: Haugeland [Haug85], Dreyfus [Drey79], Fetzer, [Fetz90], Searle [Sear84], Sterelny [Ster90], and Roger Penrose [Penr89]. There have been many more papers written on how science could not replicate consciousness – for recent views see Y.Shapiro [Shp96] and J.F.Rychlak [Ryc95].

With respect to traditional computational and symbolic based AI, Gerald Edelman [Edel89, Edel92] has maintained a biological neural net stance against any form of machine consciousness. Rosen [Darp88, p271] has made similar remarks, suggesting that only neural nets can model complex systems and this is why traditional AI has failed. Maudlin [Mau89] suggests that a computational theory of consciousness is not possible since it would lack causal structure. Finally, variations on Godel's theory of arithmetic completeness are often put forward in attempts to refute machine consciousness – see Penrose [Penr89].

As an alternative to symbolic based AI, connectionism arose from the field of neural nets by emphasising the importance of distributed representations and emergent behaviour – see Bechtel [Lyca90, p254]. The connectionism versus symbolism debate is largely about what level of representation is appropriate – see Bechtel and Abrahamsen [Bech91], and Andy Clark [Clar89]. Margaret Boden [Bode89, p10] highlights further differences: in symbolism explicit rules are programmed in order to do something; in connectionism a neural net is trained, it learns, and apparently behaves as though it was following the rules. Edelman [Edel89, p33] says consciousness is continuous – it has to obey a continuity constraint, and that symbolic approaches are doubtful because they are discontinuous whereas neural nets are continuous. However, a continuous property at one level does not imply a continuous mechanism at the lower level. Douglas Hofstadter [Hofs83, p279] espousing a connectionist view says, "the brain itself does not manipulate symbols; the brain is the medium in which the symbols are floating and in which they trigger each other...".

Nevertheless there are doubts about the representational power of connectionism. Jerry Fodor and Zenon Pylyshyn [Fod88, p63] point out that if viewed as an implementation level theory it will run into difficulties when dealing with issues relating to semantics, productivity, systematicity and compositionality. They suggest a symbolic level theory is better suited to dealing with these issues. "The question is whether the kind of activity they (representations) exhibit should be accounted for by the cognitive model or by the theory of its implementation."

Fodor [Lyca90, p282] and others (see Goel et al. [Goe88b]) have argued that language and thought exemplify properties of mind. These properties can be implemented by the brain but are seen to have a constituent structure in their own terms and can be studied abstractly, e.g. via symbolic logic and other symbolic representations. They point out that the form of representation affects the theory for better or worse, and that symbolic theories need not necessarily depend on connectionist implementations. Goel et al. say a once held belief was that "Connectionism could side-step pretty much all the representational problems and dismiss them as the bane of Symbolism." In terms of David Marr's levels [Marr82], the architecture chosen may have different primitive functions that can affect the problem solution depending on how close it is to the architectural level. Fodor and Pylyshyn summarise, "The point is that the structure of 'higher levels' of a system are rarely isomorphic, or even similar, to the structure of 'lower levels' of a system."

3.2 The Prospect of Mild AI Consciousness

Turning now to the properties of mind that mild AI seeks to replicate, *qualitative consciousness* refers to that aspect of consciousness concerned with the raw sensation the first-hand experience of consciousness has – see William Lycan [Lyca96]. The term quale has been introduced to refer to the content of specific conscious experiences, such as the sensation of redness, which is referred to as *the red quale*. It applies to the experiential content of mental experiences, such as the sensory modalities (auditory, olfactory, tactile and visual) and thought. Thus, the experiential content of specific sounds, smells, thoughts (etc.) can be referred to via an appropriate quale. This aspect of consciousness is analysed in further detail in chapter two.

Qualitative consciousness has been characterised as a problem in the philosophy of mind and to mild AI since an acceptable explanation for it has not been forthcoming despite many years of investigation – see Alan Code [Lepo91, p105], and Martin Davies & Glyn Humphreys [Dave93, p14]. For example, Colin McGinn [McGi91] has suggested that we may not even have the cognitive ability to understand a theory of consciousness, just as a goldfish could not comprehend abstract algebra. Almost all the prominent contemporary theories of consciousness, such as those of Daniel Dennett [Denn91] and Penrose [Penr89], only marginally deal with the problem of qualitative consciousness. One reason for this neglect is that there has not been an adequate formal theory for qualitative consciousness and it is said to be a 'hard' problem – see Jonathan Shear [Shea97]. This will be investigated in the next chapter.

The experience of qualitative consciousness is an emergent phenomenon, but the underlying process necessary to give rise to this is precisely determined, and can be entirely explained. While the ontological levels in a system give rise to a type of predictable emergence, as implementers, we still have to program, or configure, the right structure in the lower levels. These points are discussed later.

When asked whether a machine could ever be conscious, Minsky [Mins87, p160] said that this should be left to future designers. However, as the thesis will show, for assessing the future direction of mild AI there is much to be gained from advancing theories for consciousness, even now.

4 Strategy

The objective of the thesis is to explore the prospects and promise of mild AI. The strategy taken here is based on selecting a viable criterion for mind and analysing how a suitable digital machine can be programmed to satisfy this criterion. Fetzer's [Fetz90, p39] criterion for mind, as the capacity to make a mistake, was adopted after reviewing various candidates – see chapter two. Fetzer suggests this criterion implies the underlying system could be a semiotic system, which is defined as being a mind. This definition for mind is based on Peirce's theory of signs – see below. However, Fetzer goes on to show that, since programs are not semiotic systems and therefore not minds, the equivalence made by strong AI is inappropriate. Peirce's theory of signs and Fetzer's analysis of the Analogical Argument are now reviewed followed by some comments on how this shapes the strategy.

4.1 Peirce's Semiotic

Peirce set about developing what he called an architectonic for philosophy, a comprehensive view that encompassed pragmatism, semiotics (his theory of signs), phenomenology and metaphysics based on synechism (his theory of continuity) – see Charles Hartshorne [Hart58]. While these themes are very much interrelated, for present purposes only aspects of his semiotic are drawn upon – later chapters will highlight some parallels to the other themes.

Fundamental to Peirce's theory of signs is the thesis that a sign should be understood as a property of a semiotic process involving an irreducible triadic relation between a sign, object and interpretant. Notice that Peirce's analysis is from a phenomenological perspective, it is in this sense that a sign is to be understood as a component of a triadic whole. Quoting Carl Hausman [Haus93, p72], "A semiotic process requires that there be something that has an object for which that thing stands, an interpretant that relates it to its object, and a respect or ground that qualifies the relation between the thing functioning as a sign and its object."

Starting with what Peirce calls the *dynamical* object, this is the object as it exists independently of any interpretation by a mind – however, it could be a mental, abstract object. A ground relation then transforms an aspect of the dynamical object into a representation called the *immediate* object. This produces an effect, called the *interpretant*, in the subject acting as interpreter. A *representamen* is the product of a representation process, and when this has a mental interpretant it is called a sign – Peirce wanted to allow for non-human interpretation.

Peirce classified the semiotic characteristics of signs into three trichotomies. The first trichotomy classifies the properties of signs. These are i) a *qualisign*, which is a pure monadic quality of an object, i.e. a quale, ii) a *sinsign*, which is an individual thing or event, and iii) a *legisign*, which "is a law that is a Sign". A sinsign is dyadic in that it must embody a qualisign, and a legisign requires a sinsign in order to be instanced – see Hausman [Haus93, p86].

The second trichotomy concerns the relation of the sign to its dynamic object, that is, the nature of the sign's ground relation. These are i) an *icon*, which bears some similarity to the object, ii) an *index*, which is a cause or effect to / of the object, and iii) a *symbol*, which is associated to its object by convention. Semiotic grounds must be distinguished from causal grounds, which are discussed below.

In the third trichotomy, Peirce distinguished three types of interpretants depending on the effect they produced in the subject. These are i) a *logical* interpretant, which is either a habit-change or thought, ii) an *emotional* interpretant, which is a "feeling" or "sensuous content", and iii) an *energetic* interpretant, which leads to an act or reaction.

4.2 Analogical Argument for Strong AI

Fetzer defines mind as a semiotic system and distinguishes a hierarchy of five types according to what kind of signs are used and how they are manipulated – see Fetzer [Fetz90, p41-58]. According to the signs used the first three types are defined as i) iconic, ii) iconic and indexical, iii) iconic, indexical and symbolic. The remaining two types are defined according to how the signs are manipulated: iv) transformational, and v) metamentality. Thus, the last two types have the ability to reason logically and for criticism, respectively. The criterion for a system to be a mind then arises from its capacity to make a mistake since it may take "something to stand for something other than that for which it stands" (p40). This also implies the system is conscious according to Fetzer's definition (p81): "A sign-using system is *conscious* (with respect to signs of a certain kind) when it has both the ability to utilize signs of that kind and the capability to exercise that ability, where the presence of signs of that kind within the appropriate causal proximity would lead ... to the occurrence of *cognition*."

Fetzer characterises the strong AI position in terms of "The Basic Model" (p16). This compares human beings and digital machines by forming an analogy between stimuli, processes and responses, with inputs, programs and outputs, respectively. Fetzer recasts this explicitly in the form of the inductive Analogical Argument for strong AI (p277):

	Human Beings:		Digital Machines:
Premise 1:	Stimuli	=	Inputs
Premise 2:	Responses	=	Outputs
Then infer:			
Premise 3:	Processes	=	Programs
Premise 4:	(= Minds)		
And infer:			
Conclusion:			(= Minds)

Table 1. The Analogical Argument for Strong AI. From Fetzer [Fetz90,p277].

Premise 3 is singled out as the point at which the analogical argument for strong AI breaks down. This happens when an attempt is made to formulate the relations between stimuli, processes and responses, and inputs, programs and outputs, according to the triadic sign relation of Peirce – see **Figure 1**.

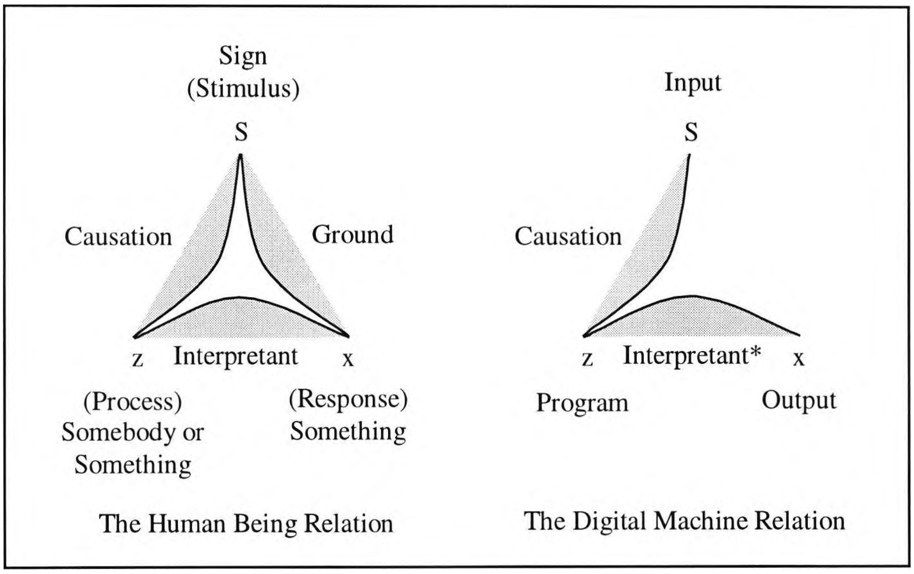


Figure 1. Comparing operation of human beings & digital machines in terms of Peirce's Sign triad. After Fetzer [Fetz90, p277-278].

The figure highlights that there is not a true semiotic ground between input and output in the case of digital machines. Hence, the relation between program and output is not a true interpretant, because of which it is designated an "interpretant*". Fetzer remarks (p278), "there may be causal connections between a cause C (call it "stimulus") and an effect E (call it "response"), but unless that causal connection obtains *because* that sign is an icon or an index or a symbol in relation to that effect (...), it cannot be a *semiotic* connection."

Fetzer [Fetz91] emphasises that the dynamical object only counts as a true sign to the sign user through the effects of the interpretant on the disposition and habits of the sign user, and that these are taken as indexical, iconic or symbolic in the semiotic sense as part of the internal sign triad relation. These are semiotic grounds rather than causal grounds since whether the sign user takes them as indexical, iconic or symbolic depends on the internal context determined by past experiences and innate factors.

Fetzer [Fetz90, p88] suggests, "if human beings are pragmatic semiotic systems, while computational devices are syntactic symbol systems instead, however, then they are incapable of replicating one another's modes of operation." They are "fundamentally different" kinds of causal system.

4.3 Counterpart to Analogical Argument for Strong AI

A counterpart to the basic model can be formulated by exploiting two oversights in the argument. Firstly, the analogical argument runs broadly as follows:

Premise 3: An equivalence is made between mental processes and executing programs.

Premise 4: From this it is inferred that computers have minds given the right program.

However, the converse does not necessarily follow:

A) Given a mind generated in a computer executing a program.

B) This does not imply equivalence between mental processes and executing programs.

In other words, there may be some other means of relating mental processes to machine processes not covered by Premise 3. One approach is the systems reply in which mind is attributed to a property of the system as a whole. This approach has been used to counter Searle's Chinese Room argument against strong AI – see Jack Copeland [Cope93, p126]. However, Searle was quick to refute this counter argument. To use the "skin-of-an-onion" analogy given by Turing [Tur50], if each layer of skin corresponds to a function of mind, which can be described in purely mechanical terms, as layers are peeled away at what point do we reach the mind?

There is a promising alternative to the systems reply approach, which also highlights the second oversight of the analogical argument. Namely, the alternative distinguishes the ontological levels that may exist in the system, where an ontological level is individuated by the laws that are in force – see chapter three and the appendix for formal details. In this regard, for the purposes of analysis the thesis takes, what might seem, an extreme stance by lumping together computers and programs as properties of the same ontological implementation level. In addition, the mind is treated as a higher ontological level with respect to its underlying implementation level.

While this may appear to be heading toward a dualism, the relation between ontological levels developed herein is amiable to analysis and implementation – it shows much promise and remains consistent when supervenience is considered – see Mellor [Mll93]. In addition, it does not rely upon or imply non-deducible emergent properties – see Claus Emmeche et al. [Emm97] and chapter three. Consequently, the version of mild AI adopted here does not presuppose Premise 3 holds. Hence, pursing the prospects and promise of mild AI will involve examining from an operational perspective how the mind might otherwise be related to the mechanism.

4.4 Theories of Content & Dispositions

Theories of content set out to explain how meaning arises from the syntactical structures in which representational theories for mentality are often couched – for a review see Georges Rey [Rey97] and Barbara von Eckardt [Ecka93]. Fetzer [Fetz90, p86] groups the various approaches according to their computational, representational or dispositional conceptualisation, and notes that these emphasise syntactical, semantical and pragmatical aspects, respectively.

Of the approaches to theories of content, only the dispositional conceptualisation seems to readily offer a solution to the problem of content determination. This approach was favoured by Peirce and is reflected in his third trichotomy, which classifies the effects of interpretants on the subject e.g. habits, feelings and reactions. Briefly put, in this context a disposition is a tendency of a subject to behave in a consistent way in certain situations – see Robert Audi [Audi95]. An interpretant is generated as part of the semiotic process amidst a web of other triadic sign processes, as such it has the potential to cause a disposition in the subject if fully realised, but being part of an occurrent process, it may be superseded.

Dispositions and the problem of content are discussed further in chapter three. One theme to arise in these subsequent chapters is an integrated view of the seemingly separate ideas on the nature of meaning, qualia, resemblance and iconic mentality. In particular, the definition of a quale developed in chapter four suggests different instances of the same quale share a common operational causal structure – a type/token relationship that contributes to the quale's role as a sign to the system. This integrated view is discussed further in relation to Peirce's semiotic in chapter six.

Fetzer [Fetz90, p288] went on to characterise semiotic and computational systems in terms of laws that concern "the logical form of the lawful relations that characterise systems of these types." Fetzer's primary interest was how the laws could be used to characterise each domain and the behaviour of the systems. For the present purposes interest lies with exploring whether mild AI can succeed and in this regard the laws are also helpful for refining the structure of semiotic systems. Laws HL10 and CL10 relate to dispositions:

$$\text{(HL10)}\ (x)(t)[B^*xt \Rightarrow (EF\,xt = u \Rightarrow M^*xt^*)].$$

Here HL10 is a human being (semiotic system) law, where B* refers to the brain and B refers to a brain-state, M* refers to the mind and M a mind-state. This law asserts, "For all x and all t, if x were a brain of kind B* at t, then exposure to environmental factors of kind EF at t would invariably bring about the acquisition of a mind of kind

M* at t*." Fetzer suggests "brains" B* should be viewed as "predispositions to acquire semiotic dispositions M* falling within some specific range..."

$$\text{(CL10) } (x)(t)[H^*xt \Rightarrow (EF\ xt = u \Rightarrow C^*xt^*)].$$

Here CL10 is a digital machine (computational system) law, where H* refers to the hardware and H refers to a hardware-state, C* refers to a computational disposition and C refers to a computational-state. So the above law asserts that a computer H* subjected to environmental factors of kind EF would invariably bring about the acquisition of a computational disposition of kind C*.

Fetzer (p288) remarks that these computational dispositions are not enough to bring about outputs as a causal consequence of inputs as required for Peirce's triad and therefore for semiotic systems. When viewed in this manner, computers would lack the right kind of dispositions to be classified as semiotic systems, and so would be unable to satisfy a dispositional theory of content determination. Chapter six examines how grounded semiotic processes might be replicated in a computer.

Law HL10 is a deterministic causal conditional, Fetzer supplements this by considering analogous laws that relate brains of kind B* and the acquisition of semiotic abilities – see [Fetz96, p109]:

$$\text{(LC-6) (a) } (z)(t)[B^*zt \Rightarrow (EF\ zt = u \Rightarrow SAzt')];$$

$$\text{(b) } (z)(t)[B^*zt \Rightarrow (EF\ zt = p \Rightarrow SAzt')].$$

Here, law LC-6b states that given a brain of kind B* and environmental factors of kind EF at t, then z will probabilistically possess a semiotic ability of kind SA at t'. Of note is that the dependency is probabilistic and that the semiotic ability "is a part but not all of those factors whose presence constitutes a mind state." The chapters that follow will seek to refine the nature of the system structure that gives rise to these laws. In particular, starting in chapter three, graded dispositions are examined as a basis for semiotic processes, suggesting that, operationally, a superposition of deterministic graded dispositions may give rise to law LC-6a, individually, and collectively appear to support law LC-6b.

4.5 Making the Case for Mild AI

Consequently, to satisfactorily rejoin and redress the above problems facing mild AI involves showing how a mind can be replicated in a suitable computer. Central to this will be examining the operational requirements for a minimal semiotic system independently of any underlying physical implementation. This involves two steps. Firstly, exploring how a system can operate in terms of semiotic processes. Secondly, exploring the system structure sufficient for generating consciousness – a necessary condition for a cognitive semiotic system. Finally, to make the case for mild AI requires determining whether a suitably programmed machine is able to instantiate such a system.

The analysis of the operational requirements for implementing semiotic processes looks to the temporal counterparts of state-transitions and stationary-representations, which are termed functional-flows and temporal-representations, respectively. This is closely aligned with a dispositional conception of system operation. These terms are defined in the appendix. Briefly, they highlight the varying structure of temporal patterns of activity in time. For the purposes of replicating mind, preserving the causal-flow structure of a mental process appears to be an important requirement. A causal-flow is a type of functional flow, this, and what is meant by 'causal', is discussed in chapter three and the appendix.

Showing how a system can operate in terms of semiotic processes would help to solve some troubling issues, for instance, the spatial-temporal limit problem and issues traditionally raised against functionalist approaches to mind, such as the modal independence problem and the interpreter regress problem – these are discussed in chapter three. A solution to these problems is particularly important to the development of semiotic systems and variants, such as the iconic state machine of Aleksander [Alek96], since it helps refine their structure and sanctions their plausibility.

The approach to developing the system structure for generating consciousness from an operational perspective is based on an analysis of systems and Peirce's triadic relation in terms of dispositions and ontological system levels using the characterisation developed in chapter three and the appendix. This helps clarify exactly between what and mind an equivalence should be drawn in Premise 3 of the Analogical Argument. It also suggests a dispositional mediating level should be inserted into the above laws of human beings and digital machines. A goal of chapter four is to examine why this mediating level might be necessary.

The justification for the introduction of ontological levels is a consequence of implementing semiotic processes in terms of instances of dispositions and establishing the right kind of causal structure between these instances. This structuring is driven by the system's need to signify reality by categorising these aspects as operational entities upon which decisions can be made. Consciousness arises through the manner in which the signified perception of reality is generated – see chapter four and the section on the requirement for system structuring for causal accessibility. This makes mind and consciousness the result of a co-ordinated occurrent system wide activity.

As a formal foundation for determining the operational requirements for a semiotic system, a characterisation of systems is developed in terms of the causal structure and ontological levels in the system. Brains and computers can be classified as numeric and symbolic systems, respectively – see chapter six. Consequently, the characterisation so developed is used to compare these two types of system and to formulate conditions under which they may give rise to equivalent ontological levels.

According to the theorems developed in the appendix, the necessity of an ontological mediating level would effectively enable a suitable computer to replicate mind. One important factor in determining whether a computer is suitable for this purpose is its performance capacity. Hence, some estimates are calculated for the performance capacity bounds that would be required to replicate a human mind.

5 Methodology

A successful case for mild AI must explain how to replicate mind in a system such that the operation, purpose, interaction and reason for every component are understood. The approach for this analysis is from an applied engineering perspective, that is to say, how could a semiotic system be built, in theory? Thus, the analysis seeks to provide an explanation on engineering grounds alone independently of introspection or neuro-philosophy. These points are now clarified.

5.1 Theories of Mind

Jeffrey Gray [Blak87, p468] issues a word of warning on how theories of consciousness can dictate the questions and answers one looks for when pursuing a scientific study of consciousness. For example, the traditional dualist view of mind and body (see chapter two) can be misleading in that it presupposes two distinct properties (one mental and one physical) that plays against our preconception of objects as being physical. Max Velmans [Vlm98] highlights the confusion that can arise when theories are based on ill-defined beliefs, as happens in investigations into the neuronal correlate of consciousness – see Francis Crick and Christof Koch [Crc97]. Part of this investigation will be unpacking our presumptions, dispositions and innate beliefs concerning the nature and meaning of terms (e.g. consciousness, mind, qualia) and our beliefs about the way things are in the world (e.g. its representational character). Nothing can be taken for granted.

Brian Smith [Smit96] warns that in developing adequate theories of computation or cognition, ontological problems concerning the methodology adopted will have to be addressed. By way of illustration the methodology of the analytic tradition is questioned on the grounds that the formalisms adopted to describe situations will shape the conclusions reached. The foundational order of mathematics itself is said to be in need of an overhaul e.g. continuity is currently defined in terms of discreteness. This battle between continuity and discreteness will surface many times in the thesis.

Consequently, bearing these presumptions in mind, algebra, elementary functions and calculus will be used for modelling purposes – see Paul Cohn [Cohn93] and Earl Swokowski [Swok79] respectively. Some additional mathematical tools will be introduced along the way – such as dynamical systems, D.K.Arrowsmith [Arro92], fuzzy logic, H.T.Nguyen, [Ngy96], group theory, D.J.S.Robinson [Rob95], and wave mechanics H.J.Pain [Pain93]. First order logic, with the normal connectives and quantifiers, will be used as a meta-language for presenting some lines of reasoning when necessary – see Moshé Machover [Mach96]. Thus, axioms will indicate assumptions of the thesis, and lemmas mark intermediary theorems. A handful of requirements for semiotic systems are tendered and represent a paraphrase of the implications of the ideas under consideration at that point.

It will also be necessary to use two types of system to present the thesis. Firstly, dynamical systems theory is used to model the processes responsible for mind and qualitative consciousness – see Gelder [Geld95]. Secondly, formal systems are used to characterise conventional computers and to reason about the relationships within the dynamical system, and between both types of system – see J.E.Savage [Svg98]. There has been much debate concerning the relative merits of these two types of systems and their use in cognitive modelling, e.g. connectionism versus symbolism – see Bechtel [Bech91]. Laws and systems are discussed further in chapter six.

The explication of cognition and consciousness in chapter two is based on Carl Hempel's [Hmp52] three levels of definition in empirical science: nominal definition, meaning analysis, and empirical analysis. Nominal (or *stipulative*) definition introduces a new word, phrase or expression as having the same meaning as some old word, phrase, or expression. Meaning analyses are reports of the established usage of words, phrases, or expressions within a language using community. Empirical analyses, by contrast, redefine the meaning of words, phrases, or expressions on the basis of the results of empirical study of the properties of samples or examples of things of that kind. The ancient Egyptians, for example, were familiar with gold as a yellow, malleable metal, but with the development of the atomic theory of matter, gold has come to be defined on the basis of its atomic structure as consisting of atoms with atomic number 79, – see Hempel [Hmp66], and Steven Chaffee [Chf91] for a discussion of concept explication.

Descriptive statements make assertions concerning the way things actually are in the world. Normative statements make assertions about ideal situations – see Henry Kyburg [Kyb91]. The thesis is concerned with an idealised normative form of mind, not necessarily human, since it is challenging material chauvinism – see Ned Block [Blc78]. However, it will necessarily draw upon descriptive accounts of the human mind and consciousness for guidance. The thesis can be interpreted as providing an explanation of the operational requirements for a mechanism for producing mind.

5.2 System Analysis & Design

This subsection highlights the important difference between levels of analysis, levels of organisation, and ontological levels with respect to systems.

System design deals with levels of analysis and levels of organisation. Marr [Marr82] proposed a three-step approach to system analysis involving the abstract analysis of the problem, algorithmic specification (showing the Turing machine influence), and implementation details. Traditional system analysis has four stages: a statement of requirements, formal specification, design proposal and implementation proposal – see Cohen et al. [Cohe86]. The system analysis approach is neutral with respect to implementation architecture. In some respects, object-oriented design is a modern version of Marr's approach. For example, Booch's revised design method breaks analysis down into four stages: logical structure, physical structure, dynamics of classes and dynamics of instances – see Ian Graham [Grah95, p212].

The second major aspect to system design is detailing the levels of organisation. Here the system is divided into subsystems according to criteria such as function. Simon [Sim96, p184] comments that in some cases the division into subsystems may be arbitrary, while in others there may be a hierarchic ordering, which can be treated as structure within a level, such as a management hierarchy.

A.Bailey [Bail98] distinguishes five levels of description: hierarchies of epistemological convenience, modelling hierarchies, explanatorily pragmatic hierarchies, group-level properties and levels of being. The first three are classified as epistemological and the latter two as ontological. Epistemological convenience adheres to the view that there is ultimately just one fundamental and complete description of the universe and all higher level descriptions could be reduced to this bottom level. Modelling hierarchies posit levels as a means of discovering structure in a domain. Explanatory pragmatics adopts terms appropriate to that being described, such as economical or psychological. Group-level properties literally concerns the properties or patterns a group of objects has when considered as a mereological whole, Bailey suggests a soccer-team is one example. Finally, levels of being is a much more contentious proposal and concerns the ontological existence of levels.

With respect to ontological levels viewed in terms of levels of description, Bailey suggests two conditions must hold for such levels to exist. Firstly, the description is irreducible since it captures patterns or generalities that cannot be expressed by a lower level description. Secondly, the description is ineliminable and therefore necessary for a full description of the universe. Bailey doubts whether any serious ontological pattern at one level will turn out to be irreducible in this way. However, very often the same thing can be described in different ways on pragmatic grounds.

The ontological level of being has implications toward descriptions of consciousness. According to Bailey's conditions on levels of descriptions, either it is explainable in physical terms, and therefore ontologically reducible, or it is irreducible. However, it does not necessarily follow that something that is ontologically irreducible cannot be indirectly explained in physical terms. For example, Searle [Sear98] suggests such a strategy for scientifically explaining the ontologically subjective – see the discussion on ontological system levels in chapter three.

6 Logical Overview of Chapters

6.1 Chapter Two: Cognition & Consciousness

The domain of mild AI covers the properties of mind. To judge the ambitions of mild AI requires having a characterisation of this domain. Hence, the starting point for this chapter is the animal kingdom since it is important to remember that minds come in many forms and operate in varied contexts. To characterise and individuate these mental phenomena requires a definition of brain and mind since they form the seat of intelligence. Further refinement necessitates definitions for cognition, consciousness, and finally qualia, whereupon the quale problem is presented. It is suggested that any complete theory of consciousness has to tackle this problem since it is closely related to the problem of content determination. A review of current theories reveals that it has yet to be explained satisfactorily. The chapter ends by adopting the semiotic conception of cognition and consciousness as guiding definitions for the subsequent chapters.

6.2 Chapter Three: Operational Requirements for Semiotic Processes

For mild AI to succeed it has to show that a semiotic system can be replicated in a machine. This requires showing how a system can operate in terms of semiotic processes. Here, Peirce's semiotic acts as a conceptual framework from a first person perspective for reasoning about signs, the interpreter, and mind. This is a pertinent point from which to explore the operational requirements for mind from a third person perspective.

This chapter suggests semiotic processes can arise in a system whose operation at a particular ontological level is based on functional-flows and temporal-representations. These terms are discussed and defined in the appendix. Three problems levelled at theories of consciousness are reviewed and used as a guide for the analysis.

A hypothetical system containing a hierarchy of three ontological levels (an implementation level, a dispositional level, and a phenomenal level) is proposed as a working framework on which to refine the operational basis of semiotic processes. A particular kind of dispositional process is investigated as a possible mechanism linking the implementation and phenomenal levels. The term "dissociated disposition" was introduced to refer to an abstraction of this mechanism. These characterise the variable dispositional potential of the processes within certain kinds of causal systems. This concept of variable dispositional potential plays a central role in this and the following chapters.

This chapter focuses on the operational nature of the processes underlying semiotic processes. Although qualia are mentioned in passing, the next chapter considers how these underlying processes are structured to produce qualia. The next chapter after that discusses the requirements for these processes to be considered semiotic and considers how a machine might be configured to replicate them.

6.3 Chapter Four: Structural Requirements for Generating Mind in a Semiotic System

The previous chapter suggested that the primary operational requirement for implementing a semiotic system is a system that has the right kind of causal structure. This chapter shows how primary-consciousness, a primitive form of semiotic-consciousness as explained in chapter two, arises in the system through the way in which it signifies reality and how this is achieved through the relationship between the ontological levels in the system.

The next step in exploring the prospects and promise of mild AI is to examine how dispositional processes are structured to produce mind. This requires determining the system framework for generating, at least, qualitative primary-consciousness. The guiding light here is the search for a framework that adequately deals with the interpreter regress problem (introduced in the last chapter).

To address the interpreter regress problem requires showing how consciousness can arise in the hypothetical system without necessitating an internal homunculus. This chapter explores the structural requirements on the systems' operation that is collectively sufficient for generating consciousness without necessitating a homunculus. This is brought together in the most important section of the chapter entitled "Requirement for System Structuring for Causal Accessibility."

6.4 Chapter Five: Implementing Causal-Flow Systems

Minds were defined as a type of semiotic system, which was shown to contain a hierarchy of three levels: implementational, dispositional and phenomenal. In turn, the dispositional level is a type of causal-system characterised by its causal-flows and their causal structure. Consequently, this chapter considers how a causal-flow system might be implemented.

Traditionally, in system design attention is paid to the permissible state transitions. With instances of dissociated dispositions, the causal structure between the order of these conventional transitions is of more importance. An instantiated dissociated disposition's causal-flow enforces a particular structure on the temporal order of these conventional transitions. Hence, programming dissociated dispositions is partly concerned with specifying this dynamic causal structure. They can be thought of as demarcating a confluence of states in a causal state space, except that at the dispositional level, the state space is open and continually changing – see Douglas Lind and Brian Marcus [Lin95].

From a reductionist point of view, dissociated dispositions could be specified by giving the state transitions at the implementation level. However, this would be a monumental task, and would miss the causal structure that was under investigation. To compound the situation, they are evolving, self-programming entities. Hence, after only a brief period, the current state of the system will be characterised by a different disposition configuration. This emphasises their evolving, malleable and transient nature - a difficult nature to program.

The goal of this chapter is to push the formal characterisation of dissociated dispositions to the point where the main requirements for a modelling technique for specifying systems based on them can be deduced. This proceeds by examining how they can be described and modelled mathematically, which prompts the need for a tailored programming language.

Part of this exploration involves a practical example that is investigated to demonstrate the applicability of modelling with dissociated dispositions. In addition, it also sheds some light on "Causal Alteration"; an important process in the dynamics of instantiated dissociated dispositions. This embodies the fundamental mechanism by which the dispositional form of a dissociated disposition is coded. This is discussed in the section on implementing instances of dissociated dispositions, and in the section on the generic Gaussian alterator.

6.5 Chapter Six: Logical Requirements for Replicating Semiotic Processes

When concern lies with the mathematical form of a system's dynamics rather than properties of the system, such as the learning and emergent capabilities typically attributed to neural nets, then in this mathematical sense, biological minds are a type of numeric system and digital computers a type of symbolic system (the terms *numeric* and *symbolic* are explained later in the chapter). For many decades these two classes of systems have been investigated, in the case of numeric systems to determine just how mind arises in certain systems, while with symbolic systems the investigations have yet to determine whether they can support mind at all.

The goal of this chapter is to determine whether semiotic processes can be replicated on a digital machine and under what conditions this might be possible. This involves two aspects: that the system operates semiotically, meaning in has the right kind of processes for the right reasons, and establishing an appropriate equivalence between numeric systems (the systems behind minds) and symbolic systems (the systems behind computers) and their ability to support the semiotic processes concerned. This, in conjunction with the results from the prior chapters is used to determine whether the semiotic properties of mind can be replicated under these limited conditions and so help determine whether mild AI can succeed.

The introduction chapter drew attention to the importance of the ground relation in semiotic processes and highlighted Fetzer's analogical argument for strong AI that suggested digital machines lacked such relations when viewed from a computational perspective. The first half of this chapter analyses Fetzer's suggestion in some detail where the focus shifts to examining the semantic grounding of semiotic processes and the logical form of the operational structure this in turn implies about the semiotic processes.

The second half analyses and compares the structure of the two classes of system and states under what conditions a symbolic system might be considered weakly equivalent to a numeric system, this would correspond to replication. The primary conditions relate to the ontological levels in the system, the perspective of the observer, and the causal structure of the system. If the semiotic properties of mind can be reproduced in a symbolic system under these conditions, then mild AI is possible. A brief calculation is presented in order to estimate the future performance capacity of digital machines and whether they would have sufficient capacity to support a replication of mind.

6.6 Chapter Seven: Conclusion

The field of AI is revisited in light of the exploration into the prospects of mild AI. Given that mild AI may show promise, the implications and impact of this on the analogical argument for strong AI is discussed. Major issues face the future of mild AI, such as determining the requirements for, and authenticating, an AI person. The prospects of mild AI are discussed.

Peirce's triadic semiotic process is categorised in terms of ontological levels. This places the sign, subject and immediate object at the phenomenal level, the interpretant at the dispositional level, and the dynamical object at the implementation level. A Peircian sign is the qualitative counterpart of an entity signified for operational purposes.

Fetzer's laws for human beings are supplemented by suggesting brains could be treated as compound systems containing a neurological system and a lawfully dependent dispositional system.

Finally, the version of mild AI supported here takes the operational view that mind is a property of an occurrent system wide process. Therefore, this version of mild AI is not susceptible to arguments levelled against the analogical argument for strong AI since it doesn't equate minds with programs (Premise 3).

A summary of the dissertation's contribution to knowledge and possible areas for future work is given.

6.7 Appendix

A formal characterisation of systems in terms of their ontological levels and causal structure is presented. Formal definitions for representations, functional-flows and the perspective of an observer in relation to a system are also presented. Finally, a specification of the practical example discussed in chapter five is given.

Chapter 2

Cognition & Consciousness

1 Overview

This chapter surveys the territory of mild artificial intelligence, a territory populated by a society of ingenious animals able to survive in complex and sometimes hostile environments. The survey will amount to providing definitions for the main areas in the field, such as the mind, cognition, consciousness, and qualia. These are the phenomena mild AI would have to recreate in machines. By exploring the relationships between these phenomena, qualia will arise as being strategically very important, and solving the quale problem a necessary precursor for the promise of mild AI.

1.1 Logical Development

The domain of AI covers the properties of mind. To judge the ambitions of mild AI requires having a characterisation of this domain. Hence, the starting point for the discussion is the animal kingdom since for ontological reasons it is important to situate cognition and consciousness as having arisen in animals through their need to survive. To characterise and individuate these mental phenomena requires a definition of brain and mind since they form the seat of intelligence. Further refinement necessitates definitions for cognition, consciousness, and finally qualia, whereupon the quale problem is presented. It is shown that any complete theory of consciousness has to tackle this problem by explaining the origin of qualitative consciousness. A review of current theories reveals that it has yet to be explained satisfactorily. The chapter ends by adopting the semiotic conception of cognition and consciousness as guiding definitions for the subsequent chapters.

2 Animal Life

One dream of workers in mild artificial intelligence is to produce creatures that can survive on an equal footing with us in society. Another dream is to create creatures that surpass us in ability – see Stan Franklin [Frnk95]. These dreams focus on the animal kingdom as the subject of study and eventual imitation. Consequently, this section presents animals and environments as the ontological basis and context with reference to which cognition and consciousness should be considered.

2.1 Animals & Environments

James Gibson [Gibs79] popularised the view that animals inhabit an environment forming an ecosystem in which a constraining, circular and reciprocating coevolution takes place. Gibson suggested that analysis could be posed at physical and ecological levels. Physical descriptions deal with physical and physiological processes, whereas ecological descriptions are concerned with the animal as an active, perceiving organism in its surrounding environment. An animal was loosely defined as a living organism having sense organs and muscles. Of more specific concern here are those animals that belong to one of the arthropod, mollusc or vertebrate phyla (classes), since these animals exhibit the most intelligent behaviour. A common feature of which is a well-developed central nervous system, in particular, those that have a cranium encasing part of the nervous system. By convention this part is called the brain – see Gordon Shepherd [Shep83b].

Picking out the important factors, Gibson's terminology emphasises animals in environments: a domain of discourse featuring animals as individual bodies, having brains, sense organs and muscles enabling them to interact with objects and other individuals in their environment. Under this description a brain is associated with a distinct *living body* and its purpose is to endow the animal with a degree of intelligence enabling it to survive whilst roaming in an *environment.*

2.2 Animals, Brains & Minds

Giving a stipulative definition for the brain in animals, as above, was relatively easy. Providing a neurological description for the physical structure of the human brain is also reasonably straightforward in theory but not practical yet – see Semir Zeki [Zeki93]. Although, some would argue that it might have as yet unknown properties or features that are incomprehensible to us – see McGinn [McGi91].

Should the definition of a brain be restricted to animals? Definitional problems arise when brains are considered as a basis for minds since, conceivably, there could be brains without minds and minds without a brain tipped central nervous system. If a computer is able to house a mind, should its processing unit be called a brain? It is tempting to define a brain as any structure that houses a mind, for in a future with artificial minds it may be necessary to refer to the underlying mechanism, which could be a form of man-made machine. In this sense a definition of brain is dependent on a definition of mind.

3 The Mind

Very rapidly the discussion has arrived at the need for a definition of mind simply in order to define brains. A rather more compelling reason is to determine what distinguishes the mental from the non-mental and so to uniquely characterise the basis of intelligent behaviour in animals. Without such a characterisation, it would be impossible to evaluate theories of consciousness proposed in the field of mild AI. Consequently, this section starts by enumerating the realm of characteristics encompassed by mind. It then reviews the main denotative definitions of mind and finally goes on to review various degrees of descriptive definitions. The perspective on mind adopted by the thesis will be explained at the end of the chapter.

3.1 Characterising the Mind

The first task in defining mind is to draw a boundary around what is to be defined. This depends on the school of psychology one subscribes to, whether it's behaviourism, functionalism, or one of the other schools, since they often espouse different elements – see Tony Malim and Ann Birch [Mali98]. Taking a middle road, cognitive psychology attempts to explain behaviour by studying the internal processes occurring in an animal. This tolerates a more liberal individuation of the mind. Thus, an acceptable starting point is, perhaps, Gilbert Ryle's [Ryl49, p61] remarks on the traditional coarse tripartite classification of the domain of mind as consisting of those mental processes as being cognitive (thinking), emotional (feeling) or conative (willing). In contemporary terms, this would correspond to the more extensive interrelated areas of cognition, consciousness and behaviour. Cognition covers internal skills such as attention, perception, memory, thinking and language. Consciousness covers sensation and perception, emotions, motivations and states such as pain and stress. Behaviour covers instinct, learning, social skills and personality development. There is much overlap amongst these areas. In addition, they only hint at the power of mind. Among its many abilities are: qualitative states, intentionality, pattern recognition, problem solving, action based on experience and generalisation, adaptation and dealing with novelty, creativity, curiosity, humour, rationality, freedom of will and justice.

In theory, for mild AI to be completely vindicated it would have to be able to imitate any characteristic of mind. However, this doesn't mean every mild AI system has to imitate all of these characteristics. Clearly, some characteristics are more readily imitated than others, but what characteristics would be acceptable for attributing success to the field of mild AI? The following section discusses in more detail the definition and properties of mind. From this cognition and consciousness arise as important characteristics for possessing a mind.

3.2 Definitions of Mind

Cornerstone to defining and proving whether mild AI is possible is a definition of mind since a strict definition of mild AI has it as the pursuit of man-made models of mind. This aspect will be discussed in the next chapter. For now consider the Encyclopaedia Britannica definition of "mind" as:

> "1) The element or complex of elements in an individual that feels, perceives, thinks, wills, and especially, reasons, 2) the conscious mental events and capabilities in an organism, 3) the organised conscious and unconscious adaptive mental activity of an organism."

The first of these statements is imprecise in that it could be referring to the brain, and the other two are circular since they refer to the mental i.e. that which pertains to the mind. For our purposes a less controversial definition is required.

Reber [Rebe86] categorises five approaches to defining mind. 1) Mind as a model for psychological theories. Again, somewhat circular, but principally taking psychology as abstracting from the biological mechanisms and instead focussing on higher level processes – see Malim and Birch [Mali98]. 2) Mind as the totality of conscious and unconscious mental experiences. Here the emphasis is on *experiences* as defining the mental. This approach is central to dualism – for a review see John Heil [Heil98]. 3) Mind as a collection of processes, typified by those processes studied under perception and cognition (see Harvey Schiffman [Schi90]), and functionalism (see Block [Blc78]). 4) Mind as equivalent to the brain, i.e. brain function. This is physicalism – see J.J.C. Smart [Smrt62]. 5) Mind as an emergent property of a complex (biological) system. This ranges from emergentism (see Searle

[Sear92]) to epiphenomenalism (see Jaegwon Kim [Kim93]). Notice that these categories are not necessarily disjoint. They highlight how mind is sometimes treated as a faculty or thing, as referring to a collection of integrated processes or capabilities, or as a particular system perspective or level.

What is needed is a more precise descriptive definition of mind. This would amount to answering one of the fundamental problems in philosophy of mind concerning what it takes for something to possess mentality – see Fetzer [Fetz96] and the discussion given in Rey [Rey97]. For example, the Encyclopaedia Britannica goes on to consider distinguishing criteria for mind, noting that thought, knowledge or self-knowledge, and purpose, seem to be common to all theories of mind. Four of the main criteria that have been suggested as being unique to mind are as follows.

3.2.1 Mind as Thinking

René Descartes [Desc85] suggested that a mind is solely a thing that thinks. The term *thought* was extended in scope to encompass everything apparent to the mental, including sensory experiences, which were then classified as a form of thinking. A modern definition of thinking positions it as one out of many cognitive processes that make up the mind, where it refers to the covert process involved in the manipulation of mental elements such as words or concepts – see Audi [Audi95]. A definition of thinking should probably mention consciousness; otherwise, it would also be applicable to computational symbolic manipulation of the Newell and Simon variety – see [Nwl82]. This requirement is addressed further in the next two criteria. For a discussion on thinking see Peter Smith and O.R. Jones [Smit86].

3.2.2 Mind as Intentionality

Roderick Chisholm [Chis57] gives an account of Brentano's thesis that intentionality is the mark of the mental. Everything mental is said to have intentionality, which refers to mental states as having an outward direction. They are *about* something and are said to have *content* – see William Lyons [Lyn95]. Martin Davies [Dave98] reviews the five modes of *aboutness* that are commonly distinguished: attitude, experiential, indicator, linguistic, and subdoxastic aboutness. The first mode refers to a class of mental states that involve an attitude toward a proposition, e.g. "I *believe* that *the weather* is *sunny*". Here the mode of intentionality is a belief attitude about the weather. Syntactically, this can be characterised as a propositional expression of the form: "I [mental attitude] that [object] is / has [property]".

The second component of an intentional mental state is its content. In the previous example, "the weather" and "sunny" are labels that refer to the underlying mental concepts. Thus, the concepts of what "the weather" and "sunny" refer to contribute to the content of the mental state. The type of a content depends on the mode of aboutness. The nature of content is more apparent when considering experiential aboutness. For example, the visual perception of an apple has as content the visual experience of the apple in contrast to attitude aboutness which would have the concept of an apple as content – see Searle [Sear83, p61].

Dennett [Denn71] draws attention to what is called the *intentional stance* as a strategy for explaining the behaviour of systems by treating them as though they had beliefs and desires. Searle [Sear92] calls this *as-if* intentionality and distinguishes it from the *intrinsic* intentionality mentioned above. These two types of intentionality are very different in that as-if intentionality does not necessarily have any content. This will prove important later in evaluating the potential success of cognitive science and the future of mild AI.

In his early work, Franz Brentano [Bren74] claimed that "all and only mental phenomena exhibit intentionality." However, Rey [Rey97] develops a computational / representational theory of thought in which the problem of explaining content in physical terms is a central issue. A critical review is also presented of other theories of content, such as co-variational, asymmetric dependencies, teleo-semantic, externalist, and narrow and wide theories. Rey suggests a combination of these theories may explain content in physical terms, although the details of this unified theory have yet to be worked out and the nature of qualitative states remains unsatisfactorily elusive. For further details on grounding content see Michael Devitt [Dev90] on narrow and wide issues, Fodor [Fod87] on psychosemantics, and Paul & Patricia Churchland [Chur83] on calibrating internal states.

Consequently, part of explicating consciousness from a physical viewpoint will involve countering Brentano and analysing the nature of intentionality in cognitive states with respect to the relationships between the various modes of aboutness and types of content, especially experiential content. For discussions in this regard see Fred Dretske [Dret80] and Robert von Gulick [Gul80].

It may well be that a common form of relationship or mechanism underlies all of these different modes and types of content. Consequently, a precise description of these relationships is necessary for judging the feasibility of mild AI. The thesis argues that for mild AI to succeed it must, at least, account for this common mechanism when applied to an exemplary mode of content. This will entail untangling the relationship between consciousness and experiential aboutness and content. This is discussed further in the section on consciousness and intentionality.

3.2.3 Mind as a Semiotic System

As discussed in chapter one, Fetzer [Fetz90] suggests that minds are semiotic systems and presents a means of classifying types of minds. This is based on Peirce's theory of semiotics, characterised by the Sign Relation, in which *something* stands for *something* for *somebody*, highlighting the importance of the observer, the representation and the thing being represented. Semiotic systems use signs that are defined operationally in terms of the habits, dispositions or tendencies they give rise to. This is in contrast to symbols in symbolic systems, as defined by Newell and Simon [Nwl76], which are just anonymous state variables at the system level, see the discussion of levels in the next chapter. Interestingly, Fetzer points out that a criterion for a thing to be a mind is that it has the capacity to make a mistake, that is, to misuse signs – see Dretske [Dret86].

Fetzer [Fetz98a, p384] presents a semiotic conception of consciousness "according to which a system is *conscious* (relative to signs of specific kinds) when it (a) possesses the ability to use signs of that kind and (b) is not incapacitated from exercising that ability." Following this with a conception of cognition: "Moreover, when properly understood, *cognition* is an effect that is brought about by a causal interaction between the presence of signs of specific kinds (within suitable causal proximity) and a system that is conscious with respect to signs of that kind in relation to its *context*, consisting of its other internal states, including pre-existing motives and beliefs."

Again, to judge mild AI requires a precise description of the form of these signs and the relationships within the system. Definitions of consciousness and cognition are considered in more detail shortly.

3.2.4 Mind as Consciousness

To Descartes [Desc85] all aspects of consciousness were a form of thinking and thinking alone was essential to mind. Therefore, mind and consciousness or cognition were in some sense synonymous.

More recently, and from a different tack, Searle [Sear92, p228] has called for an inversion of explanation, suggesting that consciousness is produced as a direct result of brain functioning and it is conceptually wrong to postulate an intermediary mental level. Thus, when people talk about mental processes they are really referring to brain functioning, and talk of the mind is referring to the results of this functioning (consciousness). Searle is doing more than simply defining away mind, a cognitive scientist could not argue that the term *functional* could be substituted for the term *mental*. Instead, Searle is suggesting that there is no place for a level of mind. It puts across the wrong causal relationships (direction of causation). Rather, the system is just intrinsically functional all the way up to consciousness.

Consequently, this leads to a criterion for a thing to have mind-like properties, notably, it possessing some degree of consciousness. Hence, for a mild AI system to have a mind it would have to have some degree of consciousness. In contrast, Rey [Rey97] proposes a computational representational theory of thought in which a theory of “Modest Mentalism” is developed that doesn’t necessitate consciousness, i.e. a mind could exist without consciousness – see below for a discussion of Rey’s theory.

3.3 Unity of Mind

Common to the above criteria for mind is the role consciousness plays, along with mind viewed as a property of a cognitive system, rather than a distinct entity. However, it is necessary to differentiate between the concepts of a unified-Self, the unity of consciousness, and a unified-mind. A unified-Self refers to our perception and conceptualisation of ourselves as the same person from one day to the next, our “capacity to think in terms of the story of one’s past life” – see John Campbell [Cmpb94]. The unity of consciousness refers to the continuity and unity of phenomenal space – see Christopher Hill [Hill91], i.e. every point of consciousness is accessible to the Self – see the next two chapters. In contrast, the unity of mind refers to the idealisation that brains house just one mind, or rather, no more than one mind can exist in consciousness (i.e. a phenomenal space) – see Thomas Nagel [Nag71]. Thus, a definition of mind as in "The Mind", requires at least a concept of Self and a concept of unity, and an explanation as to the nature and role played by consciousness. The nature of consciousness is considered in more detail below.

4 Cognition

Throughout the previous section, cognition and consciousness figured highly in defining mind. This section elaborates on the nature of cognition whereupon its centrality to mind, consciousness and therefore mild AI, will become apparent.

4.1 Defining Cognition

The word cognition comes from the Latin for "to become acquainted with, to know." The Encyclopaedia Britannica describes it as:

> "The process involved in knowing, or the act of knowing, which in its completeness includes perception and judgement. Cognition includes every mental process that can be described as an experience of knowing as distinguished from an experience of feeling or of willing. It includes, in short, all processes of consciousness by which knowledge is built up, including perceiving, recognising, conceiving, and reasoning. The essence of cognition is judgement, in which a certain object is distinguished from other objects and is characterised by some concept or concepts."

Echoing the opening section of this chapter on animal life, Terry Winograd and Fernando Flores [Flor86, p47] quote Maturana on how cognition should be understood as relating to living systems and their environment: "A cognitive system is a system whose organisation defines a domain of interactions in which it can act with relevance to the maintenance of itself, and the process of cognition is the actual (inductive) acting or behaving in this domain."

In general the above definitions of cognition are reasonably clear-cut; the difficulty lies more with the domain of cognition and whether or not it includes consciousness. This issue is discussed in the following sections.

4.2 Cognitive Science & Cognitive Psychology

There are two fields of study that are concerned with explaining cognition. The first, cognitive psychology, treats cognition as referring to those aspects of behaviour that are controlled by the higher centres of the brain, the cerebral cortex – see Anthony Sanford [San91]. With its roots in psychology, it seeks explanations in terms at the psychological level of description. The second, cognitive science, is a broader discipline and additionally includes lower level implementational descriptions, incorporating contributions from neuroscience and computer science – see Neil Stillings et al. [Stil95].

In the early days of these cognitive fields, emphasis was placed on studying human cognition rather than animals in general – see von Eckardt [Ecka93]. What is now called the classical view was based on the working hypothesis that cognition can be modelled as a representational information-processing system, and that the core of the cognitive architecture is a physical symbol system. In more recent times complementary cognitive architectures have arisen, such as connectionism and dynamical systems – see Andy Clark [Clar89] and Timothy van Gelder [Geld95] respectively.

4.3 The Domain of Cognition

The classical cognitive processes that have been studied include attention, perception, memory, language and thinking. The additional inclusion of consciousness has only recently become more popular with cognitive scientists. Here a brief summary is given for the main aspects of these classical processes – see Stillings et al. [Stil95].

Investigations into attention are concerned with how the brain allocates its finite amount of resources for performing tasks; important considerations are selection, filtering, capacity, limits and resource allocation. With perception emphasis is on pattern recognition applied to sensory modalities, particularly vision; important considerations are feature detection, classification, context, and compound patterns. Different kinds of memory have been postulated: long-term versus short term, sensory (vision etc.), language, semantic, conceptual, and procedural versus declarative. Upon this various operations are performed: chunking, association, random access and networks of activation. Language considerations cover grammar, phonology, syntax and semantics of speech and thought, discourse and its relation to memory, attention and problem solving. Finally, thinking deals with problem solving, classification, interpretation, decomposition, representation, heuristics, biases and search control. This also includes discovery, learning, understanding, mental models, intuition, deductive reasoning and imagination.

4.4 Cognition & Consciousness

At one time in the history of cognitive science, consciousness and cognition were treated as being distinct from one another and so permissibly studied independently. However, a treatment of consciousness has now become more central – see Alvin Goldman [Gol93] for a discussion on its importance. Indeed, Searle [Sear92, Sear98] suggests it is an integral and inseparable part of the mind and cognition. As mentioned above, Searle effectively replaces mind by consciousness. Tim Shallice [Shal98] reviews the debate as to whether consciousness is essential to cognition from a systems viewpoint. The review highlights that although early arguments, based on phenomena such as blind-sight and knowledge without awareness, were inconclusive, a variety of theories have been proposed that usefully incorporate or require a version of consciousness in a high-level information processing role. Some of these theories, such as Bernard Baar's, are reviewed below.

If consciousness could be related to cognition in a computational way, it would imply mild AI was possible. However, incorporating consciousness into cognitive science is turning out to be a difficult task. While the classical cognitive processes, mentioned above, are proving to be computationally complex, they are felt to be ultimately tractable to the information processing methodology or more recent approaches such as connectionism – see Block [Blc95]. However, in comparison, qualitative consciousness in particular, is proving to be much more resilient to investigation. Some have suggested it may even be an intractable "hard problem" – see Shear [Shea97]. Central to this problem is intentionality and content. The hard problem is discussed in more detail below.

4.5 Cognition & Mind

Ryle [Ryl49] draws attention to how the primary purpose of mind is often solely equated with cognition and the exercise of finding answers to questions. This brings to the forefront reasoning and language as tying together mind and cognition. On the one hand reasoning seems to be based on some kind of language of thought, as suggested by Fodor [Fod75], and on the other, minds explicitly use languages for communicating with others and internally – see Noam Chomsky [Chom87]. Quite quickly this leads to an analysis of language in which the language terms used often refer to mental particulars – see Tim Crane [Cran98]. These are internal mental phenomena that ordinarily have no spatial extension but some kind of temporal existence, such as a particular pain or thought. It is not stipulated whether mental particulars actually exist, for example, there is not necessarily a mental object corresponding to a pain. One thing these particulars do appear to have is intentionality and content.

4.6 Cognition & Intentionality

Charles Dunlop and Fetzer [Dun93] define "cognition" as any instance of a mental operation that displays intentionality. Stillings [Stil95, p342] describes the classical cognitive architecture view as being based on syntactically structured representations that have intentionality, and in the current context, the same could be said of the more recent connectionist approach. However, Searle [Sear92, p78] points out that even when these representations are causally and semantically embedded in an information-processing system, they can only be said to have metaphorical as-if, rather than intrinsic intentionality. Consequently, explaining intrinsic intentional content without resulting to epiphenomenalism is currently problematic for the information-processing system view.

5 Consciousness

An empirical analysis of consciousness would be required to determine the ultimate feasibility of mild AI. The thesis attempts to provide a partial analysis that explains a fundamental part of consciousness enabling a degree of plausibility to be attributed to mild AI. As a precursor to this the following presents denotative definitions and a meaning analysis for consciousness.

5.1 Defining Consciousness

The Encyclopaedia Britannica picks out a number of meanings for the term "consciousness" as follows:

> "1) a: The quality or state of being aware especially of something within oneself. b: The state or fact of being conscious of an external object, state, or fact. c: Awareness; especially: concern for some social or political cause. 2) The state of being characterized by sensation, emotion, volition, and thought: mind. 3) The totality of conscious states of an individual. 4) The normal state of conscious life. 5) The upper level of mental life of which the person is aware as contrasted with unconscious processes."

When one talks about explaining consciousness, it is extended to cover most of the domain of cognition as well, and just about all the other hard AI problems. Thus, today the word is used in so many ways that it has to be qualified to be meaningful. Lycan [Lyca96] lists eight different senses. In this taxonomy, qualia belong to subjective consciousness. Norton Nelkin [Nel93] suggests consciousness consists of at least three types: phenomenal, intentional and introspective states.

For the purposes of the thesis the distinction made by Edelman [Edel89, p24] is useful. In a slightly revised and neutral form one can distinguish between primary and secondary consciousness:

> "Primary consciousness may be considered to be composed of certain phenomenal experiences such as mental images, but in contrast to secondary consciousness, it is supposed to be bound to a time around the measurable present, to lack a concept of self and a concept of past and future, and to be beyond direct individual report."

Edelman uses the term 'higher-order' rather than secondary. The latter is preferred since it is neutral with respect to implementation levels – an important aspect considered in the next chapter. Secondly, this distinction should be thought of as characterising two points along a continuum rather that demarcating a dichotomy. It is suggested that the number of qualitative experiences that make up the consciousness of different creatures lies on a continuum. However, due to the structuring effect induced when signifying reality, there will be gaps in the population of realisable creatures along the continuum – see Aaron Sloman [Slo97].

Primary consciousness is defined to be that which deals with the phenomenal basis of consciousness, whereas secondary consciousness is more concerned with cognition. Sentience is a primitive form of primary consciousness involving uninterpreted sensation – see Dennett [Denn96, p84ff]. In what follows, primary-consciousness is examined rather than sentience. As was mentioned above, explaining primary consciousness is the harder problem facing the mild AI approach. Therefore, to determine whether mild AI can succeed the thesis focuses on those aspects of consciousness sufficient for producing primary consciousness.

5.2 Perception, Sensation & Consciousness

To avoid confusion this section states the meaning of perception and sensation used in the thesis. Malim and Birch [Mali98] suggest “sensation, …, is the primary process of data collection from the environment. Perception is the secondary process of interpreting these data.” Historically the senses were thought to supply the higher processes with sense-data, such as colour patches and audio tones, which were building blocks for more complex objects. Nowadays the senses are thought to merely act as receptors for physical stimuli whereupon signals are relayed through neural pathways to the perceptual systems. Therefore, it is at this higher level, following perceptual interpretation, that conscious sensations (i.e. qualia, see below) are said to arise. Jonathan Westphal [West87] comments “I agree with Ryle that sensation is actually a kind of perception, and not the other way round.” For a logical analysis of sensations see Austen Clark [Clar93].

5.3 Basis & Reason for Consciousness in Nature

What is the purpose of consciousness? Horace Barlow [Bar87, p366] suggests consciousness arose as a means for social discourse - even with one’s Self. Similarly, Nicholas Humphrey [Hum87] suggests consciousness evolved so that we could explain in simple terms our bodily state to others. Davies and Humphreys [Dave93, p8] suggest “the experience of conscious recollection is thought to be a prerequisite for intentional action, and it is intentional action that enables people not to operate in an overly data-driven way.” To quote David Armstrong (see Borst [Bors70, p79]), “And so consciousness of our own mental state becomes simply the scanning of one part of our central nervous system by another. Consciousness is a self-scanning mechanism in the central nervous system.”

5.4 First & Third Person Accounts of Consciousness

Perhaps the distinguishing feature of consciousness is that experiences are subjective – see Searle [Sear92, p94]. That is, the content of mental experiences, such as thoughts, pains, and sensations in general, are private events that are experienced exclusively by the individual. No one else can directly know what another's experiences are like – see Thomas Nagel [Nag74]. An explanation from a first person perspective concerning these subjective experiences involves descriptions based on comparative qualitative terminology, such as bitter versus sweet, hot or cold – see Clarke [Clar93]. However, descriptions from a first person perspective can never get beyond this qualitative level to the mechanism behind. In contrast, explanations from a third person perspective use objective terminology and insights from the neuro sciences enabling the nature of the underlying processes and properties of experiences to be described. For an overview of these different approaches to analysing consciousness see Guven Guzeldere [Guz97].

5.5 Intentionality & Consciousness

A vexing question is whether consciousness is necessary for intentional states. Searle [Sear92, p130] argues that through a liberal definition of intentionality, all consciousness is intentional, and conversely all intentionality is aspectual i.e. has a conscious component. This leads Searle [Sear92, p154] to suggest that certain unconscious intentional states with a certain aspectual shape must be treated as being intrinsically mental. Hence, any definition of "mental states" must encompass both conscious and unconscious states.

In contrast, Nelkin [Nel93] suggests a finer analysis will reveal that not all intentional states need have a conscious component; examples involving blind-sight are given to support this. To clarify matters Nelkin distinguishes three kinds of consciousness: "C1" (first order) intentional state awareness, "C2" introspective (second order) awareness of intentional states, and "CN" phenomenal states as alluded to by Nagel. Nelkin suggests it is only through introspection that we come to associate CN experiences with C1 states.

The distinction made by Nelkin as to the three kinds of consciousness is particularly important for it suggests a further refinement. Common to each of these kinds is a notion of consciousness. In other words, it suggests there is a common form of mechanism that makes each kind a conscious state. If so, for mild AI to succeed it would have to address this common factor. Hence, one task for the thesis is to explain the nature of this common mechanism and how a specific kind of consciousness fits in, such as qualitative consciousness (the N in CN).

6 Qualia

The prior discussion on the character of cognition and consciousness had a recurring theme, the qualitative nature of consciousness. The philosophy of mind has been pondering this aspect of consciousness for many years. Discussions in this regard eventually touch upon something philosophers call "qualia". To some philosophers qualia are central to explaining qualitative consciousness. To them this is what mild AI would ultimately have to conquer. This section introduces the term. It then goes on to state what is thought to be the main problem to be answered by any complete theory of consciousness.

6.1 Defining Qualia

The Latin term *quale*, plural *qualia*, forms the root of words such as *quality*. Its origin stems from the contrast made in conjunction with the term *quantum* when characterising the Universe. The Universe can be thought of as being made up of certain *quantities* of things, such as mass and energy – see also Tim Crane and David Wiggins [Cra98] for a discussion on universals and particulars. In addition, it has *qualitative* aspects, such as colour and beauty – see Richard Gregory [Greg89]. Science has largely just dealt with the quantitative aspects.

There are three main ways in which the term "quale" is used:

1) Scientific use with respect to sensory perception and, ironically, quantifying qualia – see Clark [Clar93] and Schiffman [Schi90].
2) To refer to, or label, the sensory experience of the mental.
3) As demarcating a kind of substantive entity or property outright.

Lycan [Lyca96, p69, p175] attributes the first modern philosophical use of the term in the sense of 2) to C.I.Lewis (1929) and C.S.Peirce (1898), in which a quale is the irreducible "introspectible monadic qualitative property of what seems to be a phenomenal individual," for example, the experience of colour, such as blueness, as opposed to a blue object. This pure phenomenal experience, which is referred to as the blue quale, appears unattached to any underlying mechanism. It is experienced as though being completely substantial and far removed from neural firings.

There has been much philosophical debate about the existence of qualia as in sense 3). For instance, Dennett [Denn88] suggests they cannot be *properties* of things, as such their existence should be denied. That is, a bar of gold has the property of having a certain weight. Cut the bar in two, and the pieces will still have a weight, but half as much. Qualia are not attributes or properties of an object in this sense. For example, cut an aeroplane in two, and it no longer has the potential property for flight. Related views include sensationalism in which sensations are treated as building blocks for all other cognitive states – see Audi [Audi95]; the monads of Leibniz – see G.H.R. Parkinson [Park88]; the secondary qualities of Locke – see John Locke [Lock75]; and the tropes of Williams – see Keith Campbell [Cmp81].

To clarify matters, the position adopted by this thesis is stated as follows:

Statement of Logical Reason for Qualia. (S-LRQ)

The term 'qualia' is used to refer to the experience of mental qualities arising in a suitable system.

This is sense 2), the least contentious sense. What is meant by a suitable system is explained in the next chapter. The statement remains neutral with respect to what qualia are. It leaves it open as to whether they are properties of things or complex processes. Herein, the term is used as a label to refer to the experience of particular mental qualities, such as the blue quale. For a discussion on qualia and consciousness see Owen Flanagan [Fla92].

6.2 The Mind-Body Problem

In the production of sensory experiences a difficulty arises because somewhere along the way physical neural activities produce phenomenal content bearing experiences. The physical gives way to the mental, mind arises from matter. The mind-body problem in its purest form is a question of what is the link between the mental and the brain: How does one arise from the other? The question is posed in a neutral way with respect to what causes what. For the purposes of the thesis it is presented as a deceptively concise statement:

Statement of Mind-Body Problem. (S-MBP)

The mind-body problem is concerned with the nature of the system relationship between the mental and the physical.

The keyword *system* reflects the perspective adopted here that a complete system is being considered.

In its extreme form, the mind and body are taken to be distinct systems with some kind of connection that correlates mental and brain states. The mental denotes the non-physical and the brain the physical. Problems then arise when the principles of physical energy conservation and causation are applied. How can the two systems influence each other and uphold these principles? This and all its implications is at the centre of the mind-body problem. Nowadays, this extreme Cartesian dualistic form of the problem has given way to materialistic mentalism – see David Rosenthal [Rsnt91], McGinn [McGi91] and Lycan [Lyca90]. The problem is now felt to be one of definition, with the notions of mind and the relationship between mind and body needing to be explained.

6.3 The Quale Problem

The thesis is concerned with a more specific form of the mind-body problem, namely, What is the mechanism by which qualia arise from the system? This section states this and then goes on to discuss related issues.

Part of the illusion of consciousness is that the world appears 'bright', so reassuringly sunny. How can neural grey-matter give rise to this? To borrow an example from Dennett [Denn91, p108], if afferent nerve signals were followed through the brain from the eyes, at what point do conscious qualia appear? Surely before the signals become efferent (outbound)? The quale problem is a more refined form of the mind-body problem. In particular, how do specific non-physical mental qualia, that have a particular sensory experience, arise from physical brain processes? A more general form of this problem is known as the hard problem, or the problem of experience – see David Chalmers [Chal96], and Shear [Shea97]. For the purposes of the thesis the following is stated:

Statement of Quale Problem. (S-QP)

The quale problem is concerned with the nature of the system mechanism that gives rise to specific experiences of mental qualities.

For example, what differentiates the mechanisms that give rise to red qualia as opposed to blue qualia, or sounds, or pain? The mind-body problem is deemed to be more general than the quale problem in that not all mental processes are said to have a qualitative aspect. One consequence of the thesis developed here will be to suggest that these two problems are facets of the same problem.

6.3.1 Quale Problems

The quale problem raises a number of topical issues in the philosophy of mind. Lycan [Lyca96, p5] itemises a dozen related problems: a) the subject and object distinction, b) immediate or at least privileged access, c) temporal and other empirical anomalies, d) how or why did consciousness evolve, and what is it for?, e) epistemology, f) inverted qualia and absent qualia, g) homogeneity or grainlessness, h) the monadic, first-order qualia of apparent phenomenal objects, i) the intrinsic perspectival point-of-view and, or first-person aspect of experience, j) funny facts, k) ineffability, and l) the "explanatory gap". Westphal [West87] presents an analysis of Wittgenstein's "Remarks on Colour" wherein specific problems relating to colour qualia are raised: "a) something can be transparent green or any other colour, but not transparent white, b) white is the lightest colour, c) grey cannot be luminous, d) there cannot be a pure brown or brown light, e) there is no blackish yellow, and f) there can be a bluish green but not a reddish green." See also Clark [Clar93, p147].

Why is redness experienced the way it is compared to blueness? Could the two be interchanged? In this regard, Dennett [Denn88] discusses the inverted quale dilemma. Peter Hacker [Hack87] notes that a plank cannot be both 1m long and 2m long. Perhaps the quale of red and blue colours is ordered in a similar way? Might someone else's experience of redness be different to yours (see Hill [Hill91] for a discussion of the other minds problem)? John Biro [Biro93] suggests we could know everything physical there is to know about a quale, nothing would be left out. Even so, we would still not know the experience until we actually experienced it. This is called the knowledge-argument problem – see F.Jackson [Jack82].

Most questions on qualia tend to centre on colour, vision and visual imagination. The other sense modalities (sound, smell etc.) are given much less attention. Nevertheless, the quale problem applies equally to these modalities. As will become apparent, questions on how one modality is distinguished from another need to be addressed. Clark [Clar93] considers sensory qualities across modalities. Notice that all these approaches treat qualia as irreducible perceptions.

6.3.2 Qualia Dependent Problems

The quale problem is a springboard to a number of other important issues in cognitive science and the philosophy of mind. Alan Code [Lepo91, p105] lists four serious problems: consciousness, intentionality, subjectivity and mental causation.

Donald Norman [Aitk90, p322] lists twelve, including belief systems, emotions and perception. All these problems depend fundamentally on a formal solution to the quale problem.

Davies and Humphreys [Dave93, p14] suggest that the problem of consciousness can be broken down into three areas: a) the notion of phenomenal consciousness, b) states with intentionality, and c) access consciousness type states. They suggest b) and c) hinge on a), although solutions to b) and c) might help reveal a solution to a). While they make a distinction between phenomenal consciousness and access consciousness, in reality, access consciousness, to be conscious at all, must be dependent on phenomenal consciousness.

6.4 Epiphenomenalism & Qualia

Phenomenalism is the view that only thoughts and qualia exist. Reality appears to us through this limiting perspective – see L.E.Goodman [Gdm92]. Robert van Gulick [Gul93] remarks that the idea of phenomenal structure of experience was introduced by Kant "in the context of rejecting the sensational theory of experience associated with traditional empiricism." Thus, phenomenal experience is more than just basic qualia, "but rather the organised cognitive experience of a world of objects and of ourselves within that world."

Somewhat in contrast to phenomenalism is epiphenomenalism, the view that qualia are merely non-causal by-products of the brain. If qualia are simply epiphenomena belonging to another dimension (information or mental space), how can the physical mind driving the information space know about them? One can introspectively experience and contemplate the red quale, which implies the underlying physical driving processes know of the existence of qualia, but how, if they are in a different dimension?

Qualia tend to be lumped together, as all being epiphenomena, or the thorn in the functionalist's model. Is it safe to assume that this is so? What if a functional explanation for a pain qualia was found? Would it have to be excluded as a quale? In a similar way to the computational hypothesis of $N \neq NP$, the quale problem might not be 'quale-complete'.

6.5 Intentionality & Qualia

Stephen Stich [Sti92] says that naturalising (i.e. explaining) intention is basically the mind-body problem and why is ‘red’ red, or a ‘C sharp’ a C sharp? Joseph Levine [Lev93] suggests that “Since the problem of qualitative character turns out to be primarily epistemological, the source of which is to be found in the peculiar nature of our cognitive representations of qualitative character, a theory of intentional content ought to explain what makes these representations so uniquely resistant to incorporation into the explanatory net of physical science.”

Searle [Sear83, p35] considers the qualitative composition of intentional states and how all intentional states would appear to contain elements of belief and desire. Searle (p43) comments that there is more to intentional states than just their intentional contents. The propositional attitudes have a qualitative component and a disposition content, e.g. the belief and desire of love, the experience of love, and the disposition actually discriminating a state as being about ‘love’. Hence, the problem of intention is dependent on, or at least related to, the quale problem.

6.6 Cognition, Consciousness & Qualia

It was mentioned above that at one time cognition and consciousness were treated as distinct entities. More recently consciousness is referred to from two standpoints: when referring to qualitative experiential mental states, and when referring to the conscious mental life of an individual. Both of these have qualitative experiential aspects and therefore relate to qualia. The second standpoint is typically treated as cognition embellished with qualitative states that border on the epiphenomenal, their function being uncertain. This uneasy situation could have arisen because of the representational basis of the theory, or it could be seen as a feature of the theory and a dismissal of qualia. The problem for mild AI is that a definition of mind may require qualitative consciousness and an explanation as to how qualia are generated. Unfortunately, the theories that suggest qualia are incidental fail to give a reason why they should be experienced at all. The task for cognitive science is to explain why qualia are generated irrespective of whether or not they are functionally redundant. Consequently, one strategy is to show that qualia have a role to play in cognition.

7 Review of Theories of Mind and Consciousness

This section reviews a range of the more popular theories for mind and consciousness that refer to qualia in order to assess the contemporary perspective toward qualitative consciousness.

7.1 Theoretical Frameworks for Mechanism of Mind

Theoretical frameworks for the mechanism of mind range from the biological through to the symbolic. The first extreme suggests that either neuro-biological processes are necessary or a causal equivalent is required to instigate mind – for a discussion of this view see Ted Honderich [Hond90] and Edelman [Edel89]. The symbolic extreme suggests that the manipulation of symbols is sufficient for mind – see Turing [Tur50] and Newell and Simon [Nwl76]. This extreme is discussed in more detail in chapter six.

These extremes are spanned by three principle approaches: dynamical systems, connectionism, and the computational theory of mind. According to Gelder [Geld95, p9] dynamical systems are typified by trajectories through a fixed dimensionality numerical phase space, whereas computational systems involve trajectories through different symbol sequences. Quantum mechanical approaches would be classified under dynamical systems. Connectionism falls in between and involves distributed representations across adaptable networks of simple processing units that are modelled on real neural networks – for a discussion of this view see Bechtel and Abrahamsen [Bech91].

The chief difference between these approaches lies with the nature of representation used in mental processes. Representational theories of mind suggest mental processes use representations of the world, and these representations have semantic content – for a review see Sterelny [Ster90]. A weak version treats representational content as being epiphenomenal. In contrast, syntactical theories of mind postulate that mental processes can be explained in terms of syntax and causal profiles with no reference to content – see Stich [Sti83]. This leads to computational / representational theories of thought (see Fodor [Fod75]), according to which representations are used, but in a purely syntactical way. Some prominent computational theories of consciousness are reviewed below.

7.2 Perceptual Field Based Theories of Consciousness

In almost all theories of consciousness there is, in some form or other, what is called the perceptual field (defined later, also called the blackboard, global workspace, and the Cartesian Theatre). In some cases, the theory is independent of the exact nature of the perceptual field. However, in others the influence of presentational concepts and terminology is reflected in the theory's view of consciousness.

7.2.1 Baars' Cognitive Theory

Baars [Baar88] develops a cognitive theory of consciousness. In this, a broad definition of "consciousness" is used in which it is treated as a distinct system to the rest of brain functioning, i.e. the 'unconscious' with a strong emphasis on cognition. The brain is composed of interacting subsystems, one of which is conscious and self-controlled. Central to the theory is the "Global Workspace", a common area where messages are relayed and broadcast, it has a strong presentational character. Here is where consciousness is said to reside (p104), although details are not given as to what makes something conscious. From this perspective it is suggested (p76) "Conscious processes have a great range of possible contents," and therefore offers an evolutionary selective advantage. This is in contrast to the functioning of unconscious components, without the intervention of conscious control, which it is suggested lead to certain drawbacks and possible errors in the operation of some tasks.

Exploring the constraints of consciousness, Baars (p83) considers the serial nature of consciousness, and suggests that we cannot have two different thoughts at the same time and be conscious of them. How much is this due to our environment and image of the Self as one? Language imposes a serial order, so words and thoughts have a sequentially defined meaning, i.e. defined on a serial grammar with hard-wired semantics. Finally, Baars (p86) makes the point that "Consciousness processes are computationally inefficient" when understood on certain accounts. When something is worked out consciously, it forces a conscious representation that is anchored by environmental constraints. Reasoning can become susceptible to errors since it involves manually applying an appropriate procedure. This paraphrases Baars' view of the brain as subsystems, the conscious subsystem being a limited sequentially oriented device, perhaps a symbolic manipulator, a recent evolutionary adjunct.

7.2.2 Edelman's Biological Theory

Edelman [Edel89, Edel92] develops a biological theory of consciousness. The theory for primary consciousness can be summed up by four main points [Edel92, p120-121]: a) Self and non-Self components, b) value-category memory, c) real time, in parallel and for each sensory modality, and d) re-entrant connections between conceptual and ongoing perceptual systems. Notice that point b) is a perceptual field with a presentational orientation. The importance of re-entrant connections and memory are repeatedly emphasised. Edelman [Edel89, p155-162] suggests the emergence of primary consciousness "results from the interaction in real time between memories of past value – category correlation's and present world input as it is categorised by global mappings (but before the components of these mappings are altered by internal states)." Qualia are described as "forms of higher-order categorisation, as relations reportable to the self and reportable to others..." [Edel92, p116]. However, this leaves much unexplained, for instance, why is the red quale, red?

Edelman [Edel89, p187] considers the need for a mental language as a prerequisite for secondary consciousness, suggesting there is a need for "a symbolic representation of the self acting on the environment and vice versa." By symbolic, Edelman means something identified with, or labelling a thing, but without any physical connection with its definition in the implementation. While language may indeed be a requirement for higher order consciousness of certain varieties, it may not be a requirement for consciousness. A system could be devised which handles planning etc., in a symbolic sense, but does not break through to consciousness. To investigate whether mild AI can succeed in some form, the thesis extends these ideas of Edelman, by formalising and explaining in detail the nature of the mechanism behind qualia.

7.2.3 Dennett's Social Theory

Dennett [Denn91] proposes a theory of consciousness from a philosophy of mind perspective, and remarks (p108) that some have supposed there must be a point between the afferent and efferent neurons at which consciousness occurs. The correlation of pain and C-fibre firing is often given as such an example – see Levine [Lev93]. This implies what Dennett (p165) calls the Cartesian Theatre, a private screening of reality for the mind's eye. Dennett dismisses this view and the existence of qualia. Instead, human consciousness is simply explained as being the result of a huge complex of memes, which are ideas, concepts or information floating around in society. However, this does not explain what makes them conscious; for example, why is the red quale, red?

"Conscious human minds are more-or-less serial virtual machines implemented - inefficiently - on the parallel hardware that evolution has provided for us," suggests Dennett [Denn91, p218] in support of strong AI. However, not all of consciousness is to do with serial thought. For example, forever present qualia, and vision in particular, are parallel processes. Finally, attention is drawn to how the brain does not have to actually fill in some aspects of reality, e.g. the blind spot. "The discontinuity of consciousness is striking because of the apparent continuity of consciousness." And quoting Minsky (p356), "Nothing can seem jerky except what is represented as jerky." This is an important point; the brain has to reconstruct every aspect of reality. This feature suitably generalised and characterised, is an essential requirement for a semiotic system as will be seen.

7.3 Computational Theories of Consciousness

This category of theories is particularly relevant to determining the fate of mild AI since they are based on underlying computational mechanisms. If one of these theories succeeds in explaining qualia, then it can be concluded that mild AI is able to accommodate one of the harder aspects of mind.

7.3.1 Rey's Computational / Representational Theory

Rey [Rey97] presents a computational / representational theory of thought and qualitative states (CRTQ), in which qualitative experiences are accommodated within the language of thought hypothesis. Thus (p308), "qualitative experience is just a particular species of propositional attitudes; and propositional attitudes are computational relations to representations in fairly integrated computational systems." Quoting Davies and Humpreys' summary of Rey [Dave93, p240]:

> "... we can include sensations within [a language of thought] picture in two steps. First, we suppose that the language of thought contains certain predications (meaning roughly: its looking red, for example) to which a subject stands in the computational relation corresponding to judging only when (or normally only when) the predication is tokenised as a direct result of output from sensory systems. Second, we suppose that tokens of these predications cause characteristic subsequent processing, and we identify

> sensory experiences (of something looking red, for example) with instances of this processing."

Rey gives definitions for relating (mental) propositional attitudes to computational operations in which computational and semantical issues are separated. This is denoted by prefixing the computational aspects with the word 'comp' in the discussion. While these definitions tell us something about what is involved, they are not operational definitions, i.e. they contain terms which need unpacking and further operational denotations. For example (p245), having defined 'sensing', an applied example is given:

> "A red sensory experience would involve comp-judging a restricted predication, 's(R)', as a result of the stimulation of predominantly L-wave sensitive cones, a comp-judgement that, by virtue of the predication's characteristic processing, produces further comp-judgements of 'warm', and 'advancing' predications."

Clearly, a more detailed account would be needed to implement this.

Rey suggests that many people make the mistake of thinking functionalism is largely about dispositional states, whereas they are 'occurrent' states. Thus (p247),

> "A sensory state ... is fully activated. ...they are best viewed not as single states but, ... as *processes* involving interactions among a variety of cognitive states. A qualitative experience is presumably a process involving the comp-judging of a certain restricted predicate, a comparison of it with certain memories, involving restricted and unrestricted predicates and other associations..."

Rey notes that just how much of such a process is required for having the sensation, in a wide and narrow sense, is an interesting issue. For example, it affects whether people can be said to have similar red quale, or if the experience of the quale is dependent on too much of their personal experience to be generalised. This is basically what the requirements for a semiotic system discussed in chapter four predicts and where it picks up from.

Finally, Rey makes a good point with respect to the 'lack of grain' in a quale sense. "What you 'see' is simply what your restricted predicates represent, no more, no less." This relates to the binding problem and qualia, such as continuity qualia for space and time. Are these qualia needed if all that we can sense is that which we have qualia for? It seems as though they are not needed.

7.3.2 Johnson-Laird's Theory of Models and Parallelism

P.N.Johnson-Laird's [John83, p473] theory suggests an important representational mechanism for cognition is the formulation of mental models, such as models of our Selves. A precise model is not necessary, nor is conscious knowledge of it. The only restriction being that, structurally, "the actual algorithm for consciousness … must be a parallel one." Parallelism is used in two senses, the model must contain models of itself that are accessible in parallel, and secondly, a Turing Machine being a serial device could only simulate the parallel model. No maintainable reason is given for this when one considers that a dependence on parallelism implies the model contains events that are causally simultaneously bound; in effect, a form of quantum mechanical action at a distance, an effect that this thesis denies is necessary.

7.3.3 Chalmers' Functionalist – Dualist Theory

Chalmer's [Chal96, p249] claims that "consciousness arises from functional organization but is not a functional state," which is called "nonreductive functionalism", a combination of functionalism and property dualism. As a representational basis for a fundamental theory, information spaces are suggested whose structure is determined by the combinatorial and relational structure of the subspaces (p276). Information can be discrete or continuous. An abstract information space is mapped to a physical system according to Bateson's slogan: "information is a *difference that makes a difference*."

Chalmer's theory is essentially a representational cause and effect model, in other words a variant of the state machine model. It focuses on the instantaneous structure of the phenomenal space, the logical structure as Clark [Clar93] would say. As Chalmer's (p235) admits, this does not explain the intrinsic nature of experiences, i.e. qualia, and so would not support the case for mild AI. In fact, there is no reason to believe that the state machine model at this level of description has qualitative experiences at all. Indeed, in a critique of the theory, Searle [Sear97, p156] points out that this leads to panpsychism: even thermostats would have some level of consciousness. The argument against state machine models is that they fall prey to the spatial-temporal limit problem – see chapter three. Briefly, in the instantaneous limit of structure with respect to time they imply the right kind of stationary structure, such as a certain picture, would be conscious. The requirements for a semiotic system developed in chapter four will show that qualia are more closely related to processes than stationary representations.

7.3.4 Aleksander's Iconic Theory

Aleksander [Alek96] has proposed a computational theory that emphasises neural state-machines. Central to the theory is how inner models of the world are built up through iconic learning. This is a form of learning for producing internal representations that preserve the functional structure of that being modelled. For related approaches to learning see Edelman and Finkel [Edl85]. Aleksander briefly touches upon qualia where, following Dennett [Denn88], it is suggested that if they were substantive things (cf. tropes) they should be dismissed. Rather, they are explained as internal representations produced through iconic learning that preserves the functional and logical structure discriminated during sensory perception. This is the functionalist perspective. Consequently, further elaboration is required, in order to fend off qualia being treated as epiphenomenal, and determine whether mild AI can succeed.

Another recent computational theory is that of Ray Jackendoff [Jack87], which emphasises modularity.

7.4 Quantum Mechanical Theories of Consciousness

The apparent intractability of consciousness to understanding has spurned some to suggest this may be because the mechanism relies on effects from quantum mechanics. These theories start by showing how the brain could operate in a quantum mechanical way and how mental states might map to quantum mechanical ones – see Henry Stapp [Stap93], and M.Jibu and K.Yasue [Jibu95]. Quantum computation is based on the quantum gate analogy of a classical logic gate, which is able to handle a superposition and entanglement of states – see A.Ekert [Eke93].

The brain's possible use of quantum mechanics has been proposed in an anti-mild AI argument. Penrose [Penr94] suggests that Godel's theory implies consciousness relies on a mechanism that is non-computable and not even amiable to simulation on computers. Otherwise, it is suggested, this would mean an algorithm could produce qualia, and a computation could experience mentality (p42). Penrose suggests the only alternative is an as yet unknown non-computable property of quantum mechanics. However, this ignores the significance of levels of description. For example, dynamical systems are not concerned with computing and instead evolve according to equations of motion. It could be that consciousness is the result of a process, which can be simulated, rather than a computation, and so the argument against mild AI would be less forceful.

7.5 Self-Awareness Theories of Consciousness

These theories place self-awareness as essential to explaining consciousness. They commonly involve, in some fashion, the availability of models of the Self to consciousness. However, it is doubtful that self-awareness is needed at all to explain qualia. For example, it is easy to imagine a zombified individual with no sense of Self, but who is still conscious. Indeed, Edelman's definition of primary consciousness has no self-aware mechanisms.

Israel Rosenfield [Rsnf92, p8] supports the idea that self-awareness is the key to consciousness. Self-awareness is distinguished from perception, which is taken as something different but perhaps dependent on self-awareness memory. Qualia are deemed to be fixed (p49) "by the dynamic qualities of [the] body image." However, not all qualia are grounded by a body image. A number of other interesting comments are made. For example, (p83) discusses coherent stimulus responses. "Hence, it is not the individual coherent response [as suggested by Edelman] that is important but the relation of different coherent responses to each other. ... It is the very process of change that rises to consciousness, that is consciousness; awareness is change, not the direct perception of stimuli." This is a subtle point. Consciousness is a temporally dependent process. It is integrally a process of change.

Rosenfield (p104) suggests "The brain creates qualities - the colours, sounds and other sensations we are conscious of - by establishing relations among stimuli." Categorisation is then linked strongly with language. "In the same way, notions of 'big' and 'little' that a child appears to acquire at about the age of three and a half are not inherent characteristics of the stimuli but abstractions that are only possible with, and that necessitate, words." One interesting point is that a child uses words about size at around three and a half years old, and words about colour six months later, suggesting colour is more abstract. This leads Rosenfield to the suggestion that language in some way enhances consciousness. However, it is important to note that qualitative consciousness does not rely on language!

Rosenfield (p120) suggests "some kind of symbolic representation" is needed in order to have ideas about things, cf. secondary consciousness. However, is there a sense in which we have a qualitative experience that some things are of the same kind, e.g. reds and blues belong to 'colours'? To understand an image must we innately know that different patches of colour represent the same quality? Finally, Rosenfield (p127) notes how perception depends on our accumulated experiences. Again, the conscious experience of reality must not be confused with the secondary conceptions of things so perceived. As colour is a secondary property, so are concepts about reality third-order properties.

7.6 Theories with Consciousness as Higher Order Thought

These theories suggest a mental state is conscious if there is a second mental state that has it as content. There are still many questions to be answered by these theories, and many terms that go unpacked. The following presents a brief review of the work by Rosenthal, a key exponent in this area.

Rosenthal [Rsnt91, p469] considers what makes a mental state conscious. "To confer consciousness of a particular mental state, the higher-order thought must be about that very mental state... So, ... the higher-order thought must be a thought that one is, oneself, in that mental state." However, what about when one reads a book? One is conscious of the words, but not conscious that it is they who is conscious of the words. Rosenthal (p470) responds,

> "We normally focus on the sensory state and not on our consciousness of it only because that consciousness consists in our having a higher-order thought, and that thought is usually not itself a conscious thought. ... For a mental state to be conscious, the corresponding higher-order thought must be a thought about oneself, that is, a thought about the mental being that is in that conscious state."

This statement highlights a problem with the higher-order view: if it is the higher-order state which makes the sensory state conscious and yet we are focused on, i.e. conscious of, the sensory state, why have a higher-order state at all? That is, what is it about the higher-order state that makes the sensory one conscious?

Rosenthal [Dave93, p211] acknowledges the above point as advanced by Brentano in terms of "performance conditions," i.e. that the mechanisms behind the first and second order thoughts are the same, so why should one and not the other be conscious? Where Rosenthal goes wrong is drawing a comparison between levels of reference in language, which he relates (unknowingly) to the (logical) manipulation rules used in our serial stream of conscious thought, and the mechanism of thought. This analogy does not extend to qualia. So paraphrasing, in Rosenthal's view we can have thoughts about thoughts in an analogous way that we have words about words, as Wittgenstein would say. However, Wittgenstein also said that these words about words are still just words. What about qualia? We do not have qualia with qualia as content. Holding such a view indicates the wrong perspective on what qualia are.

How does the higher-order state pick out a 'red' conscious state rather than a 'blue' one? Consequently, this could lead to a linking problem. Rosenthal (p472) continues, "If a sensory state's being conscious is its being accompanied by a suitable higher-order thought, that thought will be about the very quality we are conscious of. It will be a thought that one is in a state that has that quality." So, for example, first off I have a state that is a sensory state 'R1' indicating a patch, where 'R1' is a non-conscious representation for a red patch say. This state is accompanied by a higher order state 'I1' with content: *In state about R1 quality*. The Italics are to highlight that the states contain a great deal of unpacked notions, e.g. where is the quality coming from?

Rosenthal [Dave93, p209] suggests that for the higher-order account to work, one must be conscious of being in a particular mental state token type: "Perhaps a dispositional account will require not that the disposition refers to a mental-state token, but that it is a disposition to have a higher-order thought that refers to it." As it stands this will still not do since no qualia enter the picture here merely by having the 'potential' disposition to have a higher-order thought. Paraphrasing, this becomes a disposition to have a higher order thought that refers to 'red' (say). This does not lead to 'red' being experienced as red since the 'red' is only a token at that point. The disposition needs to be continually actioned. The problem with the higher-order view is that the content of the higher-order states is taken as being a conceptual kind of presentation.

A counter argument to Rosenthal's theory is this: a single higher-order thought could accompany a group of lower order thoughts; there is nothing in the theory to prevent this from happening. How can it make them all simultaneously conscious, particularly if these lower thoughts are about different things, i.e. qualia?

Finally, Rosenthal (p473) goes on to say, "For an organism to be conscious means only that it is awake, and mentally responsive to sensory stimuli." This seems to be confusing 'conscious' with 'conscious of' as in 'aware of'. For example, if one stares through a window at a landscape, and empties one's mind of thoughts, so that you just have the 'conscious' sensory experiences. That is primary consciousness. A non-conscious algorithm could achieve being awake and mentally responsive! Really, the answer is that it is not orders, but rather the perspective that is important.

7.7 Summary

All the theories reviewed found explaining qualia problematical, or else they denied their existence. The requirements for a semiotic system presented in chapter four agrees with some aspects of these theories but differs on others. For example, it agrees with Rey's suggestion that sensory states are processes involving many interactions, but doesn't support the interpretation of consciousness as a traditional form of representation. The requirements suggest this traditional perspective will make interpreting qualia difficult. An alternative perspective will be developed in the following chapters.

8 A Perspective on Cognition & Consciousness

The analysis presented in this chapter suggests mind should be treated as a set of abilities, with conscious cognition and primary qualitative consciousness being necessary. Conscious cognition is stipulated as necessary in order to distinguish mind from unconscious symbolic manipulation (although whether this is possible has yet to be proved). Only primary consciousness is required rather than secondary, since there could be minds without self-awareness. Qualitative consciousness is included because pure thought has a qualitative aspect. Adding cognition to the requirement excludes some animals that only have primary consciousness and thus who would mostly have conditioned responses/reflexes. Strictly speaking, cognition involves the manipulation of representations, although primary conscious may do as well but unconsciously. This appears to be the minimal gross requirements for mind.

With the current state of analysis the need for intentionality is undecided at this stage since the relationship between form and content is uncertain. The justification of form and content as a basis for the main supporting medium is not yet proven.

Fetzer's semiotic system criterion for mind is a good individuator. In what follows, the semiotic conception of consciousness and cognition, as defined by Fetzer, will be used as a guiding definitions for the subsequent chapters – see review in the section on mind as a semiotic system.

9 Summary

The goal of this chapter was to informally define and characterise the main features of the mind that workers in the field of mild AI will eventually have to accommodate in their artefacts if their endeavours are to be considered a success.

This started with the ontological reason for intelligence and minds as an environmental advantage to an animal's survival. The central features of a mind were seen to be cognition and consciousness. These features were linked through their dependence on a form of intentionality. Qualia were reviewed as a working term for referring to the phenomenal nature of consciousness, and were seen to be the intentional content of conscious states.

The quale problem was introduced as concerning the nature of the system mechanism that gives rise to specific experiences of mental qualities. It was conjectured that a common mechanism underlies the various forms of intentionality and that understanding this mechanism hinges on a solution to the quale problem. After reviewing a number of theories of consciousness and their perspective on explaining qualia, no conclusive solution was found. However, the semiotic conception of cognition and consciousness that was reviewed showed promise. Consequently, these conceptions were adopted as guiding definitions for the subsequent chapters.

It was suggested that a solution to the quale problem would ultimately determine the fate of mild AI. However, saying this may be premature. The job of the next chapter is to start exploring the nature of the processes that give rise to mind and qualia.

Chapter 3

Operational Requirements for Semiotic Processes

1 Overview

For mild AI to succeed it has to show that a semiotic system can be replicated in a machine. This requires showing how a system can operate in terms of semiotic processes. Here, Peirce's semiotic acts as a conceptual framework from a first person perspective for reasoning about signs, the interpreter, and mind. This is a pertinent point from which to explore the operational requirements for mind from a third person perspective.

To understand the nature of these processes it will turn out to be essential to keep in mind at all times the context or perspective of the explanation, i.e. whether it's from a first or third person perspective (e.g. phenomenal or operational), and the ontological level of the thing or property being considered. These terms and their significance are discussed shortly. In addition, unless the context implies otherwise, the term "consciousness" is to be understood in its semiotic sense, such that the term "generation of consciousness" implies a structure instantiating a sufficient network of signs that supports their use by the mind so generated. Likewise, following Fetzer [Fetz98b], "cognition occurs as a consequence of causal interactions between systems that are conscious (with respect to signs of specific kinds) and the presence of signs of those specific kinds in suitable causal proximity" – see chapter two. Therefore, the thesis is not dealing directly with those aspects of consciousness, such as sentience, awareness, and self-awareness, but rather, the more basic structure of the dispositional processes collectively responsible for producing consciousness in the semiotic sense.

1.1 Logical Development

This chapter suggests semiotic processes can arise in a system whose operation at a particular ontological level is based on functional-flows and temporal-representations. These terms are discussed later and defined in the appendix. Three problems levelled at theories of consciousness are reviewed and used as a guide for the analysis.

A hypothetical system containing a hierarchy of three ontological levels (an implementation level, a dispositional level, and a phenomenal level) is proposed as a working framework on which to refine the operational basis of semiotic processes. A particular kind of dispositional process is investigated as a possible mechanism linking the implementation and phenomenal levels. The term "dissociated disposition" was introduced to refer to an abstraction of this mechanism. These characterise the variable dispositional potential of the processes within certain kinds of causal systems. This concept of variable dispositional potential plays a central role in this and the following chapters.

This chapter focuses on the operational nature of the processes underlying semiotic processes. Although qualia are mentioned in passing, the next chapter considers how these underlying processes are structured to produce qualia. The next chapter after that discusses the requirements for these processes to be considered semiotic and considers how a machine might be configured to replicate them.

2 The Importance of Dispositions to Semiotic Processes

The introduction chapter mentioned how Peirce classified the semiotic characteristics of signs into three trichotomies, the third of which dealt with the effect the interpretant produced in the subject. Peirce termed these effects as being of a logical, emotional or energetic nature. The first term referred to habit-changes or thoughts, the second to qualia, and the third to behaviour. Fetzer extended Peirce's analysis of the role played by habits in imparting meaning to signs by showing how this can be cast in terms of dispositions – see Fetzer [Fetz90, p78] and [Fetz91]. This stemmed from Fetzer's dispositional ontology for the physical world in which the concept of a disposition was formulated with respect to a descriptive language as:

> "A predicate is *dispositional* if and only if the property it designates (a) is a tendency (of universal or statistical strength) to bring about specific outcome responses when subject to appropriate singular tests, where that property (b) is an actual physical state of some individual object or of an arrangement of objects (should it happen to be instantiated by anything at all)." – see Fetzer [Fetz77, p401].

For present purposes, the following four dispositional conceptions introduced by Fetzer [Fetz81, p40] are explanatorily useful for defining how dispositions relate to kinds, things, objects and events:

1) "(particular) *kinds of things* are specific arrangements of (permanent and transient) dispositions, independently of whether or not these distinctive sets of properties happen to be instantiated during the course of the world's history;
2) *things of* (particular) *kinds*, therefore, are instantiations of some specific arrangement of (permanent or transient) dispositions that happen to occur during the course of the world's history, independently of whether the arrangements they instantiate are object *or* property kinds;
3) *individual objects* are continuous sequences of instantiations of particular arrangements of dispositions during the course of the world's history, where any object ceases to exist as an object of a particular kind whenever it no longer instantiates the corresponding (reference class) description;
4) *singular events* are continuous sequences of instantiations of particular arrangements of dispositions during the course of the world's history, where any event ceases to exist as an event of a particular kind whenever it no longer instantiates the corresponding (reference class) description."

Fetzer distinguished between *atomic events* and *molecular events*, where a molecular event, such as the sinking of the Titanic, was a sequence of atomic events.

Of note here, is the importance of the reference class description for identifying particular kinds of dispositions. Properties are further distinguished as being either permanent or transient with respect to a reference class description. Thus, a property is a permanent property of every member of some reference class if and only if losing that property would exclude it from the class and the possession of the property is not logically entailed by the reference class description; otherwise it is a transient property – see Fetzer [Fetz81, p38]. For example, members of the reference class description "being water" would have, among their permanent dispositional properties, a boiling point of 100 ^{o}C at a pressure of one atmosphere, while being used to put out fires would be a transient property. Finally, Fetzer deals with dispositional properties where the "thing" need not be an object, and indeed, in what follows the thing will turn out to be a process. A related point is that a particular permanent property may have a variable state, for example, water always has a spatial extension, i.e. “a physical volume”, regardless of the particular sample referred to – the size of its volume would be a transient property. A property such as the variable spatial extension will be called a *graded* disposition and is discussed later.

Consequently, the task for this chapter is to explore how semiotic processes might be implemented in terms of certain kinds of dispositions. The approach taken here is to start by considering the ontological nature of physical systems and the part played by natural laws in shaping their dynamics. This matter is taken up in the next section after which the discussion returns to the dispositional basis of semiotic processes.

3 Ontological System Levels

To explore how semiotic processes might be replicated in a machine the thesis adopts the framework of ontological system levels (described shortly) and its significance to the relative perspective of the mind created in the machine versus the observer, i.e. the first and third person perspectives, respectively. The objective is to produce an ontic explanation for the form of the relationships in the system, rather than an epistemic explanation for testing whether the machine possesses a mind. Notice that to provide implementable specifications the operation of the system has to be explained from a third person perspective. This perspective is taken throughout the following chapters.

3.1 Causal Relevance Model of Explanation

The exploration is based on a causal relevance model of explanation whereby causal conditionals are central to describing the nature of the system. A causal conditional is a statement to the effect that the occurrence of an event brings about, or causes, the occurrence of a second event. Fetzer [Fetz77, p407] introduces the non-extensional “fork” operator to represent subjunctive conditionals, and annotates this with subscripts “u” and “n” to represent universal and probabilistic causal conditionals, respectively. Thus, the subjunctive conditional: (x)(t)(Kxt ==>Xxt), asserts that “For all x and all t, if x were K at t, then x would be X at t.” In contrast, the universal causal conditional: (x)(t)(Txt =u=> Oxt), asserts that “For all x and all t, subjecting x to T at t would invariably (with strength equal to u) bring about O-ing by x at t.” This describes a disposition of universal strength. The probabilistic causal conditional: (x)(t)(Txt =n=> Oxt), asserts that “For all x and all t, subjecting x to T at t would probably (with strength equal to n) bring about O-ing by x at t.” On a technical note, Fetzer [Fetz81, p110] emphasises that this conception entails that more than one outcome is possible under precisely the same complete sets of relevant conditions with constant probabilities. In what follows, interest lies with a graded form of universal dispositions, for which it is convenient to introduce a graded causal conditional "=g=>", described shortly. Finally, bearing in mind the dispositional definitions for events and things given in the previous section, the above conditionals denote lawlike sentences, when the variables remain unquantified, and they denote nomological conditionals when instantiated – see Fetzer [Fetz81, p49].

The causal relevance model of explanation is adopted for two reasons. Firstly, the evolution of physical systems is determined by natural laws, and in particular causal laws, as opposed to social or logical laws – see Fetzer [Fetz93, p22]. This means that if an explanation proceeds from first principles, the dispositional properties of the system can be used as a basis for explanations, specifically, in terms of their causal relationships, and the form of these relationships. Secondly, causal relevance is taken to refer to a version of the requirement for strict maximal specificity whereby only causally, or nomically, relevant phenomenon appear in explanations – see Fetzer [Fetz93, p77]. A less narrow conception, for example, would allow for explanations that explain why something x has a property A on the basis that A is a permanent property of everything that has property R, and x has property R. Hence, in exploring the operational requirements for semiotic processes, while there are non-causal, but nevertheless lawful relationships between brains and minds, only one kind of operational relation will be used to explain things with, i.e. causal (this follows by definition of an operational relation as the nature of the connection between a cause and its effect). This implies that all semiotic processes would have to be based on some form of causal process.

3.2 Causal System Laws, Levels and Properties

The goal of this and the next two subsections is to establish a descriptive foundation for discussing causal systems with particular emphasis on the ontological levels they contain. In this regard, consider what Newell [Nwl82] had to say about his proposed knowledge level in the context of computer system levels,

> "A level consists of a medium that is to be processed, components that provide primitive processing, laws of composition that permit components to be

> assembled into systems, and laws of behaviour that determine how system behaviour depends on the component behaviour and the system structure. … Each aspect of a level – medium, components, laws of composition and behaviour, – can be defined in terms of systems at the next level below.”

This implies there is no lowest level, but in what follows a base level is assumed – this point is discussed later. Following Newell’s example, to explore how semiotic processes might be replicated in a machine, a framework is developed whereby systems (i.e. machines) are analysed in terms of three interrelated aspects: ontological levels, laws and dispositional properties. *Ontological levels are a lawful categorisation of dispositional properties in that an ontological level is a domain of elements over which a specified set of laws intrinsically applies.* The set of laws and elements thereby demarcates a level. System laws reflect constraints from lower ontological levels. Dispositional properties reflect physical structure and instantiations. These points are discussed in more detail shortly. The next few paragraphs present an overview of the formal characterisation of systems presented in the appendix.

In characterising causal systems, the first task is to consider the universe from which the systems will be constructed in terms of the laws that are in effect, the elements it contains, and the structure on the elements induced by the laws. Consequently, in a system S_i the set of laws that are in effect are collectively referred to as the axioms of the system, and will be denoted by the set $A_i = \{a_{ij} \mid j \in 1..n_a\}$. The configuration of the elements for the system is specifiable as a set of sentences C_i, in a formal language L, under an interpretation I_i. The set of elements $E_i = \{e_{ij} \mid j \in 1..n_e\}$ upon which it is constructed is called its domain. The formal language L and the interpretation I_i are introduced as a convenient method for discussing the structure of the system.

A system is then an L-structure $S_i = \langle E_i, F_i, R_i \rangle_{C_i}$, where the subscript C_i indicates a set of sentences from L specifying the state of E_i, where E_i, F_i and R_i are sets of symbolic names for the elements, operations and relations of S_i, respectively. The set C_i describes the configuration of the elements (in terms of the F_i and R_i) into a particular instance of a system, and the requirement $C_i \models A_i$ constrains all specifiable configurations of the elements to obey the axioms. In effect, a set of axioms and a configuration of elements determine a system.

Given a base system S_i^0, a higher order system S_i^n is a structure: $S_i^n = \Xi_i^n(S_i^{n-1})_{C_i^n}$, where Ξ_i^n is a construction process on S_i^{n-1}, such that the two systems have different axiom sets ($A_i^n \neq A_i^{n-1}$), and the subscript C_i^n denotes the configuration of the elements necessary for S_i^n, $n \geq 0$. The system S_i^{n-1} from which S_i^n is constructed is called the basis for S_i^n. To distinguish between a system and the systems from which it is constructed the set of systems from which a system S_i is constructed is denoted by the set $L_i^n = \{ l_i^n \mid n \in 0..(n_l - 1)\}$, where $l_i^n \equiv S_i^n$ is called an ontological level of the system of order n. Here, an important point is that *a difference in the axiom sets needed to define a level distinguishes each ontological level in the system.* Hence, a set of elements and a set of axioms define an ontological level.

In contrast to ontological levels, a hierarchy H is a set of entities upon which is defined an arbitrary partial ordering. For example, the entities could be levels, where the ordering is over the order of the level, or a group of elements, where the ordering is over their functional organisation. It then follows that a particular system will have been constructed from a hierarchy of ontological levels relative to a base system. Notice that the elements within an ontological level may be organised according to some hierarchy, and yet all the elements will still be intrinsically under the influence of the axioms that characterise the ontological level as a whole.

3.3 Explanations From a First & Third Person Perspective

Within this characterisation of systems, it is possible to distinguish explanations from first and third person perspectives. Firstly, a notion of self-referent sentences is needed. Thus, a self-level-referent sentence r_{ij}^x of a level l_i^x in a system S_i is a sentence that refers to the elements E_i^x of S_i^x – see Smullyan [Smul94]. Consequently, given a system S_i with a set of levels L_i^n, a *first person perspective* is an interpretation via I_i^x of any self-level-referent sentence r_{ij}^x of l_i^x in S_i, where x designates a single level c. A *third person perspective* for a system S_i is an interpretation via I_i^x of any sentence s_{ij}^x, for any level x, where the sentence s_{ij}^x refers to any other system S_k, $i \neq k$. Notice that this definition does not require the system to be a person or to have consciousness. From this, the notion of an *observer* arises as a third person perspective.

The appendix gives a formal analysis and definition of ontological system levels, system laws, properties, and the first and third person perspectives.

3.4 A Note on Emergence and the Causal Ontology of Levels

In what follows two conceptions of emergence are drawn upon. The primary conception follows that of Fetzer [Fetz86, p124] whereby an emergent property is dependent upon the arrangement of conspecifics. The term "conspecifics" refers to other members of the same species, for example, people in the case of emergent properties of a social group, and atoms in the case of chemically emergent properties. The secondary conception relates to the ontology of levels and is a variation on the contemporary view as detailed by Emmeche et al. [Emm97]. Briefly, properties at a certain level of organisation that cannot be predicted from the properties found at lower levels are said to be emergent. Here prediction refers to the inability to determine the future behaviour of properties in lower level terms. Nis Baas [Baas94] distinguishes between two types of emergence: deducible emergence in which there is a deductional or computational process by which a higher level property can be determined, and observational emergence for which no such means of determination is possible.

Herein, the conception of emergent properties due to the arrangement of conspecifics expressed in terms of ontological levels was found sufficient for the purposes of exploring how mind might be replicated in a machine. In these terms, an ontological level can be seen as a domain of emergent properties that are all instantiations of arrangements of properties of lesser complexity. To pursue this requires formally refining Emmeche et al.'s ontology of levels and extending Baas's and Fetzer's emergence framework in a systems context – see Baas [Baas96]. In this regard, the appendix presents a formal characterisation of the ontological levels, laws and properties of a causal system. The criterion for distinguishing levels is that any consistent axiomatisation leads to a different intended interpretation for the semantic model that the syntax is meant to represent at a higher level of abstraction.

Besides the topological arrangement of conspecifics, other factors may contribute toward the invocation of an emergent property. Within the current context, it helps to distinguish between the primary conception of *topological* emergence, in which a new property arises from a particular arrangement of causally interacting systems (cf. Fetzer's conception), and *temporal* emergence, in which the behavioural patterns of the interacting systems as a whole is also significant. In summary, focus will be on the arrangement of instances of interacting dissociated dispositions, which map to the activity of the supporting medium (e.g. neurons), where the extent to which an element of the support is influenced by the activity of its neighbours, is a function of the activity across the interconnect of the medium. That is, physically a fixed network of elements could act as the support where causal variation in the interaction between elements depends on their internal state.

The next suggestion is that some emergent properties can be created by design by configuring certain networks of instances of dissociated dispositions (e.g. interacting patterns of neural activity). Whether these are predictive or not according to Emmeche et al. is another matter and not of primary concern here – their notation is used as a basis for formally specifying emergent properties in the appendix. Thinking about it in terms of emergent properties, in this particular case, a crucial point made in chapter four is that to get emergent mental properties requires more than just the topological arrangement of interacting causal systems, but also the right temporal causal structure on top of this. A rough analogy would be a football match, which requires the right arrangement of players, performing the right skills. In the case of mind, it is a bit more involved, in that to get emergent mental properties the causal accessibility requirement (see chapter four) requires that there is also the right kind of feedback between instances of dissociated dispositions in order for the system to be able to use its mental signs and therefore be conscious.

Returning briefly to causal conditionals, notice that while the "n-fork" operator means "A has a tendency to cause B", it does not express why this should be so, even under the requirement of maximal specificity, or a covering law – see Fetzer [Fetz93, p63]. One means of explaining the origin of the cause is in terms of processes at the ontologically lower level. For example, laws of co-existence, such as the Ideal Gas law PV = nRT, express regularities at one level that arise through the interaction of processes at a lower level – see Fetzer [Fetz81, p144]. Thus, a physical law that is nomologically obeyed by the processes at one level must have arisen from the structures at that level produced by the processes at a lower level, although, this appears to lead to a regress. Moreover, this problem affects machines and biological minds alike, since they both share the known physical universe as a common base level. In what follows this base level is taken as a given. Thus, with respect to the requirement of maximal specificity and ontological levels, the latter simply reflects the structure induced by the laws and says something about the structure of the domain of properties. In effect, an explanation satisfying the requirement of maximal specificity might be less terse when meta terms are used, such as expressing the relations between emergent properties, i.e. by considering the levels involved. The appendix presents a formal characterisation of the causal structure of causal systems.

4 Dispositional Processes

In the above review of dispositions, it was pointed out how Fetzer notes that dispositional properties need not refer to a physical object, but may refer to a process. In what follows interest lies with graded dispositional processes whose tendency is to influence the graded tendency of other dispositional processes. This leads to an operational ontology for dispositions (cf. [Fetz77]) and the following subsections.

4.1 Functional-Flows & Temporal-Representations

As a foundation for the following discussion on "graded" dispositions, some terminology is helpful from dynamic systems theory – see Norton [Nor95]. The evolution of a dynamic system can often be described by a temporal sequence of state values, such as a trajectory, or path, through its state space. This is commonly called a *flow*. Thus, in contrast to a *functional-map*, which associates pairs of input and output patterns, a *functional-flow* denotes the temporal profile of a flow, i.e. a particular temporal sequence of input or output patterns. The sequence itself may be changing, giving rise to a *dynamic-functional flow*. This leads to distinguishing between a *stationary-representation*, which can be defined at an unconnected point in time, and a *temporal-representation*, which is defined as a flow over a period of time – see the appendix for a more detailed discussion of these terms.

A causal-flow is a particular type of functional-flow that describes the causal relations in part of a process. Within a system, a complex of causal interactions may be possible giving rise to a particular causal structure between its elements. The definition of causal structure (D-CS) given in the appendix casts this in terms of the relations between property instantiations – see also in the appendix, the section on the dynamic structure of causal networks, and the section on defining function & representation.

4.2 Graded Dispositions

Graded dispositions are now examined in order to describe the continuous operation of a causal system, and modelling graded dispositions, such as the disposition to laughter. For example, when someone hears a joke, their laughter grows and subsides, rather than being an all or none, or random event. Treating a disposition as a law, in examining graded dispositions interest lies with the laws of transition between dispositions of a similar type.

An expedient way to describe graded dispositions is in terms of temporal sequences of universal dispositions. The idea being that each member of the sequence is an incremental variation of its neighbours (cf. [Fetz81, p51] – incremental changes in strength of tendency). These variations correspond to the possible grades of the dispositional property a thing may possess at any one time. A graded disposition is then defined as the set of all sequences, where set membership is determined by a reference class description – see also the discussion on defining dynamic functional flows in the appendix. Notice that the variations between sequence members need not be discrete, they could be continuous in which case the sequence becomes a continuum, i.e. a continuous flow. Hence, it is possible to have discrete or continuous graded dispositions. In addition, the mind appears to be a system that is continuously evolving (cf. [Fetz77, p415]), hence the need for incremental and continuous changes.

This conception of graded dispositions and the ensuing system dynamics conforms to Fetzer's laws of cognition LC-6a and LC-6b – see [Fetz96, p109]. For example, while the tendency of graded dispositions is deterministic rather than probabilistic, probabilistic behaviour may still appear to arise from graded dispositions. Briefly, the system evolves according to the deterministic interaction ofgraded dispositions – a particular region may involve the superposition of many graded dispositions. Meanwhile, the observed behaviour of an individual may still appear probabilistic when tested over repeated trials. This is explained as the practical difficulty an external observer would face trying to specify all the "nomically relevant properties" ([Fetz81, p113]) in order to duplicate the trial preconditions. In particular, the test individual would now have knowledge of the prior trials, which would influence their subsequent actions. When a decision's outcome rests on the interaction of many graded dispositions, a small change in the preconditions may be amplified and result in an apparent probabilistic distribution of outcomes (cf. chaotic dynamics wherein small initial changes produce diverse outcomes).

4.3 Distributive & Reversible Dispositions

Distributive dispositions are introduced by way of an example based on the artificial neural nets popularised by Hopfield in the 1980s – see Hopfield [Hop82]. These nets are capable of memorising patterns that can be later re-invoked when prompted by a similar pattern. Kosko [Ksk87b] considered pairs of Hopfield nets coupled together such that the pattern on one net would invoke a certain pattern on the other – see also Dreyfus et al. [Dry88] and Fujiwara et al. [Fji87]. Thus, relations between patterns can be programmed into these nets. So, if patterns "A" and "B" are programmed into the respective nets, such that pattern A invokes pattern B, we can say A has a tendency to cause B. These memorised patterns are often called prototypes. Notice also that each net has a tendency to evolve toward one of it's prototype memories when prompted by a similar pattern. The set of similar patterns that lead to a particular prototype memory are called the prototype's attractor set.

On a technical note, notice that these neural nets can be implemented with synchronous or asynchronous dynamics and, consequently, will produce different behaviour (for example, the synchronous version is more prone to cyclic states) – see Hopfield [Hop82]. This difference does not concern us here – we can choose either dynamics as appropriate. Rather, the difference between continuous and discrete dynamics is more important – see the appendix for a discussion of system processes.

Now consider what is happening operationally. One way to do this is to look upon the dynamical evolution of the state of these nets in purely causal terms. So, consider the change in the state of a net $s(t)$ at times t_1 and t_2:

$$\text{Net state change } \Delta s = s(t_2) - s(t_1) \approx \delta s \text{ for small } \delta t.$$

Ideally, the change produced by the evolution of the state can be thought of as having an inverse, for example, there is some $\delta s' = -\delta s$. That is, just as there may be some δs that has a tendency to drive the nets toward one prototype state, there may be some other $\delta s'$ that has a tendency to drive the nets away from a prototype state. The reason for noting the possibility of an inverse is to emphasise that the dispositional tendency exhibited by the net may be cyclic i.e. in some sense *reversible*. For example, a piece of clay can be moulded into a ball, then a slab, then a ball again etc., while for blocks of stone this would be an irreversible process. Consequently, the δs corresponds to a directional influence, or tendency toward a prototype state:

Equivalently, new state: $s(t+\delta t) = s(t) + \delta s$.

From this, the current state of a net can be thought of as consisting of 1) a position in state space, plus 2) a directional influence (the vector δs) toward some other prototype state.

To elaborate further, Fetzer in [Fetz77, p403], draws an important distinction between "*predicate constants*..., such as 'H', and the *sentential functions* that may be constructed from them, such as $\ulcorner Hx \urcorner$." The latter "exhibits the form of an event attribution." Now, for the present purposes, the 'H' can be thought of as a compound predicate. Fetzer gives a specific example of 'H' as designating "a half-life of 3.05 minutes". It is said to be compound because the tendency it designates has a certain form that dictates the things that may permissibly instantiate it. Thus, *a disposition when defined irrespective of the thing, would be the tendency to produce a change of a particular form under suitable conditions.*

Recalling the definition of a property given in the previous section, and defined formally in the appendix, see definition D-SLP, an instance of a property was effectively a function of a particular configuration of elements: $p^n_{ijx} = \omega^n_{ij} \varsigma^n_{ij}(C^n_{ijx}, E^n_i)$. Consequently, in the above neural net example, by abstracting from the operational medium, a certain category of *distributive* dispositions can be defined as a directional and potentially reversible tendency to redistribute the domain elements of some causally connected thing. A particular kind of distributive disposition would then be defined by a reference class description that specified over what elements the disposition exerted an influence, how this set may change, and the form of its influence. Notice that a further restriction may be applied such that the redistribution is always upon a fixed subset of the domain elements.

4.4 Isogenetic & Dissociated Dispositions

In the previous subsection, a category of distributive dispositions was introduced as a tendency to redistribute the domain elements of some causally connected thing. Within this category, a class of *isogenetic* distributive dispositions can be singled out according to two further refinements. Firstly, the form of the distributive dispositional tendency is itself a function of the distribution of the elements upon which the disposition is instigated. In other words, the same type of process defines the tendency of the disposition as that upon which it acts. Secondly, by abstracting from the particular kind of underlying elements, i.e. the medium, only the causal power and form of the distributive influence becomes of importance in understanding the operation of the system. This allows the analysis to focus on those kind of arrangements of properties that manifest semiotic abilities as among their emergent properties. Here, "emergent" is meant in the sense that systems as instantiations of arrangements of properties of lesser complexity (or of different properties, etc.) do not manifest them.

Consequently, a homogeneous interactive network of graded isogenetic distributive instances of dispositions can be treated as an abstract level. Referring to these as *dissociated* disposition instances (DDIs), they will be associated with an ontological system level that supports the evolution and interaction of instances of these dispositions. It can be treated as an ontological system level because it will be characterised by laws specific to that level that determine the interaction and evolution of the instantiated dispositions. Of concern in what follows is the lawful nature of these dispositions, how their instances relate to one-another, how they relate to the other levels, and finally, how they might lead to semiotic processes. Thus, in what follows dissociated dispositions and their instances will be used to refer to dispositional processes in which details of the medium have been abstracted away along with any of its irrelevant dispositional properties.

5 Minimal Semiotic System Framework

As a framework for developing the requirements for a semiotic system, a hypothetical semiotic system is proposed to which is attributed qualitative consciousness, the most primitive form of consciousness under the semiotic conception – see the discussion in chapter two. The task is to explain and refine the operation and structure of this hypothetical system. The remaining sections in this chapter explore the structure of this system, with respect to the operational nature of instances of dissociated dispositional processes, as a precursor to producing semiotic processes. The next chapter considers the gross system structure required to produce qualitative consciousness. Finally, the chapter after that studies what is required to consider these as semiotic processes.

5.1 Classifying the Levels in the Hypothetical System

Returning to the opening hypothesis, the hypothetical system is divided into three ontological levels for the purpose of analysis: an implementation level, a dispositional level, and a phenomenal level. Thus, adopting the notation used in the appendix:

Statement of Hypothetical System. (S-HS)

Let S_H denote the hypothetical system to which qualitative consciousness is attributed. Initially, S_H is hypothesised to consist of three levels: $l_H^1 \equiv l_H^I$ an implementation level, $l_H^2 \equiv l_H^D$ a dispositional level, and $l_H^3 \equiv l_H^P$ a phenomenal level.

At this stage these are just convenient levels of description, with the dispositional level initially introduced as a way of referring to the unknown process that bridges the other two. However, as the chapter develops it will be shown that they can be treated as distinct system levels in the ontological sense developed at the start of this chapter and more formally in the appendix. For example, to carry out an operational analysis of concepts would require distinguishing between qualitative properties and the operational basis of the ontologically lower dispositional level. Concepts when experienced as such are a property of the phenomenal level that is generated from the dispositional level. Hence, a web of instances of dispositions in a given operational context and in conjunction with the causal accessibility requirement (see next chapter) gives rise to the phenomenon of concepts at the phenomenal level.

5.2 Hypothesis for Dispositional Composition of Qualia

The first step in refining the system structure is to suggest that there is actually a structure to be refined. Chapter two characterised the quale problem as concerning the nature of the mechanism that produces qualia in a system. In what follows the viability of this mechanism being operationally based on instances of dissociated dispositions is analysed and to start with is hypothesised to have a compositional structure:

Hypothesis for Qualia - Dispositional Composition. (H-QDC)

Suppose that qualia, positioned at the phenomenal level, are operationally composed from instances of dissociated dispositions at the dispositional level. Then the hypothesis is that a quale is not operationally synonymous with a single dissociated disposition at the dispositional level. That is, qualia are derived from dissociated dispositions.

This suggests a quale, although itself dispositional, is an emergent property produced by a mechanism (i.e. a web of dissociated dispositions) with some discernible structure. An analogy would be an aeroplane producing the property of flight. That is, the relationship between a quale and the mechanism is not necessarily a simple identity.

It must be emphasised that the kind of dispositions pursued here refers to a type of mechanism or process. The following discussion will show that they are certain patterns of causal activity. They are not entities as such or conscious in any sense.

5.3 Modelling versus Signifying Reality

Herein, when the system is said to *model* some aspect of reality this is meant to imply the internal formulation or existence of a process that may enable predictions or decisions to be made concerning that aspect of reality – see Johnson-Laird [John83]. A model is not necessarily a faithful or complete portrayal of the aspect, and an aspect can be modelled in many different ways, e.g. via replication or emulation, as mentioned in the introduction chapter.

In general, the form in which something is modelled in the system can be either in terms of explicit representational relational structures, which are interpreted in some way, or in terms of implicit relational structures occurring between the interaction of the processes in the system. As mentioned in the opening section, this latter, implicit form of modelling offers one route to devising a sustainable theory of grounding meaning. As will be seen in the next chapter, it also forms a basis for one means of explaining how qualitative consciousness might be generated in the system. In computer science, a distinction is made between declarative and procedural knowledge and programming languages. However, notice that in the above case, modelling by way of implicit relational structures is not a computational process in that it is not about computing a result but rather the occurrent form of the causal interactions – see Fetzer [Fetz98a].

Under the semiotic conception, reality is experienced, rather than modelled, as an implicit relational structure of signs produced by an underlying network of instances of dissociated dispositions. The phrase *signified perception of reality* reflects this semiotic conception, and how the system extracts salient features from its reality which come to be represented by signs and utilised in some capacity, originally to aid survival – see Umberto Eco [Eco76]. It also emphasises that this is an occurrent process.

The reason why the system generates a "signified perception of reality" is discussed later in the chapter. One theme to arise in this and the following chapters is the suggestion that the system's signified view of reality can be operationally constructed by means of an evolving network of interacting dissociated disposition instances.

5.4 Signs, Objects & Entities

In Peirce's theory of signs, a sign corresponds to a property of a semiotic process involving an irreducible triadic relation between a sign, object and interpretant – when viewed from the phenomenal level. When viewed from an operational perspective, it is important to distinguish between properties of the sign and the perceived properties of the thing it represents. The experienced sign in itself has a wholeness quality and identity that arise from the phenomenal projection of the underlying semiotic process. That is, a sign is experienced as an individual entity of a particular type. This is a reflection on the aspects of the dynamical object that the sign represents. The dynamical object was what Peirce called the object being represented. This need not be an actual physical thing; it could be any kind of imaginable entity, including other thoughts. Metaphysics categorises entities as being either physical or abstract and either individual things, properties, relations, events, states of affairs or sets – see Audi [Audi95]. The semiotic processes underlying the sign and their structured interaction with other semiotic processes produce a signified perception of the dynamical object, which is experienced at the phenomenal level. In this sense, an experienced entity corresponds to a quale, or collection of qualia.

The point to be made here is that while the sign as experienced has a definite wholeness and identity, the thing it represents may be far less substantial, such as a belief, the notion of an action, an unrecognised thing, or a thing of puzzlement. In effect, to be able to signify reality consistently, the system has to be able to deal with entities that may not be tangible objects, recognisable or quantifiable. For these cases, signs are instigated whose type reflects the identifiability of the dynamical object. It should be stressed that this is not positing the existence of mental objects (for example, see discussions on the metaphysics of objects and Meinong in E.N.Zalta [Zal97, Zal93]).

In general, according to Peirce and Fetzer, a sign is simply anything that could stand for something (else) in some respect for somebody, as a (potential) sign; when it does stand for something (else) for somebody, then it become an (actual) sign. However, here the focus is on internal signs as used by the system, i.e. interpretants that become signs, and determining the nomic form of the mechanism by which signs are 'used'. In this regard, the above points suggest the following statement:

Statement on Operational Individuality of Mental Signs. (S-OIMS)

A mental sign resulting from a semiotic process has an identity and wholeness with respect to the system regardless of the dynamical entity being represented.

The idea is that there is a mechanism to the effect that the sensation (or property) of identity and wholeness is associated with a set of properties. Hence, in what follows, a mental sign demarcates abstract mental objects, which are sets of dispositions over which Bound and Identity dispositions are instantiated – binding is discussed shortly and later in the section on signification and the section on logical primitives and the identity disposition.

With this idea of sign individuality in mind, a sign in consciousness ranges from representing a simple thing, such as a colour patch, to a compound object such as a tree, or, for example, the notion of the sky or an event. This implies a sign is a bound collection of qualia. The above review of Fetzer's dispositional ontology suggested how kinds, things, objects and events could be defined, and therefore signified, in terms of dispositions. This means an entity in the signified portrayal of reality is produced by an interacting set of disposition instances. Therefore, in the following an experienced entity is treated as a property of an occurrent interaction of dissociated disposition instances, rather than the interpretation of an extensional representation.

5.5 Signs & Binding Problems

Following from the definition of mind as a semiotic system, the prior statement on the operational individuality of signs raises questions concerning the integration of mental content since it suggests the apparent seamlessness of mind is composed from discrete entities. This has led to a number of interrelated problems: the feature-ground binding problem, the identity problem, the grain problem and the superposition problem – for the philosophical background to these problems see Thomas Metzinger [Metz95].

The feature-ground binding problem concerns how something is perceived as a whole given its properties e.g. an apple is red and round. The identity problem concerns how something retains its identity from one moment to the next. The grain problem arises when considering sensory experiences that appear to present a subjective continuum, such as the homogeneity of a colour patch, and how this could possibly be produced by discrete neural events or finite collections of neurons. The superposition problem concerns how the system is able to support more than one bound entity without the various properties interfering – for a discussion on these problems see Ian Gold [Gold98].

At present, the feature-ground binding problem is considered the harder problem and has received the most attention. One approach championed by Francis Crick and Christof Koch [Crc90, Crc94], is based on the concept of dynamic binding via the synchronisation of neuronal discharge as a requirement for awareness. This stems from their work focusing on empirical approaches to trying to find a neuronal correlate of consciousness (NCC). In summary, they suggest "that necessary conditions for the NCC must be some neuronal activity encoded within an explicit representation with direct access to the planing stages of the brain lasting for a sufficiently long time" – see Crick and Koch [Crc95, Crc98].

Much remains to be explained from a nomic point of view with regard to the empirical results and conclusions that are arrived at by Crick and Koch. For example, what is it about neuronal synchronisation that should lead to consciousness? Why should one particular form of neuronal synchronisation produce consciousness and not another? How does this solve the binding problem and the other mental content integration problems? Is synchronisation a nomic requirement? Is the founding objective sound? That is, should we expect to be able to find a neuronal correlate of consciousness, or is this misconceived? The following addresses some of these problems and questions in passing during the investigation into the operational basis of semiotic processes.

5.6 Distinguishing Between Perceptual Fields, Spaces, Sensations, Signs & Qualia

Chapter two noted how qualia arose from a perceptual process through the interpretation of sensory data. Broadly speaking, sensory data refers to the raw information collected by the senses and relayed to the brain. This is interpreted by a perceptual process, giving rise to, in the case of conscious creatures, signs and hence conscious sensory experiences that have qualia as contents. Not all conscious experiences are of a sensory origin, some, such as thoughts, are internally generated. For this reason, a semiotic process is a more general notion than perceptual processes since it encapsulates these internally generated experiences as well. This section discusses the relationship between these terms as used in the thesis.

From a first person perspective, qualitative experiences can be categorised into finite volume elements in a closed perceptual space of fixed dimension. For example, in the case of the olfactory sensory modality, dimensions have been proposed that reflect the degree of bitterness, sourness and sweetness – see Clark [Clar93, p79], Schiffman [Schi90] and Gulick [Dave93, p153]. A path through the perceptual space corresponds to an episode of qualitative consciousness. Summarising the terminology of Clark et al., the mechanism underlying qualitative states is said to be a "perceptual field":

Definition of Perceptual Field. (D-PF)

The term "perceptual field" refers to the support mechanism responsible for generating the perceptual space.

Percepts are defined as finite regions in this field. Thus, a percept is a type of bound process. The support mechanism would be implemented in a suitable architecture. Conventionally this has been biologically based. A goal of the thesis is to determine whether alternate architectures are feasible.

The term "perceptual space" refers to the collective qualitative experiences arising from the perceptual field. The perceptual space encompasses the union of sensory modalities (according to Clark et al. [Clar93]), with qualia experiences being associated with perceptual processes, i.e. finite-volumes in their respective modal subspace. These experiences are perceived by the mind as a cohesive network of signs that the system constructs and are manifest through qualia.

Definition of Perceptual Space. (D-PS)

The term "perceptual space" refers to the total qualitative experience of reality from the first person perspective of the mind that is constructed by the system.

The perceptual space, in conjunction with internally generated experiences, forms the top, phenomenal level in the system. Thus, in accord with the review of perceptions and sensations in chapter two, qualia are based on semiotic processes, which includes perceptual processes. The perspective refers to the system level of the observer (i.e. beholder) and the nature of the percepts that determine the relational way in which reality is portrayed. This point is developed later in the chapter – see the definition of first and third person perspectives mentioned previously and in the appendix, and the requirement for a perspective view of consciousness given later in this chapter.

On a technical note, the terms *field* and *space* are used to emphasise a number of points. Firstly, a field can be thought of as forming the basis for a space. It will be seen how the dispositional level can be thought of as a field. A perceptual space is a closed pseudo metric space of fixed dimension of mixed type. The metric is a pseudo metric because the notion of quale distance is undefined between sensory modalities, instead a distance measure can be defined in terms of discrimination functions, which is non-zero only for those qualities that are members of the space, see the requirement of causal accessibility described later. The perceptual space is not a conventional spatial-temporal space since it has no spatial extension.

Although the perceptual space appears to the mind as a continuum this does not imply the underlying perceptual field is continuous too. This point is important since it could be argued that a discrete process could not give rise to a continuous process, cf. the grain-problem. A goal of the thesis is to explore whether mind can be replicated in digital, and hence dynamically discrete, machines. Part of the response to this is that determining whether something is discrete depends on the event resolution of the observer. This point is returned to when the binding problem is discussed in more detail later.

6 Three Problems for Theories of Consciousness

This section reviews three problems that any theory of consciousness would have to address. These problems help refine the requirements for a semiotic system and motivate the subsequent analysis of semiotic processes and the generation of qualitative consciousness.

6.1 The Interpreter Regress Problem

The homunculi infinite regress problem affects a theory of mind that posits an internal "little man" to explain, for example, perception. For in turn another internal homunculus would have to be posited for the first one, and so on – see Dennett [Denn78]. The interpreter regress problem is a variation on the homunculi problem applied to theories of content that involve an interpreter. If an interpreter is posited as interpreting a representation, then another inner homunculus is required to do the interpreting, and so on – see Eckardt [Ecka93, p285]. This is summarised in the following statement (using the notation from the appendix):

Statement of Interpreter Regress Problem. (S-IRP)

To interpret I^{P}_{HMi} *a representation* R^{P}_{Hij} *of an aspect of a modality* M^{P}_{Hi} *requires a further representation* R^{P}_{Hij+1} *in* I^{P}_{HMi}*, which in turn requires a further interpreter* I^{P}_{HMi+1}*, and so on. How is an infinite regress prevented?*

Here, all the terms have been placed at the phenomenal level, hence the 'P' superscript, which is one possible flaw in the argument that will be addressed later. The 'H' subscript refers to the hypothetical system, and 'M' refers to a particular modality, with the i and j subscripts picking out particular types and instances, respectively. Notice that this problem affects the unconscious homunculus proposed by Crick and Koch [Crc99].

Eckardt (p291, 297) suggests the cognitive science solution to the regress problem centres on distinguishing between a representation and the state of a representation. In accordance with Peirce, this translates into defining the nature of the interpreter as the effect of the interpretant on the subject, thus avoiding the regress. Consequently, interpretation of a representation corresponds to the effect it has on the disposition or habits of the subject. However, one problem remains, explaining why this should give rise to qualitative consciousness. This leads to the next two problems.

6.2 The Modal Functional Independence Problem

Sunsets and symphonies cause a spectacular range of sensory experiences. Both are remarkably different, and yet objectively founded on the same neural principles. From these principles arise many neural mechanisms, but why should one neural mechanism give rise to sunsets and another to symphonies? Here two very different sensory modalities are being compared to emphasise that it is the form of the nomic connection between the mechanism and experience that is under investigation. Two shades of the same colour or two tones would have sufficed equally well. The modal functional independence problem is a refined form of the inverted qualia dilemma discussed in chapter two. From the opening hypothesis H-QDC, the qualia from different sensory modalities are suggested to be produced from mechanisms characterised by appropriate sets of disposition instances. Consider a regular array of sensory inputs feeding into an interacting network of dissociated disposition instances, see Figure 2.

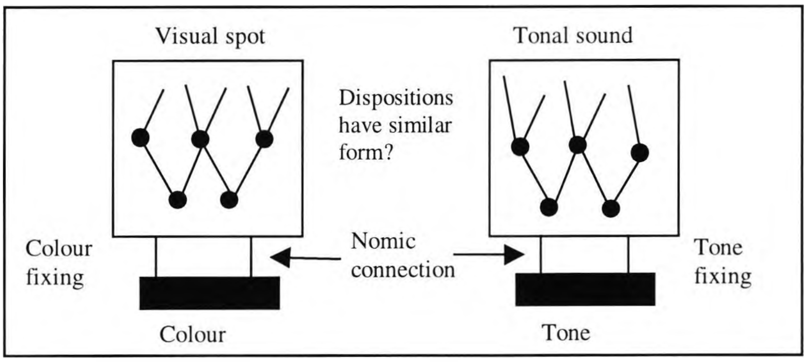


Figure 2. Comparing functionality of sensory modalities.

Comparing modalities by dimensional and functional difference as a route to defining uniqueness of modalities. The diagram shows mechanisms boxed up. Different qualia arise despite the apparent functional-map equivalence of inputs when interpreted as a stationary-representation from within the box.

Although the networks of instances of dispositions for the two modalities appear to receive equivalent inputs in a stationary-representation and functional-map sense, they give rise to different qualia via some "fixing" mechanism, e.g. a pin prick of the neural system causes a universal disposition to have a pain sensation. The problem concerns the *form* of the fixing mechanism that produces qualia at a particular moment in time. For example, imagine a sensory deprivation experiment in which the subject can only see a flashing red and blue light and hear an alternating tone. This is summarised in the following statement:

Statement of Modal Independence Problem. (S-MIP)

For two independent modalities M^D_{H1}, M^D_{H2} *with functionally isomorphic stationary-representational domains and codomains defined on* l^I_H *and* l^F_H *respectively, why are the respective modal interpretations* I^P_{HM1} *and* I^P_{HM2} *experienced differently if they are independent from a first person perspective?*

In this form, it is clear how important a consideration of the ontological levels and perspective involved are. Here functional isomorphism is meant in its conventional mathematical sense – see the appendix. Put another way, given the apparent theoretical external functional-map equivalence of the mechanism underlying qualia for different modalities when interpreted as stationary-representations, why are the qualia experienced as being different? That is, what differentiates the mechanisms that give rise to different qualia? What mechanism would be required to produce a red quale say, and how would this differ from that for producing a blue quale? Part of problem is to determine the nature of the interpretation I^P_H and the applicability of stationary-representations to the functioning of the creature.

6.3 The Spatial-Temporal Limit Problem

This problem concerns how well a theory of consciousness answers questions about consciousness, such as, How and where is consciousness located in the supporting system? Its origins stem from attacks on functionalism as being too weak a claim for a basis for mind. For example, the claim by Hinckfuss that a pail of water in the sun might have fleeting moments of consciousness since through all the atomic interactions taking place the required functional correlation's might be instantiated – see Sterelny [Ster90, p8]. It is also related to the grain-problem. The purpose of the spatial-temporal limit problem is to test whether the explanation for consciousness a theory provides can support counterfactuals. Consequently, a model of consciousness must explain what happens to consciousness when the processes involved are generalised, perturbed and reduced in number or speed etc. Conversely, pushing a model to these extremes helps reveal how and where consciousness is generated.

To illustrate this, consider taking any model of consciousness, such as some of the ideas developed herein, or a simplified version of Aleksander's [Alek96] iconic neural state machine from his theory of artificial consciousness. To remain general, the evolution of each component, or state machine, in the system can be represented by a system of differential equations. In turn the evolution of a network of components or state machines can be represented by the composition of these systems of equations. Assume that this can be written in first-order vector form as $\dot{X} = F(X)$, where X is a vector of variables and F the parameterised system of equations (i.e. the vector field) – see Alec Norton [Geld95].

Applying spatial-temporal limits to this system involves inspecting what happens to consciousness when variables are eliminated or scaled and when the dynamics move from discrete to continuous. This raises questions, such as, is consciousness correlated with a certain set of variables (cf. Crick & Koch's NCC) or processes, and if so, is consciousness still present over some infinitesimal interval of time, $\delta t \rightarrow 0$? The problem here is that as smaller and smaller time intervals are considered, the system's state effectively becomes frozen leaving only its spatial structure as being responsible for producing consciousness. However, this cannot be so since it would imply the right kind of solid structure could be conscious, e.g. a stone sculpture. The response to this is to say that consciousness has a temporal extension of some sort. So, stating the problem more formally using the notation from the appendix:

Statement of Spatial-Temporal Limit Problem. (S-STLP)

Given a set of property instances P^{D}_{Hjx} *that lead to the generation of consciousness at level* l^{P}_{H}*, a theory of consciousness must remain counterfactually coherent for all relevant limits, such as the size of the property set,* $|x| \rightarrow 0$*, and the continuity of consciousness across time intervals,* $\delta t \rightarrow 0$*.*

The advantage of expressing the problem in this semi-formal form is that it draws attention to some of the assumptions being made, such as the level of the processes involved. Here, the property instances, which have been placed at the dispositional level, refer to the processes a theory posits as responsible for generating consciousness. Defending against the spatial-temporal limit problem involves accounting for the necessity of the particular property set and the temporal extension of consciousness. If a theory is to be considered adequate, it should be possible to deduce from it what happens to consciousness around these limits.

7 Examining an Operational Basis for Semiotic Processes

The task set by Fetzer, in his analysis of the laws of human beings and digital machines as discussed in chapter one, is to show how a machine might have the right kind of dispositions in a semiotic sense. That is, that the manifestation of a disposition, such as a habit, is a consequence of causal relations that conform to Peirce's triadic semiotic process. This equates, at least, to showing whether semiotic processes can be implemented in a machine.

With regard to the production of consciousness and cognition, Fetzer [Fetz98b] has suggested a process whereby some source brings about a pattern of activation of neural nodes, which may or may not be familiar to the system. Their familiarity is roughly indexed by the ease with which they are subsumed by corresponding concepts, which are sets of habits of mind and habits of action. The subsumption of these patterns by suitable concepts (which takes place within some context of pre-existing states, including motive and beliefs of that system) yields cognition as its effect. This section and the next chapter attempt to characterise the structure of these patterns of activation from an operational perspective by approximating them through the superposition of discernible dissociated disposition primitives. The superposition of fuzzy sets of disposition instances is examined further in chapter five.

Following from the opening discussion on graded dispositions, it is plausible to suggest that an instance of a dissociated disposition can be implemented through a mechanism involving temporal-representations and functional-flows. From the statement for the hypothetical system (S-HS), this mechanism is referred to as the dispositional level. The purpose of this section is to analyse the viability of this suggestion and to see how this helps refine the nature of the level. To guide the analysis the previous problems are considered (in reverse order).

7.1 Redressing the Spatial-Temporal Limit Problem

This section suggests the spatial-temporal limit problem can be partly redressed by observing that the mechanism of mind is more likely to be based on temporal-representations and functional-flows rather than functional-maps and stationary-representations.

7.1.1 Implementation Combinatorial Complexity

Starting at the implementation level, it follows from the cellular neural structure of the brain that the apparent sophistication of qualitative consciousness arises, in part, through the large-scale composition of relatively simple processes rather than the interaction of a few complicated processes. Hence, a defence of the spatial-limit problem is unlikely if it suggests a compositionally irreducible holistic property. This leads to postulating that instances of dissociated dispositions are expected to be simple rather than complex processes.

A troublesome point is that it would appear as though it takes an intricate set of dispositions to uniquely determine and distinguish a sensory modality. To some extent, this view assumes a hierarchy wherein the low-level dissociated dispositions instantiations are built upon to give top-level qualia. However, the system has to recreate the experience of the three fundamental features of the perceived reality: time, continuity, and extension, which are inherent in each modality. This suggests that the parallel topologic structure of networks of disposition instances across the modality is also significant. This raises a question as to whether the network structure is wide and shallow, rather than narrow and deep with respect to disposition instances per organisational layer versus a hierarchy of layers. This question is returned to in the next chapter.

7.1.2 Implementational Basis of Dissociated Dispositions

When graded dispositions were discussed at the start of this chapter, Hopfield's neural net was used as an example implementation of one kind of disposition. However, this was an idealised neural network supporting an inflexible set of dispositions. Consequently, one route to refining the operational nature of instances of dissociated dispositions, when viewed as transient properties, is to consider other ways in which neurons might support them. There are two ways in which this might be possible: a) directly, if an instance of a disposition is paired with a neuron, such as a simple input/output device that instantiates a single-case disposition to produce output O when subject to input I, and b) indirectly, if the neuron acts as the support for the disposition. The first way is ruled out because it is at the wrong level, the functional-flow of the instantiated dissociated disposition would have to be equated directly with neural events, i.e. each neural event would have to express a dispositional tendency that reflected the tendency of the dissociated disposition's reference class description. This is nomically and physically unlikely since, firstly, all neural events are, broadly speaking, atomic dispositional events of the same kind, and secondly, it would require synchronising individual neural events in the system. This would be susceptible to signal degradation and noise. To avoid these operational difficulties techniques such as frequency coding may be used. *In general, an instance of a dissociated disposition requires a support mechanism that is able to sustain a range of dissociated dispositional tendencies and is further able to support the faithful execution of these tendencies, which implies a degree of dynamic stability at that level of description.*

Consequently, the second way, in which a neuron acts as a support for an instance of a dissociated disposition, is more appropriate since it equates functional-flows with the statistical behaviour of the neuron, see Figure 3.

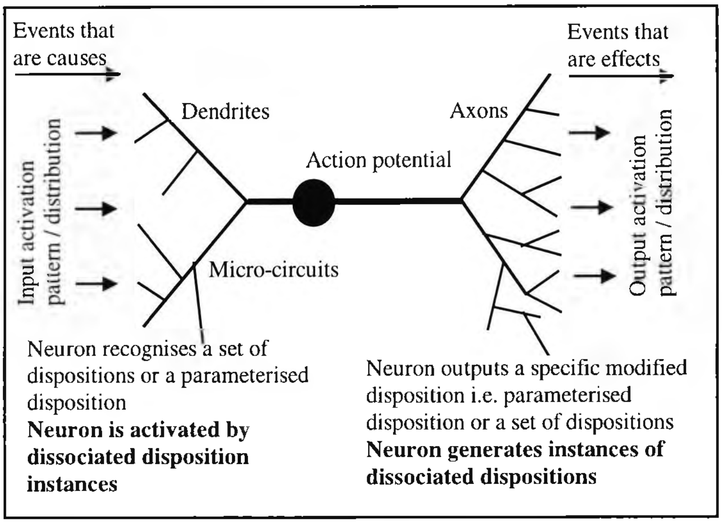


Figure 3. What is the relationship between neurons and instances of dissociated dispositions? Is a neuron paired with a disposition instance, or does it act as its support?

According to this view, the neuron responds to a set of dissociated dispositions influences or a parameterised dissociated disposition influence, in a continuous fashion, as statistical spatial-temporal patterns across its dendritic system. The neuron transforms this via soma processes, again in a continuous fashion, while modulating the nature of the dissociated disposition instance, or set of disposition instances, currently playing across its axonic outputs. In effect, the dissociated disposition instance, when activated by a certain input pattern of neural activity, brings about a certain output pattern for the system.

There is a potential problem with the second possibility, of a neuron acting as a support for instances of dissociated dispositions: it is too specific and, therefore, also susceptible to signal noise in some cases. While the large number of dendritic and axonic inputs and outputs may result in statistical continuous behaviour, the bottleneck is the soma, it propagates signals in a discrete sequential fashion. What happens if the neuron fails? Does an instance of a dissociated disposition supported by one neuron have enough excitatory power to activate subsequent disposition instances? That is, can a single neuron support going from inputs of kind x to outputs of kind y for a single neuron system? What of operational constraints arising from lateral neural processes, or constraints that are more complex and require a small network to implement? Therefore, for some dissociated disposition types a better match would be to have an assembly of neurons as the support:

Statement on Dissociated Disposition Instance Subassembly Support. (S-DDISS)
An instance of a dissociated disposition is operationally supported by an assembly of processes at the implementation level, where processes are functionally analogous to neurons with a high interconnect able to support a range of potential dissociated disposition tendencies.

In effect, an assembly of processes (cf. patterns of neuronal activation) will have robust properties that single processes may not possess, which is one of the reasons for preferring distributed representations.

7.1.3 Instances of Dissociated Dispositions as Patterns of Parameter Variations

If an instance of a dissociated disposition is represented by a set of parameter variations to a neural net, should the disposition instance be treated as a transitory pattern of activity where the pattern profile defines the type of dispositional tendency? Alternatively, does the parameter set characterise a graded disposition where parameter changes represent variations to the nature of the tendency? If a disposition instance is viewed as a spatial-temporal profile, then is it correct to view it moving through the system, so that a neural net transforms it as it progresses? One limitation is that a stable quale needs to be defined in a stable, perhaps localised way. If the instance of the dissociated disposition moves, it would have to chase its associated (i.e. causally connected) dissociated disposition instances through the system! Therefore, it is operationally reasonable to suggest instances of dissociated dispositions are a localised spatial-temporal profile.

Statement on Localised Nature of Dissociated Disposition Instances. (S-LNDDI)

Instances of dissociated dispositions are localised to a specific region of the supporting medium.

However, see the section on modelling instances of dissociated dispositions in chapter five. Consequently, if the underlying structure is poly-dispositional, i.e. it can support different types of dissociated dispositions, then some of the parameters have to stand for the type of disposition. The type refers to the nature of the dispositional tendency – disposition primitives are considered in the next chapter. There may even be degrees of poly-dispositions that depend on the function and context.

7.1.4 Requirement for Structured Mechanisms for Dissociated Disposition Instances

Examining the phenomenal level, logically the qualia of sensory modalities should be independently realisable since, for example, it's implausible to suggest deaf people see things differently. This suggests dividing into two classes the disposition primitives upon which the particular modalities are constructed: a common set for signifying system relative universals (space, time, order etc.), and a modality specific set. Since modalities appear to be possible independently of each other, semiotic grounding is probably not inter-modal. In addition, since the primitive universals mentioned above are common to more than one modality it suggests they can be ground independently of a modality. Notice, though, that these primitives may require an extensive operational structure spanning the system. The next chapter examines some of these primitives.

The mechanism must rely on independent, internal, relative and relational processes – more will be said about relations and processes in chapter six. The independence requirement follows by noting that colours can be perceived in isolation. The internally ground requirement because modal experiences can persist long after the external stimuli. The relative requirement is a logical necessity – see Hering's colour model below. Finally, countering inverted qualia arguments, the relational requirement requires, when viewing the sequences of a graded dissociated disposition as functional flows, a functional-flow ordering exists or is definable on the functional-flow basis. To achieve this the dissociated disposition requirements suggest the qualia axes (i.e. perceptual space dimensions) are typed by sets of dissociated disposition instances, where the sets change with position along the axis – see below.

Requirement for Structured Mechanisms for Instances of Dissociated Dispositions. (R-SMIDD)

The mechanism behind the instantiation of dissociated dispositions relies on independent, internal, relative and relational mechanisms.

Notice that this does not prevent particular qualia being associated with a set of universal dissociated disposition primitives. Here qualia are said to be *associated* with instances of dissociated dispositions because it cannot be assumed at this stage that they are generated directly by them. This is a weaker claim than the NCC hypothesis. In fact, the next chapter suggests that qualia are more naturally associated with a structured dispositional process.

7.1.5 The Constant & Evolving Nature of Qualia

Suppose the conscious experience of a quale has a temporal extension, and that it is possible to discriminate between the presence and absence of the experience. However, suppose the qualia is normally experienced extending over time with no notion of its absence, e.g. forever in light, or more subtly, the perception and subsequent qualia experience of spatial depth. For example, we seldom contemplate the absence of space in a visual experience. Consequently, is a semiotic and therefore dispositional-based signification of time needed in order to be semiotically conscious of a quale?

It would seem that a notion of time is not required for producing semiotic states and that only a stability of states is required, e.g. an entity in the real world is tracked consistently by occurrent instantiations of dissociated dispositions while evoking qualia. Hence, *a constant quale must be produced by an occurrent process* – in that to produce what is experienced as a constant sensation a network of dissociated disposition instantiations is continually interacting. For example, an analogy could be drawn with particles in a medium oscillating, not necessarily synchronously, to produce standing waves. This would fit in with a functional-flow view of dissociated dispositions. However, this is not a stationary fixed point in a memory. It is a conceptual problem in that instances of the dissociated dispositions can only exist entwined with the passage of time, being based on statistics of changing causal processes. What are the implications of this for dealing with time?

The above points suggest that, operationally, the temporal pattern of the evolution of instances of dissociated dispositions is significant. This lends itself to a description in terms of temporal-representations and functional-flows. However, to completely redress the spatial-temporal limit problem requires considering these factors in conjunction with conditions required for generating consciousness. These conditions are examined in the next chapter.

7.2 Redressing the Modal Functional Independence Problem

The modal functional independence problem concerns the qualitative aspects of the contents of consciousness. In what follows, the dispositional processes behind these contents are examined and treated as pre-qualitative perceptual entities. The next chapter suggests how qualia, and therefore, qualitative consciousness arises from the operational structuring of these processes.

This section redresses the modal functional independence problem by considering how the dispositional level might be operationally structured in order to give rise to sensory modalities that are qualitatively distinct. In this regard, the nature of a perceptual entity is examined from an operational perspective. The notion of entities leads quickly to a discussion of signs and the binding problem. This, combined with the analysis of the contents of consciousness, suggests that semiotic processes can be implemented as networks of instantiated dissociated dispositions in terms of temporal-representations and functional-flows.

7.2.1 Requirement for the Necessary Categorisation of Signs

The binding problem is often considered from the bottom up: how is a collection of properties integrated into an experienced whole? Looking at the problem in this way can lead to the properties being taken as primitives that are already in consciousness with layers of beliefs added later when grouping them into compound signs. This results in a notion of binding direction and potential discord, with properties becoming conscious first and having some subsequent possibility of being incorporated into signs.

In contrast to this bottom-up approach, it is suggested, with respect to the operational structure needed to generate consciousness, i.e. the operational ability to use signs, that a mechanism is in place to the effect that every discriminable point in the perceptual space belong to a sign of some kind, and that the sign's kind is also characteristic of a category manufactured and determined by the system to categorise the entities making up the mind's reality. An equivalent way to say this is that perceptual space is a categorisation of signs. Some signs may belong to an 'unknown' category. For example, the mind may see an object from a dimly lit, obtuse angle, and experience it as a 'something' until it is able to categorise the object more accurately. In Peirce's terminology, this would be analogous to an "abnormal" sign – see Fetzer [Fetz91, p62]. An entity cannot be experienced in consciousness without having been operationally portrayed by a sign and therefore categorised in some way.

Requirement for the Necessary Categorical Nature of Signs. (R-NCNS)

As an operational prerequisite for the generation of consciousness (i.e. the ability to use signs), the system will have constructed a signified reality in which it has signified entities by networks of interacting signs whose kind (dispositionally & collectively) portrays the category of the entity. Consequently, only signs, whose supporting mechanism includes categorical dispositional properties, are consciously experienced.

The act of categorisation also imposes a unity (or individuality) on signs (in addition to their identity). However, notice that the category boundaries could be graded i.e. fuzzy categories. In terms of category organisation, the perceptual space has a fixed or innate organisation, it is an overlaid hierarchy of categorised signs – the justification for this statement is explored later. The major implication of this categorisation requirement, coupled with this organisation assumption, is that there is no direction of sign entry or possible discord. In addition, signs would have an equal prominence irrespective of their category. This also implies that semiotic consciousness is not due to a regress of higher order beliefs, suggesting instead that the network of signs has a shallow hierarchy of fixed depth. Hence, a primitive colour patch has the same prominence as a compound sign, the respective points in the perceptual space are impelled accordingly. Therefore, in regard to the binding problem, primitive entities will either be categorised as independent signs, or additionally categorised as part of a compound sign.

Statement of Phenomenal Level Sign Completeness. (S-PLSC)

There is simply no place at the phenomenal level for uncategorised signs, and hence qualia that do not belong to a category or individual sign.

This statement suggests that, even for vague sensations, such as dull aches and faint whispers, that to be conscious, i.e. to be a usable sign in the system, the mechanism supporting the sign will include dispositions that categorise the sign with respect to the signified reality. Chapter four refines the dispositional structure of the categorisation process.

7.2.2 Atomic Signs versus Compound Signs

An experienced red spot has perceived properties (as opposed to its real properties, or the properties of the instances of the dissociated dispositions that signify it) and in turn qualia for depth, luminosity, colour and so forth, although it is only experienced as a single sensation. In contrast, a compound sign, such as the image of a tree, contains many signs. Therefore, an atomic sign is defined as a set of qualia that are irreducibly bound – irreducible binding is discussed shortly. The expression "network of instantiated dissociated dispositions" will be used to refer to the set of entangled dispositions instances that are immediately responsible for supporting an atomic or compound sign (in conjunction with the gross process structuring needed to generated qualitative consciousness as discussed in the next chapter).

Atomic and compound signs can be treated as building blocks for the phenomenal level. With atomic signs, the instantiation of their dissociated dispositions is fixed by the environment, e.g. a visual spot always has a certain set of disposition instances. With compound signs, inclusion in the signified reality is fixed by environmental and innate knowledge factors. A compound sign could be displaced by other signs, a change in interpretation dictated by the environment, or from decisions made in the post processes, such as selectively attending to its component signs. Essentially, the set size of the sign is not affected by introspection, but only via the environment and a shift in attention. For example, a tree is a tree, unless it is suddenly removed from the scene or the creature attends to a particular aspect, thus changing the contents of the scene.

7.2.3 Entity Based Beliefs

There is a distinction to be made between beliefs about (mental) entities or signs and beliefs about the phenomenal content of signs, i.e. qualia. The signified perception of the Self in the signified representation of reality can only experience, from a first person perspective, direct beliefs about entities categorised as signs, and not individual qualia. Remember that a quale is a phenomenal aspect of a sign. This effectively imposes a degree of granularity on the composition of qualitative consciousness. However, as will be shown later, this granularity can be graded and fuzzy, with signs blending in two ways: through continuous semiotics (see chapter five) and by the absence of qualia for experiencing discontinuities.

It is important to understand the system relationships involved between an entity perceived in consciousness and the qualia that make up the experienced entity. The following distinction is made:

Statement of Entity Based Beliefs. (S-EBB)
The system has (direct) beliefs about the entity being signified (as an atomic whole) rather than its individual qualia.
This implies a degree of system structuring is in place such that beliefs can only ever be about categorised entities (i.e. signs). Here an entity refers to a bound set of instantiated dissociated dispositions with associated networks of disposition instances that portray the identity of the entity. It is possible that by introspection, the signified representation of the Self may include inferred beliefs about the component qualia of an entity, but these in turn will be distinct signs in themselves. Notice that while the signified Self does not directly have beliefs about the sensation of, say, the depth of a red spot, it still experiences the depth as part of the signified perception of reality.

7.2.4 Binding Qualia into Entities: The Binding Problem

The figure-ground and grain binding problems concern how the properties of signs and qualia, respectively, (i.e. the phenomenal manifestation of the underlying network of dispositions), are bound into an entity experienced as a whole. The following suggests this can be achieved operationally by the actions of an instance of a dissociated disposition for binding: D_B. However, the primary problem concerns the nomic manner in which the set of dissociated disposition instances over which D_B ranges is determined in general from a first person perspective. Secondly, it is concerned with how the D_B's are represented in a dynamic system from an operational perspective.

7.2.4.1 Binding: More than a Parallel Process

It is suggested that the mechanism by which the components of an entity are experienced as a whole takes place in two stages. Firstly, through parallelism, all the component qualia are experienced at the same instance. This would support Crick and Koch's suggestion for synchronous events. Secondly, in constructing a utilitarian signified perception of reality, a dispositional constraint is forced on the signs to the effect of instilling the experience of a single instigating source for the component qualia.

Is there a need for a binding disposition, or is binding possible simply through parallelism and synchronous activity? The latter may seem to be the case, but a binding disposition is still needed. To demonstrate this, consider the quale for a visual patch of colour. It has a partial set of primary properties modelled by instances of dissociated dispositions, such as {colour RGBY, spatial position, brightness, texture, and continuity}. By introspection, the spot may be experienced as being at location X, of colour Y, texture Z etc. All these apparent attributes, while discernible, are part of the spots secondary qualia. They are part of secondary consciousness and inferred after the experience, after the moment when the primary properties are bound. Consequently, a mechanism must be in place that associates these primary properties together. Therefore, a binding mechanism is at work despite the parallelism of the initial experiences. Hence, the reason for a binding disposition is stated as follows:

Statement of Necessity for Binding Dispositions. (S-NBD)

Binding dispositions are necessary in order to associate qualia into entities so that secondary qualia may be produced.

This is a type of logical binding where what is bound depends on the objectives of the system, in contrast, nomic binding would be a constraint imposed by the laws of the system. The statement could have been cast in a more fundamental way, in that semiotic processes involve the interaction of signs implying some operational recognition and encoding of the unity of a sign to enable its subsequent capacity to influence other sign processes. However, binding need not be an all or none process, crisp binding is only necessary when distinct entities need to be comparatively discriminated. Although the converse does not follow in that diffuse entities (cf. Fetzer's remarks on unfamiliar patterns of neural activity) may need to be crisply bound during the categorisation process.

7.2.4.2 Irreducible versus Associative Binding

Are there different types of binding dispositions? For example, an atomic sign, such as the smell of a rose, or a just discernible colour patch, must be distinguished from a compound sign such as a rose, its petals, stem and so on. For instance, a petal has a spatial extension since many colour patches per petal can be discerned. This indicates two types of binding: the first, irreducible binding, in which the (dispositionally generated) properties that are bound are not directly influential on (or accessible to) secondary beliefs, and the second, associative binding, where the mind can deduce, upon reflection, that the components of an entity have been bound.

7.2.4.3 Instances of Dissociated Dispositions are Processes, Not Values

An instance of the binding disposition, D_B, does not have an output value such as "D_B=1" meaning "entity", or "D_B=0" meaning "no entity". Rather, the instance of the disposition is embedded within a causal cyclic path in the system so that its invocation gives rise to the experience of an entity – this is discussed shortly. Dissociated dispositions are characterised in terms of functional-flows rather than a measured value (cf. H.H.Pattee's [Pat96] notion of measurement).

7.2.5 Requirement for a Continuous Coding Constraint

In typical situations, the evolution of instances of dissociated dispositions is constrained by the need to represent the continuous or graded change in the properties of experienced entities. For example, imagine a blue expanse, the sky, and slowly changing a few areas in brightness or colour whereupon patches start to be discerned. It is more natural to think of some kind of analogue process from which a few patches become distinguishable, rather than a jump in analogue processing wherein the new patch boundaries and properties have to be freshly coded. This implies a requirement for a continuous coding constraint is in effect with regard to binding:

Requirement for a Continuous Coding Constraint. (R-CCC)

Wherever possible, stationary-representations and functional-flows should be capable of incremental adjustment following minor input variations, rather than requiring complete recoding.

7.2.6 Binding: A Question of Structure or Hierarchies?

This subsection starts to consider some of the system structuring required for constructing the signified perception of reality from instances of dissociated dispositions. The next chapter pursues this topic in more detail.

Different kinds of things (objects and properties in the world) induce different patterns of neural activation, which come to signify specific signs for those systems through the acquisition of corresponding habits of mind and habits of action which become those signs' meaning. In order to replicate this semiotic process requires characterising the form of these patterns in some succinct, implementable way. In general, signs will be produced by different sets (and therefore networks) of instances of dissociated dispositions, e.g. to account for the form of the different grounds (iconic, indexical & symbolic) and their particular qualities. Hence, the set size will vary depending on the compound sign. When constructing the signified reality there could be two levels of organisation. Firstly, binding of atomic signs in the perceptual field based on fixed sized disposition sets and a fixed hierarchical structure for assimilating compound signs. Secondly, binding of atomic signs into structures and compound signs based on variable sized disposition sets. A variable set suggests a fluid structure, that instances of dissociated dispositions are composed on the fly. This means that within the system the set must be continually evolving. Hence, instances of dispositions would have to be represented such that they could be related to each other in a generic manner; a physical mechanism flexible enough to support a multitude of disposition types, cardinalities, and the ability to smoothly transform between connected subsets (e.g. Lie groups) – cf. the requirement for continuous coding.

Trying to construct signified representations of reality from instances of dissociated dispositions across distributed structures raises a number of questions. Firstly, consider the operational problems, independently of any particular animal or machine, when processing a dynamic aspect of reality. For example, the organisation of processes in visual consciousness might be expected to involve a highly pipelined process wherein the visual scene is broken down into its component objects in a systematic fashion. Alternatively, it could involve a process wherein representations are housed on varying sub-processes, where the relationship between objects is part of the coding. This might lead to the need to bind across distributed representations.

How stationary is this coding, physically, when the field of view is panned? For example, is a tracked object moved in a continuous fashion onto topologically neighbouring processes, or is it always 'played' by the same instances of dissociated dispositions fixed in the process topology structure, with its position and relationship to other objects being captured by topologically bound disposition instances? Hence, are the flows of dissociated disposition instances constrained in a topological closed fashion, e.g. on a torus? For a small area of the scene, the set of objects enclosed may change in number very quickly. Is there a finite amount of object labelling resources per unit area of the topology structure, which in simple scenes are under used? The following sections and the next chapter address some of these questions.

7.2.7 The Logic of Binding

This and the next few sections return to addressing the binding problem more directly.

To recap, given a set of instances of dissociated dispositions underlying a quale, how are they bound into an experienced whole? The grouping may have a natural (nomic) or an artificial (habitual) or even an accidental (coincidental) basis, not to mention those that have a logical basis (in syntax or in semantics). It was suggested that a unifying D_B bind them. Its qualitative experience is of unification. Now this leads to another problem. Given the experience of the binding D_B, how does the mind know the D_B refers to its set to be unified? That is, $D_B \rightarrow \{D_1, D_2, D_i,\ldots,D_n\}$. How does the mind know D_B refers to the D_i in the experience of unification? Conversely, how does the mind know a D_i is bound by D_B? This is a weaker problem. Here the nomic form of the relation between the instances of the dissociated dispositions and the phenomenal aspects are of interest. Some logical constraints on D_B are:

1) The D_B binds the D_i , implying the D_i could have been unbound.
2) The D_B binds the D_i , implying they are not bound to other D_m or D_B's.
3) The D_B is an absolute unification, as opposed to being relative to the signified perception of the Self, i.e. the D_i are treated as referring to a distinct thing independent of a Self, e.g. a colour patch can be referred to as an ‘it’.

Put succinctly, at the phenomenal level, on the one hand a qualitative experience is generated to the effect “this D_B relates to these D_i”, and on the other what is over ridingly experienced is “that entity,” i.e. the D_i and D_B in parallel as one. It is the D_B that gives this illusion. Hence, the D_B is doing something like “these D_i denote this entity.” The actual entity referred to is defined contextually.

Consequently, the binding problem can be rephrased as: D_B is equivalent to “these D_i denote an entity,” for example “these D_i denote a colour spot.” This involves resolving how “these D_i” are specified. More generally, explaining the operational mechanism, terms and relationships behind the parallel experience of “these D_i denote an atomic sign,” and “this D_B relates to these D_i.” There are at least two possibilities.

Firstly, there could be a topology of processes where each atomic sign always has a fixed D_n set to which a D_B, within the set, refers. Here, irreducible atomic signs and associated qualia are being dealt with, not compound signs. Consequently, how are compound signs dealt with by D_B's? Secondly, binding could rely on a form of topologic proximity. The D_B effectively brings about the experience "The D_i within this topological sphere describe an entity." For instance, one approach would be to tag all D_i with a location indicator and diffusely bind localised D_i. In the extreme, this would reduce to the topology approach.

7.2.8 Irreducible Binds

An important step toward explaining how binding is achieved involves understanding the nature of the beliefs that signify an entity is bound. This section considers the binding problem from this perspective.

The above logic implies that binding becomes a matter of association: "This D_0 is associated with these D_i," where D_0 denotes an instantiated dissociated disposition whose purpose is to indicate the entity's unity or wholeness. This de-referencing is achieved on a demand-driven basis by an attentional process, i.e. a pre or post process that links D_i to D_0 or D_0 to D_i as necessary when required. The nature of the association might be iconic, indexical or symbolic, where which occurs depends on the system's context of other properties.

Hence, in parallel, an atomic sign is experienced as: the entity portrayed by D_0, and the D_n set, with a parallel D_B, resolved by feedback processes, which gives the sensation of "these D_i bound to D_0". This is made equivalent to "this D_0 is bound to these D_i" by the first person perspective view imposed on the signified perception of reality (see the statement below on the perspective view of consciousness), i.e. the experience: $D_0 \xrightarrow{bound-to} D_i \equiv D_i \xrightarrow{bound-to} D_0$, where the first person perspective view represented through signs by the system forces the equivalence. This makes the binding D_B irreducible. Remember to differentiate the sensory experience from the mechanism. The $D_0 \xrightarrow{bound-to} D_i \equiv D_i \xrightarrow{bound-to} D_0$ relation is forced to an illusionary bi-directional one: $D_0 \leftrightarrow D_i$, even though the mechanism is "Entity $\rightarrow D_i$". Hence, "These D_i bound to D_0" becomes "Are-bound: D_i, D_0,...", which is analogous to "f(X)" as in "f applied to X."

The previous paragraph suggests an answer to the binding problem may be found by observing that since the system controls the perspective on reality experienced by the mind, and what inferential mechanisms are available to the mind, it can dictate what relations and beliefs the mind is able to experience between an entity, its properties and the signified perception of the Self. To sum up:

Statement of Irreducible Binding. (S-IB)

The primary beliefs about qualia that the mind has conform to a signified perception of reality in which a directional perspective is imposed and maintained by a logically asymmetric, non-systematic set of dissociated disposition instances and corresponding qualia.

For example, to elaborate further on the apparent asymmetry in the experiences as divulged to the mind, consider a classical two-valued propositional logic. In order to have the concept of "A & B are bound," the mind would also need the concept of "A & C are not bound." That is, if there is a set of instantiated dissociated dispositions, the mind can say neither they are bound nor not bound unless it can conceptually discriminate between the two situations. Therefore, in terms of qualia, both bound and not bound qualia would be required. Alternatively, since the system controls which logic it follows, a compound sign would be sufficient such that the mind only has the experience of boundness when contextually appropriate. Every concept has to be constructed. For example, if there exists a mechanism for the belief that "X is bound", if it is the case that X is not bound, it cannot be inferred that "X is not bound", unless there is a mechanism for this as well. That is, bound is a unidirectional true only belief: it cannot be inferred "not bound" even if it is the case that "not bound" (this would be uninstantiated). The mind would have to have the belief and supporting mechanism for "not bound". This is in contrast to the systematicity suggested by Fodor and Pylyshyn – see Fodor [Fod88].

As a further example, consider a solution to the "Next-to" binding problem: "A is Next-to B." Applying the above reasoning, "These A & B are bound by Next-to" becomes "D_B: (A,B,Next-to)," i.e. a (first order) binding disposition for instances of the dissociated dispositions A, B and Next-to. The terms in this expression are all instances of dissociated dispositions i.e. processes, they are not solitary neural states nor terms in a propositional calculus. In other words, D_B corresponds to a functional-flow influence on subsequent instantiated dispositions, which has the desired binding effect.

7.2.9 Requirement for Operational Decisional Singularisation

The previous sections suggested that one reason why binding appears as a problem stems from taking a first-person perspective and analysing the problem from the bottom up, going from unbound properties to an experienced whole. This section suggests that in the system, the phenomenal level is *only* able to contain the perception of entities in the first place.

Binding should be raised to a central feature of what the system does. It is trying to identify and categorise entities in the environment. Hence, binding is not just a side issue, but an important task for survival. It is more important to the system to identify an entity as a whole, rather that as an unconnected collection of properties. In some sense, the entity is an emergent property of the sensory properties by which the entity was first identified, i.e. a particular arrangement of permanent reference properties – see Fetzer [Fetz86, p100]. To the system, it is as though there is an order of perception: an entity is experienced followed by the experience of its properties either as part of the entity or as entities in their own right. This means that qualia, as the phenomenal content of signs, are therefore necessarily part of the conscious experience of an entity. In other words, a quale, which is part of the phenomenal level as a consequence of a certain semiotic process, has to be part of an entity since all signs nomically have an identity in the system. Remember, qualia are best understood as dispositions of a specific kind that reflect the nature of the respective sign they are part of, and consequently, being part of a sign for the system, they will influence post processes – this is necessary for them to be part of the phenomenal level and is discussed in the next chapter. Therefore, the binding problem should be interpreted from the other direction: *as an operational issue rather than a phenomenal one*. This is raised to the following requirement:

Requirement for Operational Decisional Singularisation. (R-ODS)

The primary task of modal processing is singularisation, i.e. the assimilation of instances of dissociated dispositions into signs as part of a decisional mechanism for inducing subsequent processes of an established beneficial nature.

Singularisation includes integrating properties from different modalities into signs (and therefore qualia) making up a unified compound entity, e.g. binding sound emanation to the visual location of a physical object. Singularisation can be compared with traditional classification theory. Notice that the unity-of-consciousness, is achieved through a different process, specifically, by signifying the Self as a unified perspective on the signified reality.

Viewing binding as a route to singularisation, the task for survival is to identify and discriminate entities in the environment. *Hence, binding is understood slightly differently: in a modal field, an atomic sign is not the building block for entities, but the smallest discriminable entity.* The modal field is processed with respect to discriminating and identifying entities. This means the binding problem has to be looked at from the other direction. Consequently, start from the premise that the phenomenal level only ever contains the perception of entities. Some entities might be of an unknown type or mere aspects of other atomic signs. Focusing on a compound entity leads to its component qualia. A compound entity and its sub-components would all be contained by the same phenomenal level at the same time.

There can be many overlays of instances of dissociated dispositions during singularisation – a hierarchy of binding and association dispositions. The first overlay to bind the properties signified by other instances of dissociated dispositions, the second to associate the bound dispositions with an entity label at that stage, e.g. colour patches, tree trunks and leaves, tree, forest. Although this is structured as an apparent hierarchy, to the system this can be a flat level, i.e. instances of dispositions all at the same level. Singularisation could represent entities as a mere list, overlaid in parallel in the topology structure.

To sum up, the dispositional perspective treats entities as primitives of the phenomenal level. Every quale is bound to an entity. Attentional mechanisms would deal with entities e.g. they focus on an entity, and then recursively focus on its sub-properties, leading to its sub-entities and qualia.

7.2.10 Requirement for Reconstruction of Reality

Once entities have been identified as such (i.e. through the invocation of particular dispositions), they have to be categorised for the purposes of the system, such as friend, foe or food. The justification for this is discussed shortly, for now it is stated as follows:

Statement on the Origin of the Signified Perception of Reality. (S-OSPR)

A goal of the system is to operationally categorise entities detected in the environment in order for it to make decisions. This amounts to the implementation and dispositional levels being so configured such that, during the course of their normal operation, a signified perception of reality is constructed through the interaction of instances of dispositions at the dispositional level that is manifest as a network of signs at the phenomenal level and includes the Self as a compound entity. This signified reality corresponds to what the mind experiences as (the) phenomenal space.

Remember that entities are embodied as networks of interacting instances of dissociated dispositions. One of the requirements for a system level developed in the appendix is that all configurations of the system must satisfy the axioms of the system. This can be applied to the coherency of the signified reality, which could be argued to entail having to signify a proper subset of reality necessary for the creature's survival, e.g. the signified reality must meet certain levels of consistency, assuming non-malfunctioning, in order for reliable predictions to be made on average. So, the sign user, necessarily, at least as an idealisation within the present context, is assumed to be consistent, thus:

Requirement for Reconstruction of Reality. (R-RR)

The system has to reconstruct every facet of reality that is instrumental to providing it with a consistent signified perception of reality.

The word "reconstruction" is used rather than "representation" to emphasise there is no inner homunculus to which the signified reality is presented. Instead, there is the dispositional level, the network of interacting dissociated disposition instances it supports, and how these generate the perspective view of consciousness, described shortly – also see comments on functional representations below. The word "instrumental" figures in the requirement as a filter for irrelevant aspects of reality – cf. requirement for maximal-specificity in Fetzer [Fetz93, p60]. Finally, by "consistent" is meant that in operational terms the system is mechanistically deterministically consistent in that, for example, a stimulus X will always produce some primitive signs Yi (with some deterministic or probabilistic tendency), in a given context Z.

Reconstruction must be from first principles since nothing can be assumed *a priori*. Therefore, as well as reconstructing obvious qualia, such as colour, other qualia have to be set up to account for the dispositional impact of entity properties such as 'conceptuality' and 'temporality'. This dispositional process instils the characteristics of these properties into the signified entity as dispositional capabilities rather than second order beliefs, such as the concept of concepts, or the concept of time. It also implies that the content of perceptual dimensions must be ground and calibrated. The system even has to dispositionally reconstruct the notion of metric distance and difference for the dimension scales.

The system may reconstruct all aspects of reality that are instrumental to consciousness, but not all of these aspects will be conscious. There are some aspects of the signified reality that the mind normally does not have direct believes about, for example the lack of awareness of the fact that every visual object in a scene is also accompanied by a transparent category labelling – see section on singularisation. Moreover, there are those aspects that are not signified explicitly, for example physical laws and constraints (cf. Gulick's semantic transparency). These have an implicit influence in governing the shape of the phenomenal space etc.

7.2.10.1 Qualitative Continuum

It is suggested that the number of qualitative experiences that make up the consciousness of different creatures lie on a continuum. Qualitative consciousness is composed of a myriad of qualia and their supporting dispositions. These incrementally give rise to a unified sense of reality. The sophistication of consciousness develops as more interrelated qualia are added. Due to the structuring effect of the requirement to signify reality, there may be gaps in the population of realisable creatures along the continuum – see Sloman [Slo97].

7.2.10.2 Resolution of Consciousness

This highlights the difference between consciousness as a continuous activity, and the resolution of consciousness. Resolution refers to the discriminatory power of the system at categorising entities from which the signified reality is constructed. For example, the mind may only perceive events that are separated by a few milliseconds, or visually discriminate points separated by a few hundredths of a millimetre. Conversely, just as the system is responsible for signifying what is perceived as disjoint entities, so it must produce a signified process for what the mind believes is continuous.

7.2.11 Requirement for Functional-Flow Signified Perception of Reality

The requirement for reconstruction from first principles in conjunction with the conclusions in the spatial-temporal limit section suggests the following requirement:

Requirement for Functional-Flow Signified Perception of Reality (R-FFSPR)

The nature of representation (at the dispositional level) of the signified perception of reality in the system is through functional-flows and temporal-representations.

Notice that a functional-flow is not with respect to a subject or interpreter. This requirement suggests that the representation of the instances of dissociated dispositions will be as causal sequences (functional-flows) and that from this aspects of reality will be signified as evolving causal structures (temporal-representations).

Reality constraints on the mechanism have a structuring effect. The integrity of the mechanism in signifying reality and its perspective on reality has been driven by evolution to arrive at a 'realistic' encoding. This is summed up by the following statement, which is concerned with the nomic direction of the relational connectives set up in the system, and should not be mistaken as referring to epiphenomena:

Statement on First Person Perspective View of Consciousness. (S-FPPVC)

The experience of consciousness (the phenomenal level signified reality), from the first person perspective, is made to experientially appear, by way of relational connectives, as though from a (contemporary) representation third person perspective by the (dispositional level) mechanism because this is an evolutionary economical representation of it (the environmentally situated creature/system) in reality.

A contemporary representation is a stationary representation viewed from a third person perspective. Notice that *evolutionary economical* also implies that, through adaptation, certain aspects of the signified reality are in some sense a faithful portrayal of reality. Following Fetzer's definition of semiotic consciousness, as the ability to use signs, this statement suggests that a particular relational perspective will be imposed on the ability that determines the form and manner by which the signs can be used. This point is returned to in chapter six when the grounding relation is discussed.

Chapter four will use this "presentational" signified perception of reality to explore the system structure required to generate qualitative consciousness. This view fits in with Kant's ideas on how the nature of the experienced reality is determined by the perspective imposed by the mind – see Fetzer [Fetz96, p114].

7.2.12 Requirement for Entity Distinguishability

This section draws together the above points in regard to the modal functional independence problem. Firstly, from an operational point of view for two modalities to be functionally equivalent means they have the same input and output dimensionality, and the *same causal structure* as well – i.e. causal functional-flow – see the section on the dynamic structure of causal networks in the appendix. Remember that the purpose of the instantiated dissociated dispositions functional-flow is to signify reality, to signify the modality. To be equivalent the modalities would have to be the same things in reality. Therefore, it is not possible to have functionally equivalent modalities without them leading to the same sensory experiences and signifying the same things in reality. This would place them in overlapping regions of the perceptual space. Hence, the following requirement:

Requirement for Entity Distinguishability. (R-ED)

To be distinguishable in consciousness from the first person perspective, qualia must have different causal structures and in turn be at different topological locations in the signified perception of reality, otherwise they would be perceived as the same thing in the signified reality. Conversely, they would be indistinguishable aspects of reality.

Notice that locations need not be with respect to spatial dimensions, for example with a mouthful of flavours, the flavours tend to be lumped together as originating from the same spatial location. Nevertheless, each flavour has a location in "flavour space" in conjunction with the set of qualia that portray its degree of sweetness, bitterness and sourness etc. Finally, a leap has been taken here from functional-flows to causal structures, the justification for this is discussed in the next chapter in the section on causal accessibility.

7.2.12.1 Structural Basis of Qualia

Following from the above requirement, the relationship between a quale type and its underlying operational structure can be treated in two ways. Either a mechanism with a particular universal structure gives rise to instances of dissociated dispositions leading to qualia of a certain type, or that:

Statement of Quale Structure Dependence. (S-QSD)

A facet of reality will have a particular functional-flow in the signified reality and this will determine the operational structure that underlies a quale.

This implies a quale may always require the same operational structure, depending on context. Therefore, in this sense, a particular structure may be associated with a particular quale. It follows that the openness of the space of possible qualia is only limited by the diversity of functional-flows achievable in an environment. The dependence of qualia on operational structure also implies the qualia space is universal in that the same structure in different systems will produce the same qualia.

7.2.12.2 Phenomenal Space Quale Axes Types

Following from the above statement, two modalities might have a qualitative principal basis with the same dimensionality, e.g. suppose taste has four dimensions (sweet, bitter, sour, saline), not necessarily orthogonal, and colour has four dimensions (RGBY) – for example as in the case of some types of birds. However, the basis axes will have different types (i.e. units). What is more, progression along an axis, or in phenomenal space, produces a dissociated disposition type change along the axis. That is, *sweet* has a set of disposition instances associated with it, while *bitter* has another set.

Statement of Phenomenal Space Quale Axes Types. (S-PSQAT)

The perceptual space type of a perceptual space axis may undergo a progressive change in its type across its domain. This corresponds to functional-flow changes in the perceptual field mechanism.

Notice that there is some common overlap, and that there are comparative dispositions for the sensation of distance between points on an axis. Hence, as a taste moves from bitter to sweet there is a corresponding continuous change in the dissociated disposition basis. It is not just a change in a neuron's firing rate. It is a change in disposition basis. The instantiated dissociated dispositions underlying the signification of the axes will be different for different modalities and across the axis within a modality. What would be the semantic interpretation of an axis' type?

Therefore, there are two reasons, at least, why the modalities appear differently. Firstly, because they have different axis types. Secondly, because they are different in terms of their causal structure (causal functional-flow), rather than their apparent external function. Remember, if they had the same causal structure in the same context, they would be the same aspects of reality, and so correspond to the same thing in the signified reality. That is, they would be environmentally indistinguishable to the system. This is summed up in the following statement:

Statement of the Reason for Modal Functional Independence. (S-RMFI)
Qualia for independent modalities appear qualitatively distinct from a first person perspective since they are signified in the system by operationally distinguishable networks of functional-flows.

8 Summary

A hypothetical system containing a hierarchy of three ontological levels (an implementation level, a dispositional level, and a phenomenal level) was proposed as a working framework on which to refine the nomic and operational nature of semiotic processes. The term "dissociated disposition" was introduced to refer to an abstraction of the mechanism linking the implementation and phenomenal levels.

Three problems for theories of consciousness were introduced and used as a guide for the analysis. These were the spatial-temporal limit problem, the modal functional independence problem, and the interpreter regress problem. The next chapter addresses this third problem. Addressing the first two problems, the nature of semiotic signs was examined from an operational perspective. This, combined with the analysis of the contents of consciousness, suggests that semiotic processes can be implemented as networks of instantiated dissociated dispositions in terms of temporal-representations and functional-flows. With regard to the modal functional independence problem, from an operational perspective, two modalities appear qualitatively distinct since they are signified by operationally distinguishable networks of causal functional-flows.

This chapter focused on the operational basis of semiotic processes. The main result was to suggest that as far as the operational requirements are concerned, the primary requirement for implementing semiotic processes is replicating their causal structure. However, this still leaves many questions unanswered, such as the nature of the semiotic grounding relation – this question is returned to in chapter six. The next task is to examine how this operational basis for semiotic processes might be structured to produce minds.

Chapter 4

Structural Requirements for Generating Mind in a Semiotic System

1 Overview

The previous chapter suggested that the primary operational requirement for implementing a semiotic system is a system that has the right kind of causal structure. This chapter shows how primary-consciousness, a primitive form of semiotic-consciousness as explained in chapter two, arises in the system through the way in which it signifies reality and how this is achieved through the relationship between the ontological levels in the system.

1.1 Logical Development

The next step in exploring the prospects and promise of mild AI is to examine how dispositional processes are structured to produce mind. This requires determining the system framework for generating, at least, qualitative primary-consciousness. The guiding light here is the search for a framework that adequately deals with the interpreter regress problem (introduced in the last chapter).

To address the interpreter regress problem requires showing how consciousness can arise in the hypothetical system without necessitating an internal homunculus. This chapter explores the structural requirements on the systems' operation, in regard to enabling it to use signs, which might be sufficient for generating consciousness without necessitating a homunculus. This is brought together in the most important section of the chapter entitled "Requirement for System Structuring for Causal Accessibility."

2 Constructing the Signified Perception of Reality

In the previous chapter, the importance of the system having a signified perception of reality as a basis for making decisions to aid survival was discussed. This section explores the operational structure, primarily at the dispositional level, of this network of signs within the system.

2.1 The Category Structure

Combining the requirement for the necessary categorisation of signs, the requirement for operational decisional singularisation, and the suggestion for an underlying topology structure, as developed in the last chapter, leads to the grander suggestion that there is a dedicated category structure in the system across which entities are categorised. Noting that consciousness is always with respect to specific signs, a category structure is suggested as a common structure for semiotic processes and that signs have a category property modelled through instances of dissociated dispositions. This section explores this suggestion.

One consequence of the statement on the first person perspective view of consciousness presented in chapter three (see S-FPPPVC) is that in the signified reality, the creature's perspective on reality is signified as though it was in reality, not as reality relative to a distinct Self. Reality appears to have a constancy about it as the creature moves through it. The perception of reality is fixed relative to the centre of the creature's reality perspective. It is useful to remember this point when trying to model the dynamic nature of singularisation, i.e. when new signs and orderings are produced after the perspective changes, but they can change, even quite rapidly.

Consider a visual scene: each visual entity and its dissociated dispositions instances, D_i, have to be spatially fixed, with respect to the horizontal, vertical and depth, through a network of signs that make up the signified reality. That is, there is a D_m , when suitably embedded in the signified reality, that gives rise to the sense of "Next-to" for visual entities. This primitive form of D_m has to cater for the binding of D_i. In the topology structure, with parallelism of experiences, D_m could express local topological relations along with other dispositions for properties such as: boundary, centre, mass, size, type. This implies the following:

1) The instances of dissociated dispositions must be semantically self-contained to some degree since they are localised to a region of the support that may bring about dispositions for different modalities. However, a particular region of the support may have general and specific functionality, e.g. a general meaning attributing mechanism and a specific area that adapts to portray different properties, such as Next-to, Is-bitter.
2) The structure of an instance of a dissociated disposition may change completely, thereby giving rise to a different disposition. Suppose an instance of a dissociated

disposition is represented by the symbols: $D_F(X)$, where D_F describes its causal form and X is an n dimensional vector denoting any impinging dispositions. Then both D_F and X may change in form and dimension, respectively. Notice that any internal structure within D_F is topologically degenerate, e.g. compare this with the tree structure of a neuron where a term may be any function on its dendrites.

3) There is some kind of entity hierarchy on top of the modal perceptual field, which raises questions concerning the structural nature of the instances of the dissociated dispositions producing this hierarchy.

Above it was suggested that the categorisation mechanism involves some kind of underlying fixed topology structure. In light of the requirement for singularisation and what has just been said, this suggests singularisation is supported by a fixed structure that embodies the categorisation hierarchy. This would consist of specific structures for each category, e.g. the Self, the body and external reality. Each category has further structure for entity types and modal integration. The structures would have some degree of similarity corresponding to the nature of the category and the generality of the type structure. In the hypothetical system, this topology structure is called “the category structure”:

Statement on the Necessity for the Category Structure. (S-NCS)

In the hypothetical system, a structure is necessary upon which a categorisation of entities is formed during the system development of the signified perception of reality.

2.2 Building the Signified Perception of Reality in the Category Structure

This section highlights the mechanistic difference between stationary-representations and temporal-representations. How both mechanisms come into play in the category structure for vision is considered. Notice that this mechanistic type difference is the main reason for advancing a distinct dispositional level.

Building the signified perception of reality can be broken down into two tasks. The first involves representing (via stationary representations) the structure of reality directly through the modalities. This is at the perceptual field (implementation) level. For example, in the visual modality this would lead to representations for the primitive shapes, but devoid of colour, depth and compound objects. The second task is performed by the category structure, which constructs a signified perception of reality, in terms of interacting instances of dissociated dispositions, by integrating these modal representations and by adding secondary qualities derived from innate structure, knowledge and beliefs. Figure 4 outlines the steps involved.

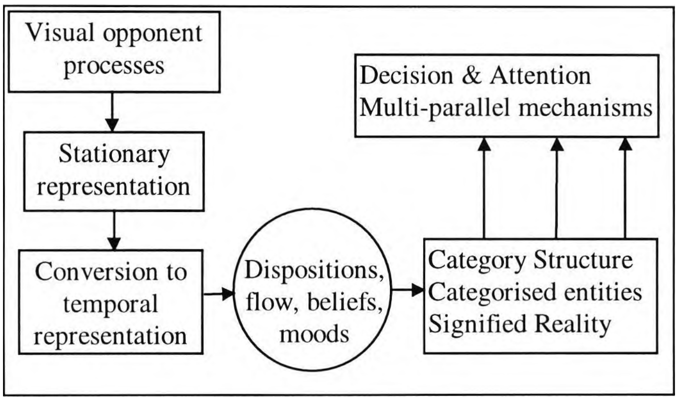


Figure 4. Transduction of stationary-representations to functional-flows.
Emphasising how visual opponent processes are representations that have to be converted to instances of dissociated dispositions, i.e. functional-flows on route to the category structure for the purpose of producing semiotic processes.

Statement on Structure Transduction to Functional-Flows. (S-STFR)

In the hypothetical system, there is a system level change from stationary-representational structures at the implementation level to functional-flows at the dispositional level.

This is a very significant statement and will influence much of what follows. It denotes a transition from explicit to implicit representations positioned at different ontological levels. Consequently, a goal of singularisation is to recombine modalities into a signified perception of reality, wherein entities, now categorised, have dispositional properties added: sound, motion and colour as *knowledge indicators* – this is discussed later in the section on generating sensory experiences. Therefore, in Figure 4, the categorisation is typed not by modality, but by the functional-flow and the utility in reality, all with respect to the creature. What are the types of categories? How are they conveyed to the decision and attention mechanisms? Is colour a distinct category, with particular primaries being subcategories, or a collection of categories?

A categorised entity has a set of categorised properties: e.g. Form (size, location) and Belief (mood - colour, utility - friend, foe or food). This suggests some of the steps involved when signifying two different modalities may be of a similar form. For example, comparing hearing and vision, sounds can often be thought of as emanating from a non-specific location. They can be detached from their source. Similarly, physical spectral colours are detached from their source, but then reattached, this time as true psychological colours (i.e. mood indicators), via locational qualia as part of singularisation binding, or just as a "surface filled with mood indicating colour" belief about the object – see the section below on generating sensory experiences for a definition of psychological and spectral colours.

2.3 Requirement for Environmental Interaction

Building the signified perception of reality is an ongoing, continuous activity – at least from the first person perspective of the phenomenal level. This leads to some comparatively straightforward, although strict conditions on the dynamics of the signification process:

Requirement for Environmental Interaction. (R-EI)

In order for the hypothetical system to integrate with its environment, the rate at which reality is signified, and its subsequent consciousness experience, has to match the actions of the individual and perceived events in the environment.

One could imagine a creature that took years to respond to events in its world. Environmental events that took place at a faster pace would be a blur, and those that took place at a slower pace might not even be noticed. What this requirement means is that the consciously experienced signified reality must temporally track anything that is perceived. Hence, the rate of modelling is largely determined by the performance of the individual. These practical constraints on the signification process will raise pragmatic issues as to how well the creature can get around in the world via its signified reality.

2.4 The Building Blocks of the Signified Reality

Without loss of generality, the instantaneous functional-flow of an instance of a dissociated disposition can be modelled as the composition of a number of instances of primitive dissociated dispositions. Table 2 presents a classification of a possible set of primitives. Since these are graded dispositions a multi-valued logic is appropriate, rather than a binary logic. For example, a quantised version of the nondenumerably infinite system of J.Lukasiewicz [Resc69] – see the appendix. However, since the functional-flow of an instance of a dissociated disposition evolves, i.e. it is a dynamic functional-flow, these primitives need to be parametrically specified – see the next chapter for details.

Class One	AND-$\{D_1 \wedge D_2 \wedge \ldots \wedge D_n\}$ OR-$\{D_1 \vee D_2 \vee \ldots \vee D_n\}$ NOT-{D}
Class Two	ORDER-{X,Y,Z} on a set {D}, Z is the type. RELATED {A related to B}, a subclass of ORDER?
Class Three	BOUND-{… some set of dissociated disposition instances} IDENTITY-{These disposition instances are an entity}

Table 2. An initial basis for the flow primitives of dissociated dispositions.
Notice that this does not include primitives for structural constraint mechanisms, such as disposition normalisation.

2.4.1 Class One: AND, OR, NOT

In a similar fashion to logic circuits, these primitives would be used to construct a particular causal structure.

The AND dissociated disposition has two possible operational roles:

1) As a causal connective. For example, D_X depends on $D_Y \wedge D_Z$. The form of the dependence is then some function: $D_X(t+1) = f(S, f_A(D_Y,D_Z))$, where S is structure.
2) To logically AND instances of other dispositions together, either pseudo truth wise, or as an implication.

Notice that AND is partly synonymous with BOUND. It is just that the consequent differs:

1) For BOUND the consequent is an atomic or compound sign.
2) For AND the consequent is not a sign. In other words it is a reducible entity, or has no causal integrity, i.e. there is not a direct cyclic causal path between its consequent and antecedent – meaning other dispositional processes may influence its consequent. The consequent dissociated disposition instance takes part in fixing the causal integrity of compound and atomic signs.

The most important operational role of OR is in specifying the causal dependence of instances of dissociated dispositions. What are the logical causal dependencies? Does causal dependency for a quale imply a direct cyclic causal path i.e. that there is a dominant cyclic causal path underlying the quale process? What about nodes which source or sink causal influences?

While a NOT operator is certainly at play in the support of dissociated dispositions, and in guiding causality, its purpose at a higher level is more tenuous. The signified perception of reality has to portray what is the case, rather than what is not.

2.4.2 Class Two: ORDER, RELATED

This class would naturally be based on the first, although some kind of dynamic parametric coding would be required in order to capture the transitory nature of dissociated dispositions.

The ORDER dissociated disposition is discussed in the following sections.

The RELATED dissociated disposition has two possible operational roles:

1) An experienced relation, e.g. A is next-to B, in some context, A is associated-to B. Although this could just be a variant on an unordered type.
2) A bound relation, e.g. properties related to an entity bound into an atomic sign.

RELATED appears to be used in the category structure to integrate modalities, e.g. one entity next to another. Although, it could be the case that all associations are some kind of typed ordering. Relations could just be a combination of $D_{(UN)ORDER}$ and D_{BOUND}. Notice the difference between NOT D_{ORDER} and $D_{UNORDERED}$ – see the discussion of ORDER below.

2.4.3 Class Three: BOUND, IDENTITY

The goal of the BOUND dissociated disposition is to bind entities, via a dispositional mechanism in conjunction with the structuring mechanism, discussed later, into an atomic or compound sign, i.e. *an experienced whole*, for the purpose of reducing, or consolidating information. Suppose D_B binds into an atomic sign a set of dissociated disposition instances that are modelling, though signification, properties of the environment: D_B-$\{D_1, \ldots, D_i, \ldots, D_n\}$. The sign is characterised by the operations supported on it, and the following would also hold:

1) Any singularisation operation on the D_i treats them as a whole.
2) There are no operations that can directly individuate them to other atomic signs. Notice that the D_i are discriminated from their kind, but not as atomic signs. This is enforced (accomplished) through the nature of the dissociated disposition signification of reality.
3) A bound atomic sign becomes a virtual primitive (abstract) entity to subsequent layers.

The goal of the IDENTITY dissociated disposition is to produce in the signified reality (in conjunction with the structuring processes – see below) the experience of something being an entity and to identify an entity by its purpose in the signified reality. This is required at the first level of singularisation, i.e. qualia in the primary field are experienced as typed atomic signs only, not as compound signs. Therefore,

1) Entities that have an identity have a set of dispositions which are context, domain and modally specific, e.g. constrained by reality.
2) There will be causal flows specific to a category layer in the hierarchy: entity connectives and orderings. These flows and constraints for category layers can be compared to those in other layers.

Thus, the D_I's will be relatively complex compared to the other primitive dispositions. It is necessary to separate the aspects of a D_I into those unique to an entity, and those general to all. However, the IDENTITY disposition could be a binding disposition, or atomic sign acting in an entity indicator role.

Related to the quale of identity and objectness is the quale for completeness indicating that an entity does not directly causally influence dissociated disposition instances of other spatiotemporally separated entities.

2.5 Nature of Dissociated Dispositions for Signifying Ordering's

The previous classification of dissociated disposition primitives distinguished an instance of a dissociated disposition, D_{ORDER}, for signifying orderings amongst entities. This primitive turns out to be analytically very useful for refining the system structure for generating consciousness. The following subsections examine the nature of the D_{ORDER} dissociated disposition for arbitrary order types.

2.5.1 The Need for an Experience of Order

With regard to order derived properties for visual qualities, such as proximity, space and colour (also called secondary properties), some mechanism has to determine the relative property states for the scene and then assign them to the visual objects. Subsequent derived properties may be based on these, such as the closest object, and the brightest object. Via an attention mechanism (briefly discussed later), focus can be directed on the winner, or next, or lowest object etc. Consequently, the winner is able to play a larger role in the dispositional feedback process – see below.

Now consider two objects such that their relative visual depth slowly changes. What structure represents, in general, their relative depth? They have different X,Y orders, but how is the depth hierarchy structured? Without binocular vision and knowledge cues, it becomes difficult to judge the depth of objects. However, the same objects are still perceived. That is, depth is added as a property (functional-flow) in a post primary field stage. Consequently, the binding set has changed for the objects.

2.5.2 Functional-Flow Typed Orders

This section suggests that in the system an ordering primitive consists of a conventional order determining mechanism and a type indicating mechanism.

Take depth, primitively it could be represented as a hierarchy of relations, such as "Before-X,Y", with the creature's position as reference. However, the dispositional functional-flow of "Before-X,Y" would still have to be unpacked. By symmetry, what would be the form of the dissociated dispositions and associated system structure for relations such as "Next-to-X,Y", and "Part-of-X,Y"? How would they get their particular functional-flow? Contrast this with giving meaning to "Tone-Higher-X,Y". Is it acceptable to even consider "Before-X,Y" as a valid dissociated disposition or primitive? Is it really the correct perspective for vision? It is used to explain depth perception and is on a par with "Next-to-X,Y", "Solid-X", "Colour-X" etc. Primitively these are, in an evolution sense, set-theoretic relational orderings. This leads to the following statement.

Statement on Functional-Flow Typed Orders. (S-FRTO)

In the hypothetical system, relational orderings are implemented by instances of dissociated dispositions where the signification of each type of ordering is modelled as a specific form of functional-flow enforced by a set of dissociated disposition instances.

Suppose that the order relation is parameterised with respect to the type of order, such as "Order-X,Y,Z", where Z is the type of order, and "Before-X,Y" ≡ "Order-X, Y, Before". This could be a fixed structure wherein entities are initially ordered via a syntactic mechanism, so that the highest or nearest entity is put first in the hierarchy. Alternatively, entities could be tagged with a rank or magnitude attribute. Consequently, the mechanism for depth representation, and in general ordering, could be handled in purely syntactic terms.

This still leaves explaining the experiential attribution of "this structure signifies type of order Z". This is not a semantics question, but rather the experience of depth as opposed to the experience of some other order type, e.g. colour. Hence, what is the form of the dissociated disposition functional-flow that leads to the experience for the typed order relation between, say, colours? The order relation is not a binary relationship, but is it necessarily quartic: RGBY, or a sextuplet: RGBYBW, or higher?

2.5.3 Representing General D_{ORDER}'s

Assume that the dispositional level is presented with an ordered array of properties, plus their respective contexts, what is the mechanism for general D_{ORDER}'s? For example, D_{ORDER} is then used to produce the experience of an ordering, e.g. "this is the {brightest, biggest, closest} entity or set of entities", or "these objects are next-to each other." This latter example is a special kind of typed ordering. Notice how the ordering may apply to sets giving joint winners, which indicates a parallel process. Also notice how each of these typed orders relies on the capabilities of the corresponding dispositional flow, e.g. the flow for biggest would access the flows of instantiated dissociated dispositions for spatial extension etc.

Consequently, the mechanistic requirements for general D_{ORDER}'s involves a logical (relational) mechanism for expressing orders as dispositional flows. In addition, this must be distinguished from the flows of an instance of a dissociated disposition that capture the experienced type of the order. Consider a test case. Imagine a landscape scene, defocus so as to comprehend the whole scene as one. Focus on the experienced depth of all objects in the scene, and the spatial extension of objects. It is as though a depth ordering is sensed between all objects – some may be at the same depth (heterogeneous ordering). In any case, a spatial ordering is still experienced across the scene. It seems as though there is a parallel set of experienced senses, typically arranged in a radial fashion around the focus of attention.

2.5.4 Operational Modelling Issues for D_{ORDER}

Mind design driven by evolution and survival of the fittest would favour models that allowed the creature to solve problems. Regardless of the way in which this is achieved, if the model adequately enables the system to solve the problems it was intended to address, it might be said to be adequate, accurate, or even true at that level, to that extent or that degree of specificity. The following discusses some issues related to representing orderings, a central issue being the manner by which entities in the ordering are referred to. While this appears as an operational problem to mind designers at present, nature may have bypassed this issue altogether by some scheme that was found to be adequate.

Firstly, if an ordered list is proposed for representing the ordered elements referred to by D_{ORDER}, then the referent needs to be referred back to. If an n-arity relation tree is used, the tree needs to be composed quickly. What if the experience of depth is demand driven? In that, the 'depth' experience is only experienced when attended to, i.e. requested? Even so, a casual glance at a scene gives the impression of a coherent depth, i.e. lots of parallel depth qualia.

Suppose some mechanism is devised for representing depth order, e.g. from neural nets such as Winner-Take-All (WTA), or On-Centre-Off-Surround (OCOS) nets – see Stephen Grossberg [Gss75b]. If based on the relative magnitude of a scalar, a comparison across objects of the OCOS needs to be conducted. However, recall that the system builds a 'virtual' signified perception of the world with the creature's perspective relative to it. Hence, orders could be predetermined and represented as constant, with the signified perception of the creature's perspective moving in this larger model. Consequently:

Requirement for the Topological Ordering of Entities. (R-TOE)
In analogy to the requirement for the necessary categorisation of signs, only entities that have been topologically ordered are consciously experienced.

In effect, this is suggesting the entities have to be *topologically located* (via a network of interacting signs) in the signified perception of reality. When updating the signified reality there will also be tight timing constraints to achieve this in real time. This is a pragmatic issue with respect to how well the mental model enables the creature to get around in the world.

Consider the relative position of objects in space represented as Cartesian X, Y, Z displacements. In this physical space, two objects are next to each other, only if they are. Generalising to other modalities, this suggests entities could be organised so that they have a set of context relations, such as "Next-to: A, B". However, this would only work for their topologically local environment, and there may be more distant relations, such as: a) distant in a metric sense, b) distant in a physical sense. It might be that two objects topologically distant in reality are mapped to a structure, in which they are physically distant, so that a series of causal relations would be required to link them across the structure. For these more distant relations a more complex mechanism would be required, such as a simple inference mechanism tied in with attention, e.g. it might obey normal transitivity: "If $a < b \wedge b < c$ then $a < c$." Compare this to Minsky and Papert's [Mins69] argument concerning the connectivity of objects.

Therefore, remembering that the system has to signify all aspects of reality, even the notion of space and Next-to, how are the entities of a modality and their relations signified in a functional-flow purpose preserving way? Consider a collection of entities. How are referents in relations referred to? The signification must be causally integrated and dynamically reconfigurable. The instances of the dissociated dispositions for the entities have a physical implementation location in the system, although the disposition instance set may be distributed, as is the effect of the flow of an instantiated dissociated disposition. If the referent problem is resolved by the proximity of instances of dissociated dispositions, then a topologic map would be expected in the system. This is not obviously so for vision, although it could be a higher order topologic map (of four, five and many more dimensions) projected onto two or three dimensions via Peano type curves, e.g. the body is topologically mapped.

Consequently, the physical proximity of instances of dissociated dispositions does not solve the referent problem. Instead, chained interactions between the underlying causal functional-flows of the dissociated disposition instances is relied upon. However, such a mechanism may be operationally limited to physically short distances (due to physical connectivity constraints) – in terms of neural columns – or whatever the base unit is. Due to the highly concurrent nature of the reality being signified and its interdependencies, evolution would be expected to favour spatially local solutions to the referent problem, but not necessarily topology preserving.

2.5.5 An Operational Model for D_{ORDER}

This section suggests there are two distinct stages involved when operationally modelling D_{ORDER}.

Suppose D_{ORDER} is represented by a hierarchy of binary relations: $a < b$ etc. How, for a general scene with a process topology structure, would the hierarchy be operationally distributed (for a certain sane, rational agent)? Consider the case where the most near and distant entities change position in the field – such as a rapidly approaching ambulance. How does the hierarchy operationally model this and the change? See Figure 5.

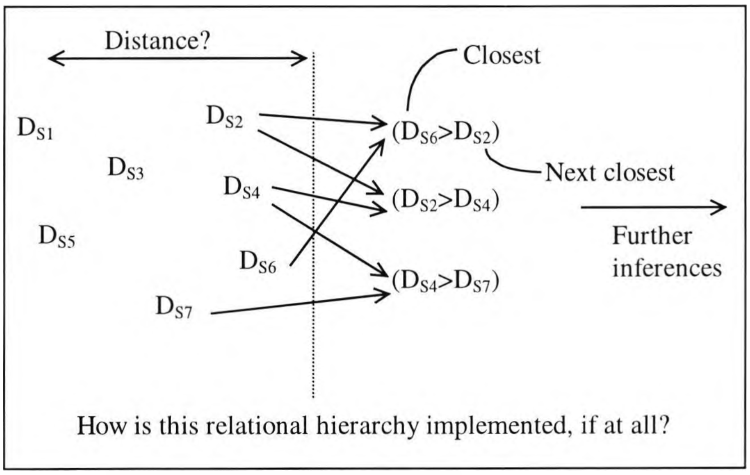


Figure 5. Ordering Signs via binary ordering relations.
Here D_{Sx} denotes a dispositional process giving rise to a sign.

The relation used to order objects, e.g. Greater-than, Equal-to, Less-than, will be flow dependent. A WTA neural net could easily pick out distinct entities. However, the creature may be able to relationally compare any two entities in the field, or at least neighbours, which suggests a self-organising structure. What of the structure for higher arity relations, such as tertiary relations?

Consequently, a low-level mechanism must already have deduced the order and then represented this in some way, e.g. by configuring the hierarchy, or has it? What is this process for determining order? Remembering that the end result has to be implemented via a combination of structure and dispositional functional-flows, this suggests the following statement:

Statement on Stages in Implementing Dispositional Ordering. (S-SIDO)

There are two aspects to implementing instances of dissociated dispositional ordering:

1) *The mechanism by which Order is determined, i.e. an implementation level mechanism.*
2) *The mechanism by which Order is portrayed, i.e. functional-flow basis of quale.*

The latter stage also has its own representation issues: how is it modelled at the dispositional level, considering that it will influence the subsequent decision mechanisms?

3 A Theory for the Domain of Dissociated Dispositions

From the above it follows that D_{ORDER} implies a prior mechanism to determine order, which is then presented to instances of dissociated dispositions. Consequently, the support mechanism for instances of dissociated dispositions, and their transformation by the support, needs to be separated from the evolution of the instances themselves. This implies the next statement:

Statement on Triplicity of System Levels. (S-TSL)

In the hypothetical system, the following two processes can be separated and treated as distinct:

1) *The mechanism of presentation of the physical reality to instances of dissociated dispositions.*
2) *The calculus of the dissociated disposition functional-flows giving rise to qualia.*

An important implication of this is that the nature of the mechanism does not matter so long as it supports the *causal* nature of instantiated dissociated dispositions. That is, the calculus of dissociated dispositions is concerned with the causal interaction and structure of functional-flows. A summary of the evidence for this gathered from the previous sections is as follows:

Firstly, functional-flows do not have to be ground in an absolute, universal set of primitives. Any mechanism that enables the instantiation of a dissociated disposition with a specific causal functional-flow will do – see the section on redressing the spatial-temporal limit problem. On the other hand, it makes determining the set of possible dissociated dispositions at the dispositional level more difficult. For example, compare this with the analogous case of trying to set bounds on the possible logic functions constructable from arbitrary combinations of logic gates. Studying biological neural structures with the aid of theoretical models would help determine the dissociated dispositions used in nervous systems.

Secondly, the system requires that in order for properties of reality to be signified and subsequently experienced, they must be presented in an orchestrated manner to dissociated disposition instances by some implementation level mechanism – see the requirement for reconstruction from first principles.

Thirdly, there is a mechanistic type change from conventional representations in the transduction to functional-flows – see the statement on structure transduction to functional-flows.

Fourthly, the thesis of mechanistic chauvinism has yet to prove that mind is necessarily limited to biologically based creatures. Otherwise, by definition, mild AI would be impossible.

Consequently, it can be concluded that the domain of dissociated dispositions and the logical structures instigated by instances of the dispositions on this domain form a distinct ontological level. A formal derivation of this would allow it to be raised to a theorem of the form:

Dissociated Disposition Level Theorem. (T-DDL)

Instances of dissociated dispositions exist at a distinct ontological level of functional-flows and temporal-representations based on a lower causally structured ontological level.

The lower level is necessarily causally structural, meaning it acts as a medium which determines the domain of dispositional functional-flows open to the evolution of the instantiated dissociated dispositions.

The theorem suggests that the mechanism by which structures and representations are achieved is largely independent of the dispositional level except that it must satisfy the necessary inter-level contingencies, e.g. time constraints and reality integrity. That is, the structures and representations must be a faithful portrayal of given aspects of reality with a high degree of accuracy and correlation – see the statement on perspective view of consciousness. What this means:

1) The dispositional level can be treated largely independently of the particular mechanism by which aspects of reality are presented to the level. Thus, semantically grounding the presentation mechanisms is not an issue. Semantics is a phenomenal space level concept achieved through the functional-flows of instantiated dissociated disposition networks – this point is returned to in chapter six.
2) The dispositional level just requires that any presentation from the implementation level is 'faithful', i.e. under the same context an event in the real world causes the same presentation to be evoked, ignoring effects of learning and so on for now.
3) The presentation mechanism need not be biological.
4) The presentation mechanism is a preconscious task.

By extension, all stationary-representational aspects of reality and thought can be placed at the presentation/implementation level. Qualitative consciousness is then a higher level arising within the system as an integral operational process for dealing with temporal-representations. What are the implications of this for primary consciousness? Does the presence of consciousness necessarily imply a dispositional level?

4 The Causal Structure for Generating Mind

This is the most important section in this chapter. The purpose of this section is to suggest how the system might be structured such that qualitative consciousness arises. This culminates in the first half of an explanation for how semiotic processes might arise from dispositional processes, presented in the subsection on the requirement for system structuring for causal accessibility; chapter six presents the second half.

4.1 Requirement for Observer and System Level Relative Properties

This requirement is a reminder that the meaning and significance of something is relative to the system level and perspective:

Requirement for Observer and System Level Relative Properties. (R-OSLRP)

The relative perspective of the observer, as theoretician or the mind within the system, whether from a third or first person perspective, and whether from an inter or intra ontological level, will determine what properties of the system they are able to observe and how those properties are manifest to them.

There is a difference between contingent and exclusive properties of levels. In particular, a property of a higher level may be contingent on a lower level mechanism. However, the property may be an exclusive property of the higher level. In addition, what an observer labels as a property of one level, may only be recognisable as such either from that level, a higher level, any higher level, or exclusively to a particular level. This is summed up by saying:

Statement on Extension of Qualitative Consciousness. (S-EQC)

Experiences (i.e. qualitative consciousness) are an exclusive property of the phenomenal level.

This implies that a semiotic process, such as one thought bringing about another thought, should be characterised in terms of the structure of the instances of dissociated dispositions that gives rise to the experience of this process (a dispositional construction of this personal aspect of reality) and the dispositional influence of the signified reality within the system. The structure between instances of dissociated dispositions for achieving this is discussed shortly.

Secondly, to recap, dealing with the question of perspective, recall that representing implies a perspective is taken by the subject toward that represented. The system's sense of perspective can only be experienced consciously by it through signs as qualia in the perceptual space. As suggested in the statement on the first person perspective view of consciousness in the last chapter, reality is made to appear as though the qualia in the perceptual space are from a representational perspective in relation to the creature. What is important here is not that it has a certain kind of qualia that are experienced like so, but the apparent relationship it believes it experiences between the sense of reality and this sense of perspective. Everything it is conscious of has to be revealed to it through qualia (and therefore signs) in the perceptual space.

Now, dealing with the change of levels, there is the apparent first person perspective of the conscious creature, and the third person perspective of the mechanism. These perspectives are from the phenomenal and implementational levels, respectively. The contemporary stance is that either qualitative consciousness is a representation for the mechanism (or system, or entity in system, i.e. has the mechanism as subject), or else it risks becoming epiphenomenal. Not much of a choice, until one realises what the change in levels also allows. It was explained above that the dispositional level does not treat representation in the contemporary fashion. There is no subject or object, just an interacting network of instances of dissociated dispositions. From the external perspective of the theorist, all that can be done is to point to an area and say that the structure and operation of this area corresponds to a system (wherein the syntax-content problem has been solved) with these kind of representational qualities – at the same time acknowledging the fallibility of the system. However, to rekindle some of the connectionist arguments, clearly the two perspectives are operationally different. The realisation is this:

Statement on Perspective and Relative Laws at each Level. (S-PRLL)

What appears (from a third person perspective) to be a representation for one level is not necessarily (and almost certainly is not) a representation, or the same representation, to (an observer at) another level.

As was explained, there can even be a different notion of representation at each level. This point will be returned to when causal accessibility is discussed shortly and again in chapter six.

4.2 Causal Significance of Instantiated Dissociated Dispositions

This section discusses the operational importance of the causal influence of an instantiated dissociated disposition on other instances of dissociated dispositions. That is, that they are functional-flows embodied as causal-flows, a particular type of functional-flow. In addition, they have a particular causal direction and are causally irreducible.

The support medium and the current modal activity determine an instantiated dissociated disposition's causal influence and consequently its causal structure. Mechanisms that contribute to its causal structure are constraints imposed by the support arising from real world representational requirements. For example, in monochromatic vision there is a consistency constraint across the instantiated dissociated disposition set for the visual field. Typically, a constraint would be something like a normalisation of a disposition's influence upon a set of recipient dissociated disposition instances.

An instantiated dissociated disposition has two principal means of influence, one is on the instances of its immediate neighbouring dispositions, and the other is its overall impact on the system. The first kind of influence may cause an immediate response from the neighbours (time scale less than 10ms), whereas the repercussions of the system impact could be an order of magnitude longer. Its causal role is identified with the first means of influence. This acts in the direction of causation, which is the same as the conventional direction of presentation. In contrast, the sensed (signified) direction of perception is opposite to this. From the analysis of the binding problem it was explained that the system signifies the creature in reality as though it was viewing reality – see the statement on the first person perspective view of consciousness (S-FPPVC). Consequently, this leads to the statement:

Statement on Direction of Dissociated Disposition Flows. (S-DDDF)

The causal direction of influence of an instance of a dissociated disposition's flow is the same as the conventional direction of presentation but opposite to the signified direction of perception.

What makes this statement interesting is what it implies about determinism. The important point being that determinism is normally discussed relative to a Self, which is normally positioned as a homunculus in the system. However, in the hypothetical system considered here, the Self is a signified set of sensed beliefs arising from a model produced by a distributed network of instances of dissociated dispositions. That is, there is a system level change in perspective – this is discussed in the next section.

The causal determinacy at play is sketched in Figure 6. This shows pseudo constant flows (e.g. stable causal patterns) with respect to occurrent activity in the system feedback loops. These are important constituents of the processes leading to atomic signs and D_B's for compound signs.

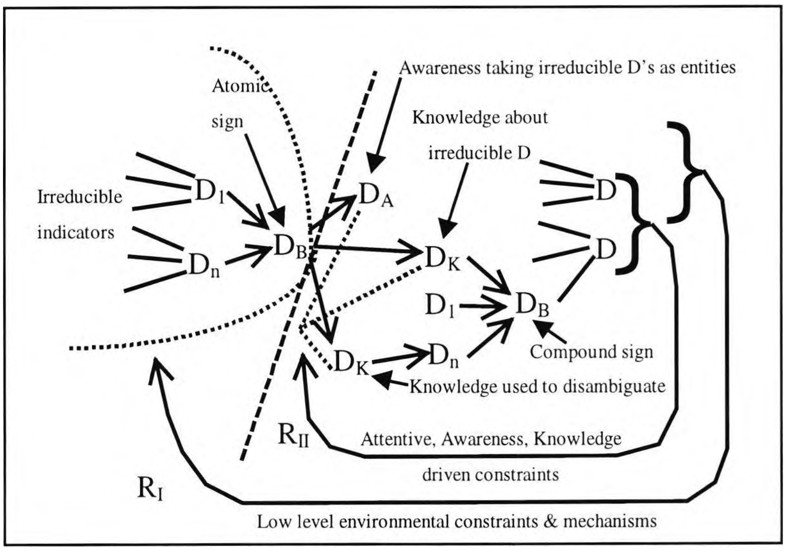


Figure 6. Overview of causal pathways in the system.
This shows the causal pathways set up via interaction of instances of dissociated dispositions in the system. Notice the two levels of recurrency, R_I and R_{II}.

Notice that there are at least two principle levels of recurrency, i.e. in the main system loop at the dispositional level: R_I and R_{II}. It could be postulated that this is a consequence of the causal constancy requirement. The next section and chapter six discuss the structure and dispositional relations within this loop in more detail.

Since an instantiated dissociated disposition is part of consciousness it is irreducible, i.e. it's a dependent part of the signified reality, as all instantiated dissociated dispositions are. This leads to the following statement:

Statement on Instantiated Dissociated Disposition Causal Irreducibility. (S-IDDCI)

To be irreducible means an instantiated dissociated disposition's causal significance (e.g. its meaning) cannot be reduced without impairing the quality of consciousness.

Whether the experience dependent on a missing dissociated disposition instance will be noticed is another matter. This depends on two factors: compensation for the missing instance by others, and that there are no beliefs for absent qualia, i.e. a quale is only experienced and noticed when it is part of consciousness.

Irreducibility implies that, in any context for the instantiated dissociated disposition, it has at least a non-linear, or binary linear flow: $D \equiv f(x_1,x_2,\ldots)$, where f is a non-linear flow function with an arity of at least two. Or, $D \equiv f(a, b)$, where f is linear, but a and b are dissociated dispositions. Causal irreducibility implies dissociated dispositions are at least two-arity, and therefore in general, if something is conscious then its prior state cannot be determined exactly given only its current state. That is, there may be more than one causal pathway to the current state.

Hence, an instantiated dissociated disposition corresponds to a dynamic causal-flow (a type of functional-flow, in the manner defined in the appendix) identified by its principal flow, e.g. an eigenvector:

Statement on Nature of Instances of Dissociated Dispositions. (S-NIDD)

An instantiated dissociated disposition corresponds to a dynamic causal-flow characterised by its principal flow.

This statement will be refined in the next chapter, which looks at the implementation details of instantiated dissociated dispositions.

4.3 Requirement for System Structuring for Causal Accessibility

This is the most important subsection in this chapter. The purpose of this section is to suggest how the system might be structured such that qualitative consciousness arises. The first half of an explanation is presented for how semiotic processes might arise from dispositional processes, chapter six presents the second half.

To begin with, consider the semiotic "conception of consciousness according to which a system is *conscious* (relative to signs of specific kinds) when it (a) possesses the ability to use signs of that kind and (b) is not incapacitated from exercising that ability" – see Fetzer [Fetz98a, p384]. Thus, the system must be structured such that it is able to use signs. To expand on this involves explaining how the system can use signs (i.e. by what mechanisms and form) when semiotic processes are based on instances of dissociated dispositions. This returns the discussion to the interpreter regress problem and entails determining who or what the sign user is, where the interpreter and interpretant fit in, and how they relate to the signs. An important part in addressing these points will be to suggest there is no "third person" interpreter, but mere dispositional processes from start to finish. In place of the interpreter is a dispositionally driven attentional mechanism in conjunction with a presentationally oriented signified perception of reality.

Inherent in all dispositions is their potential to causally bring about some effect. The structure of the causal pathways between instantiated dissociated dispositions turns out to be the primary concern in constructing a mind according to the dispositional approach developed here. A refinement on this is the causal influence an instance of a dissociated disposition has within the system in terms of its local neighbours and on a more global scale. This leads to the main system-structuring requirement that fixes signs and closes the phenomenal space (in an emergent sense):

Requirement for System Structuring for Causal Accessibility. (R-SSCA)

Causal accessibility refers to the potential existence of mutual causal connections between signs in the system necessary for topologically situating them (through signification) in the perceived reality.

Three operational implications follow from this statement with regard to signs, in contrast to their Peircian phenomena. Firstly, if a sign arises in the system, then it has an implied potential purpose, i.e. it has a causal function to bring about some effect, even if this is simply to register an external state of affairs. Secondly, the sign must have some potential to influence the attention mechanism through its neighbours. Thirdly, the sign must be situated (dispositionally) in the signified reality with respect to the attention mechanism.

Casting this in Peircian terms involves explaining what the dispositional equivalent of the interpreter would be. Recall that the statement on the first person perspective view of consciousness stated that entities are perceived as though they occur as presentations in relation to the mind of the creature. However, this doesn't necessarily mean that the system has to represent, through signification, the subject presented to, such as a Self. It simply requires that the entity has "presentational" properties that give it presentational dispositions, such as relations to other entities and properties of the creature. Recalling the statement on the direction of dissociated disposition flows, the experienced causal direction of attention is as though an attendant (i.e. interpreter) is presented with entities that it selectively attends. However, the direction of causation is opposite to this, with the entities dispositionally influencing dispositional attentional mechanisms. These, according to the attentional criterion embodied in the dispositional tendencies, lead to the generation of experience enhanced instances of secondary dissociated dispositions that add more detailed knowledge to the perceived entities being focused on. The attention mechanism is discussed further later – see Figure 8.

In effect, the interpreter, as the sign user, has been replaced by a dispositionally (non-presentational) driven attention mechanism whereupon entities in reality are portrayed, through signification, as though they were presentational things. That is, signs have a presentational property expressed as dispositional tendencies. This naturally leads to treating Peirce's 'Interpretant' as the dispositional tendencies of signs, i.e. the nature of the influence of the underlying instantiated dissociated dispositions on recipient semiotic processes. This is in accord with the conclusions reached by Fetzer – see Fetzer [Fetz98, p383].

The causal accessibility requirement will now be used to explain what makes something primary-conscious. The approach is broadly as follows: signs, and therefore qualia, are part of a categorised entity, the entity is signified via a network of instantiated dissociated dispositions, and the entity is topologically located in the perceptual space. These points will now be unpacked. The points made in the following boxed paragraphs are particularly important. Recall that qualia are the phenomenal aspect of signs, which are integrated causal functional-flow encodings of properties of reality.

To start with, a brief overview of the mechanism is presented. By the requirement for the necessary categorisation of signs, all qualia as properties of signs belong to an entity. For the sake of explanation, assume that there is a singularisation mechanism and a categorisation structure into which entities are situated. An entity is a collection of qualia and therefore signs, which are modelled by a number of entwined dissociated disposition instances. Hence, an entity corresponds to a network of instantiated dissociated dispositions situated in the category structure. Integrated with the category structure is the fixing-mechanism. This mechanism topologically locates and encodes (i.e. through signification) entities via relational qualia in the perceptual space. There are additional mechanisms that will be mentioned in passing, such as the pre-mechanisms which, among other things, process and integrate sensory stimuli, and post-mechanisms, such as decision and attention mechanisms. Now, to translate what this means.

In the following the three ontological levels as distinguished in the hypothetical system come into play: 1) the implementation level, 2) the dispositional level, and 3) the phenomenal level. By the requirement for observer and system level relative properties, the experience of consciousness only means something at the phenomenal level. Now consider the following two points: A) from the statement on the first person perspective view of consciousness, there is an innate sense, i.e. qualia, of the perspective of reality in relation to which things are conscious, cf. the contemporary view of representation, e.g. "*I* am conscious of *X*", and B) from the requirement for the topological ordering of entities, signs whose phenomenal aspects are in consciousness are located with reference to a specific aspect of the signified perception of reality, e.g. the cat is on the mat. Hence, a fixing-mechanism generates perspective *orienting* (point A) and reality *locating* (point B) qualia for an entity.

Starting at the phenomenal level there are conscious entities, where the entities are experienced with reference to the signified perception of reality and according to a representational perspective view because they are situated by the fixing-mechanism, i.e. the contemporary view of presentation is qualitatively (i.e. in a quale fashion) enforced on the entity. Only categorised entities get fixed by the fixing-mechanism, i.e. entities situated in the category structure. Whereas potential qualia, which are not part of the category structure, are not fixed, so they do not enter the main signified perception of reality in the system, i.e. the perceptual space, and therefore are not primary-conscious to anything nor are they qualia. That is, to be primary-conscious at the phenomenal level requires the entity to be fixed in the perceptual space signified perception of reality by the fixing-mechanism. To be conscious the entity has to be topologically located in the signified reality by the fixing-mechanism. In other words, to be conscious the entity must have a *functional-flow* and *topological location* in the signified reality, remembering that topological location is also fixed (i.e. signified) by qualia through the functional-flow of their corresponding sign. This rules out an arbitrary mechanism giving rise to qualia. In any case, it is the mind at the perceptual space level that is conscious, not the mechanism or instances of dissociated dispositions, but the mind signified by the causal structure created by the instantiated dissociated dispositions. Only this mind so produced is implicitly able to use signs.

The important point here, besides taking the correct perspective and level, is that to be primary-conscious involves the system generating the delusion that an entity has a presentational orientation within the signified reality and that it has a topological location in this network of signs that is causally accessible to the attentional mechanisms. Spurious qualia produced by any old mechanism cannot arise because they would not be topologically located in a viable signified perception of reality: there would be no signified mind with the power to use the signs and to experience consciousness of them (through further semiotic processes)!

Statement on the Reason for the Consciousness of Qualia. (S-RCQ)

A quale is conscious because its instantiated dissociated disposition support is ***causally*** *and topologically* ***located*** *in the perceptual space and it is signified in a presentational* ***oriented*** *manner.*

See above for the significance of the emphasised terms: causally, located and oriented.

Finally, a few side issues. Firstly, the fixing-mechanism effectively closes the perceptual space, it determines its dimensionality, because only entities that are fixed become accessible to consciousness. Secondly, primary consciousness is based largely on a feed-forward mechanism in theory, although for reasons of resource efficiency feedback is useful. Lastly, feedback is required for secondary consciousness. This is partly resource determined but also for interactive cognitive mechanisms such as shifting attention and thinking, e.g. decisions based on a thought fed back to determine the next thought. Notice that the feedback need only happen from the decision mechanism back to pre-processes and then into the singularisation mechanism. However, it makes sense for decisions to be based on results of the singularisation mechanism. These mechanisms are discussed briefly later.

5 Generating Sensory Experiences

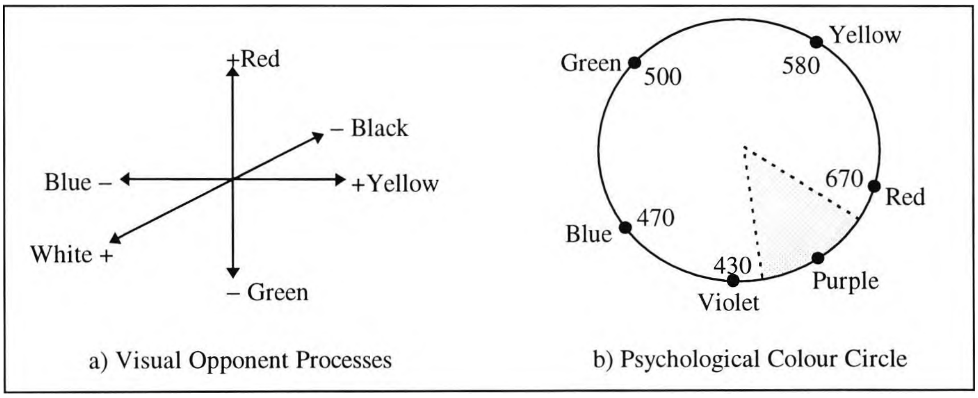
This section applies the implications of the dispositional level perspective on the interpreter regress problem to definitions of qualia and sensory experiences. It provides an explanation for the production of sensory experiences in terms of the requirements developed above. It starts by contrasting the traditional and dispositional level perspectives. For the purposes of illustration, an explanation is given for colour sensory experiences. The explanation can be readily extended to other sensory modalities.

According to the dispositional perspective, entities and their properties experienced in the perceptual space arise from causal dynamic functional-flows (changing patterns of influence) rather than stationary representations and functional-mappings. As such, they evolve and interact in time. For example, colour becomes a dynamic property like sound, even though most colours appear constant. Just as a tone can fade, so too, can a colour fade. A constant colour is analogous to a constant pitched tone. Notice how the original physical properties, light waves and sound waves, are dynamic and continuously variable.

Hence, rather than asking, why is red ‘red’, the question should be posed as, what network of instantiated dissociated dispositions (i.e. causal functional-flows) make red ‘red’? One approach to investigate is to contrast the functional-flows of the psychological colours. For example, consider a black and white world in these terms. Suppose experiences lacked brightness, white would be experienced as light rather than dark. What functional-flow underlies brightness and lightness?

5.1 Representational Disparity Between Spectral and Psychological Colours

A distinction can be made between spectral and psychological colours. Spectral colour refers to the wavelength of light from the visible spectrum that stimulates the production of a colour experience. These stimuli can be represented as the relative firing rates of three opponent neural mechanisms as in the Visual Opponent Process model of Hering – see Figure 7a, and Clark [Clar93, p150]. Psychological colour refers to the qualitative experience of a colour. These can be represented as points on a circle – see Figure 7b. When equating these two types of colour there is a discontinuity in the spectral wavelength between red and violet at which the psychological purple colours are located. Therefore, these two types of colour cannot be directly dependent, some further mechanism must mediate between the neural mechanism in the Hering model and the experienced colour.



a) Visual Opponent Processes b) Psychological Colour Circle

Figure 7. Representing colour.

a) Hering's model of spectral colour representation as three axes corresponding to deviation from equilibrium of neuron firing rate. b) Proximity of psychological colours. The numbers correspond to spectral wavelength in n.m. The shaded segment corresponds to pure psychological colours that have no spectral equivalent – see Clark [Clar93, p150, p123].

Following from the statement on structure transduction to functional flows (S-STFR) given near the beginning of this chapter, it does not necessarily follow that in the signified perception of reality all properties are determined directly by the affiliated modality. For example, with respect to colour:

1) It is used directly, *as a stationary-representation*, to construct a spatial model of the objects of reality.
2) It is used, as a *temporal-representation in a functional-flow way*, in combination with innate beliefs, to add information to objects in the signified reality.

Compare this to how sound is added to an object. Notice that point 2) is a separate process to 1). In point 2), colour is an indicator, that has been calibrated with objects, not with respect to spectral frequency, but because the object had a utility to the animal during evolution, which then developed a 'mood' toward objects with that spectral property. This 'mood' is now experienced as colour. Part of the delusion is that the system 'repaints' objects in the signified perception of reality with this mood. The system does not have to representationally fill every point in a (homogeneous) region with colour, it just instigates the dispositional belief that the region is so 'coloured'. This is part of the categorisation process and underlies the reason for colour constancy. Hence the following:

Statement on Quale Constancy. (S-QC)

In the hypothetical system, qualia, being functional-flow encodings of aspects of reality, remain unaffected by secondary environmental effects that do not impinge on the quale's functional-flow.

5.2 What Are Sensory Experiences?

As part of the signification process, qualia are encodings of an aspect of reality. Why is red 'red', why is it experienced as 'red' rather than 'blue'? As to why a quale is experienced as such, the question is, what is its causal functional-flow in the signified perception of reality? Hence, there is the role red models, but what is the connection between this role and a red sensory experience being 'red'? According to the dispositional perspective:

Definition of a Quale. (D-Q)

A quale is a categorised causal-flow encoding of an aspect of reality in a causal structure based on instantiated dissociated dispositions, which is manifested as a region in the perceptual space of the signified perception of reality.

Consider the meaning of "sensory experience". From consciousness, there is an innate belief of what "to experience" means, e.g. what the sensory experience of red corresponds to. How much is this attributable to the perspective imposed by the signified perception of reality? For example, the system is deluded into having a folk-psychology belief about sensory experiences. The property or mechanism to which a sensory experience corresponds is the causal-flow of the thing experienced in the signified reality. According to the dispositional perspective:

Definition of Sensory Experience. (D-SE)

To have a sensory experience means the conscious comprehension at the phenomenal level of the causal-flow the quale serves in the signified perception of reality.

Notice, though, that the conscious comprehension is the sensation itself, and is due to the underlying dispositional mechanism in conjunction with the points made in the prior section (i.e. causal, located and oriented). In addition, notice how folk-psychology had been burdening it with a stationary-representation sense, not a dispositional functional-flow sense.

Consequently, the distinctiveness of qualia is due to the relative difference in their underlying functional-flows. For example, visual space is non-linear in that discrimination is more sensitive near the principal axes in the Hering model, and due to evolution certain flows come to be associated with colours: red – hot, blue – cold, green – nature. The exact quale experience is determined by its functional-flow, which means 'hot' would always be associated with redness rather than blueness. This leads to the statement:

Statement for Distinctiveness of Qualia. (S-DQ)

The relative distinctiveness of qualia is a consequence of the contrast between the functional-flows of the dissociated disposition instances that give rise to them.

5.3 The Psychological & Environmental Role of Colours

This section emphasises the importance of the causal-flow of colours to the dispositional perspective. It recasts the statement for the reason for modal functional independence in terms of the requirement for object distinguishability.

The visual opponent processes model (VOPM) is only a stationary-representational coding of the reality property of the wavelength of light, i.e. spectral colour. The coding imposes some constraints, structure and form on the perception of reality. This has to be converted to a functional-flow embedded in a perceptual space to be conscious, although it can be used as a representation to unconscious decision mechanisms. This is required anyway by the conversion mechanism. Is the VOPM in some way a part of the constraining support for an instantiated dissociated disposition structure? What is the role, causal influence and significance of colour to the signified reality? Do primary psychological colours (RGBYBW) have a main characteristic flow that is interrelated to the others? Are there degrees of functional-flow as a colour changes from green to red, say? Colours are dynamic ongoing events.

Therefore, to break the topologic symmetry of the axes in the VOPM, two axes need to be fixed and each axis given a distinct dispositional type, i.e. the axes must correspond to different dispositional tendencies. Each axis must have a unique causal-structure in relation to the others (and absolutely?), i.e. ground in some structural way so that the whole axis system cannot be rotated, otherwise they would be signifying the same things. This suggests the role of colours is considered in relation to (defined by) beliefs and attitudes, such as emotions.

5.4 The Sensory Experience of Colours

This section brings the above points together and provides an explanation within the dispositional perspective as to why colours are experienced as such.

Remember that the qualitative sensory experience of something is a property of the phenomenal level and it is unique because of its unique functional-flow at the dispositional level. To be distinguishable in conscious, things must have a different functional-flow if they are properties of the same entity, otherwise they would occupy the same causal flow in the network of instantiated dissociated dispositions for that entity. They would have an identical effect on the decision and attention mechanisms.

Hence, a colour is experienced in the way it is because of a set of instantiated dissociated dispositions that instil a functional-flow sensation to do with innate beliefs correlated with the spectral colour. For example, red is ‘red’ because of innate beliefs concerned with danger and so forth. It would be interesting to determine the nature of the innate beliefs we hold for colours, some possibilities are: black signifies emptiness, white signifies neutralness, green - naturalness, blue - space, red - pay attention to object, yellow - unease or happiness? Hence, blue and red are experienced as different because they have different dispositional flows – in contrast to the trivial manoeuvre of changing the word that stands for red from ‘red’ to ‘blue’. They cannot be inverted in a wide or narrow sense. It is the causal-flow that is experienced as a colour at the phenomenal level. Functional-flows cannot be inverted, not without inverting the experienced colour.

Therefore, the sensory experience of red is due to a specific causal structure, independently of other colours. However, now they are related, ordered, and their dissociated disposition basis allows a smooth progression between each. This could imply colours belong to the same category, i.e. share some types of dissociated dispositions and support mechanism. The shared types of dissociated dispositions may simply be the locational and surface filling ones. Contrast this with how we cannot transform between sounds and colour.

Consequently, colour literally acts as a knowledge indicator in a causal functional-flow way to the decision and attention mechanisms (cf. causal role theories of meaning). It is a set of instantiated dissociated dispositions about knowledge kinds that belong to the same equivalence class defined by the original spectral colour. Therefore, the phenomenon of a particular colour is due to a particular set of instantiated dissociated dispositions that capture the evolved beliefs. This means early creatures would have seen the world in a colourless way, depending on the relative rates of development of consciousness and their knowledge about reality.

The above explains how, once the structure of beliefs for red is known, it can be re-created in an artificial creature, and not fall foul to inverted spectra. Hence, by discerning the causal structure of the psychological colour primaries, and maybe other hues, an artificial creature with similar qualia experiences to our own could be created. Notice, though, that the creature may place a different significance on the colour. Consequently, a particular structure can be pointed at and said to give rise to the sensation of red, rather than blue. However, this does not explain why red has to give rise to the particular sensory experience of 'red', but see the next requirement.

This still leaves some open questions. For instance what of creatures with more than four psychological colour primaries? What do they signify? What determines the essence of the beliefs corresponding to a colour primary? Is this an open set? If we had more spectral primaries, would there be any more psychological primaries? This leads to the following requirement:

Requirement for Universal Quale-Disposition Structure Correspondence. (R-UQDSC)

A specific network of instantiated dissociated dispositions in a particular system will correspond to a universally unique quale subspace. The exact perceptual space sensation experienced by the creature's perspective on to the signified reality can only be experienced by the creature and other systems that possess a sufficient embodiment of an equivalent network of instantiated dissociated dispositions.

A corollary being that a quale can be identified with a particular topologically invariant generic network of causal interactions obtained by parameterising its corresponding network of instantiated dissociated dispositions. This raises the possibility that modalities could have similar qualia if the parameterised networks were equivalent. However, the complexity of a particular network would make their structure highly specific and thus reduce the probability of common inter-modal quale.

Therefore, an explanation as to why colours are nomically experienced the way they are is as follows:

There is a universal psychological colour space, which is closed, i.e. all psychological colours, for any creature, would be a subspace of this. Specific colours correspond to a unique causal structure in this space. Then, coincidentally, specific colours are experienced from this space because the causal structure supporting the functional-flow beliefs held about spectral colour overlaps the causal structure in the space corresponding to the psychological colour. This reaches an irreducible relation wherein there is a universal assignment of colour to causal structure that is as inseparable as Form is to shape.

Notice that in pragmatic terms, a creature would develop a set of dissociated dispositions for certain objects encountered in its environment. The sophistication of these dispositions would develop according to the adaptive value of the object. This would lead to secondary dispositions being added to reflect the survival significance of the object, which would in turn contribute to determining the region of the universal psychological colour space the network of instantiated dispositions overlaps. For example, a triangular form is a planar shape with three edges. Similarly, the colour red would be identified with a particular causal structure in a particular context. This would then only leave the problem of determining what causal functional-flow structure corresponds to a psychological colour, which, theoretically, could be determined from empirical neurological experiments.

This explanation should be contrasted with Nagel's [Nag74] argument on how we could never know what the sensations are like of another creature. It should be stressed that this is not suggesting there is an unknowable aspect to qualia, for every aspect is explainable. Rather, it suggests that ultimately a full explanation of the mechanism of qualia, while complete, would not confer to an individual the experience of the quale. However, we would know what beliefs it related to and how it compared in a discriminatory way to the other colours, but we could only experience it by sensing it directly. Notice the implications of this for the knowledge-gap argument.

6 Outline of Hypothetical System Structure

This section presents an overview of the gross structure of the hypothetical system. This is shown diagrammatically in Figure 8.

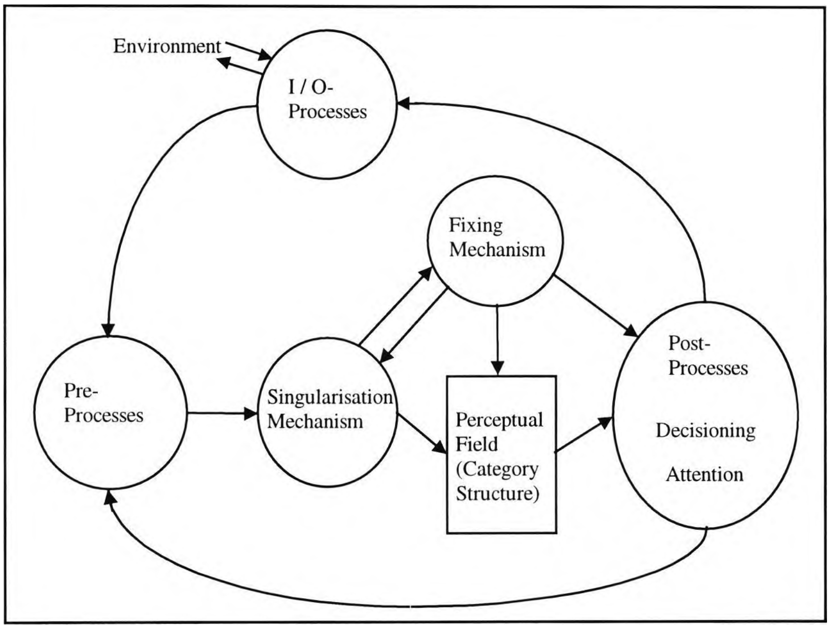


Figure 8. Logical view of main components and pathways in hypothetical system.

Notice that components are actually highly interconnected, integrated and distributed. This structure is a logical portrayal of the main components needed to explain qualia.

From the diagram of the gross hypothetical system structure it is very easy to mistakenly imagine a periodic wave of signals sweeping around the pre-process, perceptual field, post-process loop in correspondence to each new state of consciousness. This is not so. Consciousness, as the ability to use signs, is a property of the phenomenal level arising from the graded interaction of networks of instantiated dissociated dispositions at the dispositional level. It is better to imagine a cloud of activity rather than surges of signals.

The components themselves are shown as modules, although in fact they are highly interconnected, integrated and distributed networks.

Statement of Hypothetical System Structure. (S-HSS)

The Hypothetical System Structure is a feedback process featuring four essential mechanisms: Singularisation, Fixing, Category Structure and Decisioning. It is also surrounded by pre, post and i/o processes.

6.1 I/O Processes

This encompasses low level processing for dealing with efferent and afferent signals (i.e. input and output) from and to body sensors, muscles and glands. Typically, this will be pattern transformations and autonomous processing.

6.2 Pre - Processes

This encompasses all the grunt work, such as pattern recognition, analysis of sensory modalities in order to detect form, exploiting past experience (i.e. memory) to categorise, extrapolate and predict. These processes are highly integrated with the post-processes and the perceptual field mechanisms. This is in contrast to modular structures proposed in the past – see Jackendoff's original view [Jack87].

6.3 Post - Processes

When talking about primary conscious, this will encompass a decision mechanism and conditioned responses that are driven by survival. The attention mechanism is a sub-process that directs the focus of consciousness. When talking about secondary consciousness, the post-processes will also include mechanisms for higher cognitive skills, such as problem solving, 'controlled' thought, language, and imagination.

6.4 Fixing Mechanism

This is an essential component of consciousness. The fixing mechanism is highly integrated with the singularisation mechanism and category structure. Its job is to relationally fix, via qualia, entities in the signified perception of reality. It does this in two respects: by location in reality, and by orienting them in a presentational manner. Notice that it is this fixing process that binds entities into consciousness, and which helps qualitatively unify consciousness in conjunction with the post- and pre- process loop.

6.5 Singularisation Mechanism

The singularisation mechanism is the corner stone of the system. An important job for the system is to aid the survival of the animal, which means identifying food, friends, and foes. The singularisation mechanism's primary task is to categorise entities encountered in the environment so that they can influence the decision mechanism and aid survival.

6.6 Category Structure (Perceptual Field)

It was explained in chapter three that the perceptual field is the support mechanism for the perceptual space. This is equivalent to the category structure over which categorised entities are situated. Remember that entities are the phenomenal projection of networks of instantiated dissociated dispositions, so there is only a weak analogy between conventional frames and entities, and the category structure. Finally, notice that as a creature's level of consciousness increases then so too does the complexity of the category structure as it becomes a more detailed signified perception of reality.

6.7 The Category Structure and Attention Mechanism

The next stage in the system structure would be an attention mechanism based on activity in the category structure. This section looks briefly at this mechanism.

Think of attention as a concurrent cone of senses focused on an aspect of reality. The mechanism sits on top of the category structure. The increased sense of awareness experienced with convergence on cone centre is due to additional layers of instances of dissociated dispositions that embody the attention mechanism. These layers involve concurrent sets of instances of dissociated dispositions with similar prominence in the attention cone, i.e. a prominence related to their proximity to the centre of the cone. As the area of attention converges on the centre, i.e. becomes more focused, the number of instantiated dissociated dispositions increases, per unit area, to achieve increased awareness for that area – see Figure 9.

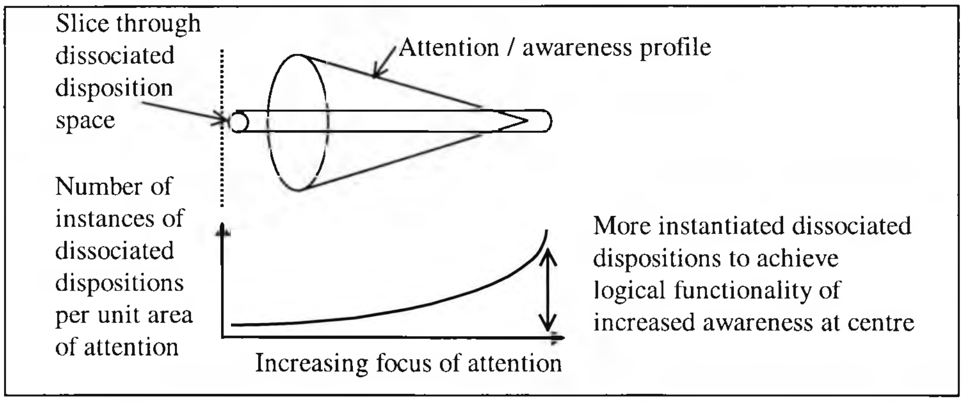


Figure 9. Dependence of number of instantiated dissociated dispositions with focus of attention.

A grandmother cell at the centre of awareness is not being advanced. In fact, there will be a multitude of entangled dissociated disposition instances at the centre, physically spread over a few system structures. Some of these areas will be specialised, while others will have a more fluid parametric causal-flow structure. This is because all entities in consciousness will be categorised in the category structure, which imposes a levelling hierarchy on all entities. In addition, it suggests that the category structure instantiated dissociated disposition density follow the attention cone profile.

As the centre of the attention cone is approached, the creature is able to sense more detail i.e. a more detailed signification. In addition, the centre has a control mechanism, linked to knowledge and curiosity senses, which direct the focus of attention, in a number of ways, e.g. a spatial sense, level of detail, and type of focus. The goal for survival being to detect movement and important objects, or to refine an entity of abstraction if thinking – see Figure 10.

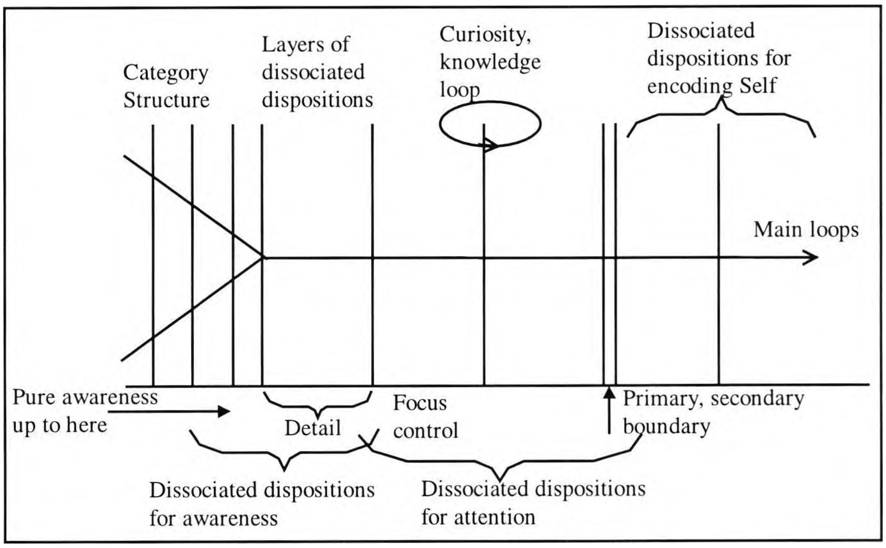


Figure 10. Layers of dissociated dispositions in the Category Structure - Attention Mechanism.

Iconic, indexical, and symbolic transitions would be effected by shifting from an initial pattern of instantiated dissociated disposition activation, PD1, say to another, PD2, when the association is iconic, PD3, say, when it is indexical, and PD4, say, when it is symbolic, as a function of the system's other inner states.

6.8 Dissociated Disposition Phases In the Category Structure

This section examines some of the phases involved in the classification of a small part of visual experiences. This is a preliminary investigation into the organisation of the category structure.

The following treats the category structure as a number of overlaid phases. A detailed qualitative experience would be built up from these phases. Remember that the phases are not perceived as such, but result in more detailed experiences. Each new phase adds more detail (through signs) to the qualitative experience associated with an entity in the category structure – see Figure 11.

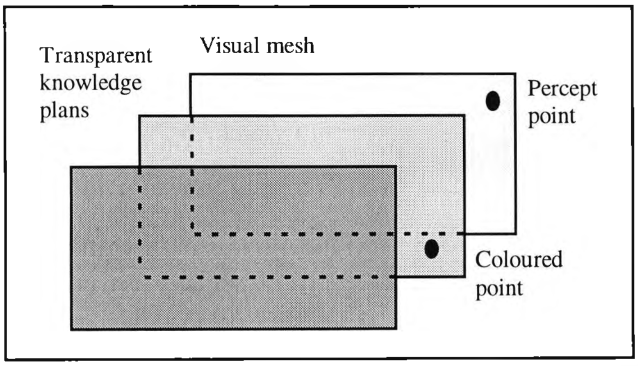


Figure 11. Logical portrayal of category entity hierarchy.

As a first attempt at specifying the nature and dependencies between phases, consider the following description:

6.8.1 Visual Topology Phase:

$\exists Q.V_{PP} \equiv \{D_{Depth}, D_{Vertical}, D_{Horizontal}, D_{Colour}, D_{Texture}, D_{Binding}\}$,

where Q is a quale corresponding to a V_{PP}, a visual point percept, and D_X are instances of typed dissociated dispositions. Note that a visual point percept will correspond to a compound sign. D_{Depth}, $D_{Vertical}$ and $D_{Horizontal}$ are instances of location dispositions, they give the modality its gross structure (gross sensory experience). D_{Colour} and $D_{Texture}$ are primitive properties, they give a modality fine structure (fine sensory experience). Notice that we tend to be only aware of the fine structure. Informal definitions of dissociated dispositions would be something like:

$D_{Depth} \cong$ (Role: rank in topology and relation to other D_X, i.e. who it affects in causal chain).

$D_{Vertical}$, $D_{Horizontal} \cong$ (Role: radial measure rank).

6.8.2 Object and Knowledge:

$\exists Q.V_{OP} \equiv \{D_{PDepth}, D_{Connective}, \ldots\}$,

where V_{OP} is a vision object percept point for a V_{PP}. D_{PDepth} is the perceived aesthetic depth, and $D_{Connective}$ is a spatial connective indicating parts of object:

$D_{PDepth} \cong$ (Role: context cues, e.g. atmosphere, environmental surround etc.).

$D_{Connective} \cong$ (Role: connects neighbours, only includes those it is a Part-of).

6.8.3 Singularisation:

$\exists\{Q_i, i = 1..n\}.V_{COP} \equiv \{D_{BindObject}, \cup \{D_{PPO's}, D_{BindVCOPs}, D_{Order}\} \forall i\}$, meaning, there is a set of qualia bound into a perceived compound object percept V_{COP}, where the union is a hierarchy of the component objects, which may be compound. Hence, D_{PPO} is a perceived point object bound together by $D_{BindVCOPs}$, and ordered in the main object by D_{Order}:

$D_{PPO} \cong$ (Role: has V_{PPs} as objects, activity of causal-flow modulated by awareness).

$D_{Order} \cong$ (Role: relative to line of sight and attention).

6.8.4 Knowledge Percepts:

$\exists\{Q_i, i = 1..n\}.V_{KP} \equiv \{D_{Bind_i} \cup \{D_{COP}, D_{KL}, D_{Label}\} \forall i\}$,

where D_{COP} is a compound object disposition, D_{KL} refers to the knowledge's functional-role, it could be its name, an attribute etc., D_{Label} refers to the network of instantiated dispositions and mechanisms delineating this component of the percept. Some of the possible attributes are: “know object”, “like object”, “important object”, “normal object” etc.

6.8.5 Self Percepts:

$\exists\{Q_i, i = 1..n\}.V_{SP} \equiv \{D_{Bind_i} \cup \{ D_{Control} \} \forall i\}$,

where $D_{Control}$ refers to dispositions taking part in control and signifying the Self, e.g. awareness and attention.

6.9 Generic & Knowledge Dissociated Dispositions

In the previous section, knowledge percepts and related dispositions were presented as part of the category structure. The interplay of knowledge via instances of dissociated dispositions and the underlying structures is an important subject. However, in this thesis it is taken to fall within the domain of secondary consciousness and pre-qualitative processes. Therefore, it is not directly applicable to these investigations. This brief section considers the nature of knowledge dissociated dispositions.

Consider the stages in a scene analysis scenario. The initial stages involve image constructors, such as edge, patch, surface, depth and colour, which are independent of the nature of the scene. The later stages involve object interpretation that is carried out with the aid of domain dependent knowledge. Are there generic or specific dissociated dispositions for dealing with knowledge? For example, consider the concept of ‘wheel’. Is there a specific dissociated disposition that plays a causal-role for the concept of wheel, or can a generic disposition be 'programmed' to have this causal-flow? Are there meta dissociated dispositions for processing knowledge? Can the class of knowledge dissociated dispositions be handled with generic dissociated dispositions? Recall that colours are postulated to be knowledge indicators generated by a tailored set of instantiated dissociated dispositions.

7 Summary

The objective of this chapter was to suggest how mind might arise from semiotic processes based on dissociated dispositions when structured in the right way. The interpreter regress problem was used as a guide to refine this structure and was motivated by the semiotic conception of consciousness as the ability to use signs. The analysis started by suggesting there was a need for a category structure upon which to build the signified perception of reality. This implied an operational transduction from stationary representations at the implementation level to functional-flows at the dispositional level. This led to an informal theorem for the ontological level triplicity of the hypothetical system.

The justification for the introduction of ontological levels was a consequence of implementing semiotic processes in terms of causal-flows and establishing the right kind of causal structure between flows. This structuring is driven by the system's need to signify reality by categorising these aspects as operational objects upon which decisions can be made.

The section on the requirement for system structuring for causal accessibility, the most important section in this chapter, presented the first half of an explanation for how semiotic processes might arise from dissociated disposition processes – the second half is presented in chapter six. Therein, consciousness was suggested to arise through the presentational manner in which the signified perception of reality is generated. This was summed up by suggesting a quale is conscious because its instantiated dissociated disposition support is causally and topologically located in the perceptual space and it is signified in a presentational oriented manner. This makes mind the result of a co-ordinated ongoing system wide activity. The remainder of the chapter considered the implications of this and suggested how it could lead to explanations for the origin of sensory experiences.

The task for the next chapter is to determine whether a semiotic system can be practically implemented in a digital machine. Leaving chapter six to explore how semiotic processes might be ground in a machine and so conclude the second half of the explanation for how semiotic processes might arise from dissociated dispositions.

Chapter 5

Implementing Causal-Flow Systems

1 Overview

Minds were defined as a type of semiotic system, which was shown to contain a hierarchy of three levels: implementational, dispositional and phenomenal. In turn, the dispositional level is a type of causal system characterised by it's causal-flows and their causal structure. Consequently, this chapter considers how a causal-flow system might be implemented.

Traditionally, in system design attention is paid to the permissible state transitions. With instances of dissociated dispositions, the causal structure between the order of these conventional transitions is of more importance. An instantiated dissociated disposition's causal-flow enforces a particular structure on the temporal order of these conventional transitions. Hence, programming dissociated dispositions is partly concerned with specifying this dynamic causal structure. They can be thought of as demarcating a confluence of states in a causal state space, except that at the dispositional level, the state space is open and continually changing – see Douglas Lind and Brian Marcus [Lin95].

From a reductionist point of view, dissociated dispositions could be specified by giving the state transitions at the implementation level. However, this would be a monumental task, and would miss the causal structure that was under investigation. To compound the situation, they are evolving, self-programming entities. Hence, after only a brief period, the current state of the system will be characterised by a different disposition configuration. This emphasises their evolving, malleable and transient nature - a difficult nature to program.

1.1 Logical Development

The goal of this chapter is to push the formal characterisation of dissociated dispositions to the point where the main requirements for a modelling technique for specifying systems based on them can be deduced. This proceeds by examining how they can be described and modelled mathematically, which prompts the need for a tailored programming language.

Part of this exploration involves a practical example that is investigated to demonstrate the applicability of modelling with dissociated dispositions. In addition, it also sheds some light on "Causal Alteration" – an important process in the dynamics of instantiated dissociated dispositions. This embodies the fundamental mechanism by which the dispositional form of a dissociated disposition is coded. This is discussed in the section on implementing instances of dissociated dispositions, and in the section on the generic Gaussian alterator.

2 Notation & Terminology

This section introduces the notational conventions used in the following and some terminology from group theory and fuzzy set theory.

Lower case letters are used to denote atomic values, these may be distinct operators, scalars and parameter values or activity patterns that are treated as a value. Upper case letters denote sets, tuples, parameters or variables. Two types of operators are distinguished: those referred to by lower case letters denote distinct operators, and those denoted by upper case letters with a subscript 'f', e.g. M_f, are operators defined as a fuzzy set (strictly speaking, a fuzzy tuple) of distinct operators. Greek letters will be used to denote functions, e.g. ψ, μ, ϕ. Definitions are denoted by $=_{df}$, as in $\psi(a, b) =_{df} a+b$.

2.1 Group Theory Terminology

A small amount of terminology is used from group theory. To start with, a *monoid* is defined as a (mathematical) system comprising a set of elements S, a neutral element e, and a binary operation μ. A *group* is a monoid in which every element is invertible. A group *operator* is an element that permutes the elements of a group – see Cohn [Cohn93, p42] and the appendix for details.

An important aspect, which is exploited here, is the distinction between properties of the group, and the nature of its elements. Group theory is primarily concerned with the characteristics of the group and operations on the group, rather than the nature of the elements. In computational terms, there is a denotational semantics for the group, at the group level, and an operational (or denotational) semantics for the elements of the group and what they stand for. A large part of the former is already defined for us in conventional group theory, leaving us to specify the particular group used and the notational conventions adopted. One aspect that will have to be dealt with carefully is that the group elements may operate on the group itself, consequently complicating the semantic distinction somewhat. This, and defining the nature of the elements, is the main objective of the chapter.

2.2 Fuzzy Set Theory Terminology

Fuzzy logic is based on the principle that a situation x can simultaneously have a number of properties s_i, from a set of elements S. The extent to which it has each property is given by a corresponding set of deterministic membership functions μ_i. The range of the membership functions is typically normalised to [0, 1]. A fuzzy value is a tuple of membership values. Hence, if G denotes a set of elements, G_f denotes the fuzzy tuple of G's elements. If it is necessary to distinguish fuzzy from normal values, the latter are called crisp values. There are standard functions for converting between fuzzy and crisp values, such as the centroid. Similarly, there are equivalent fuzzy logical connectives for AND, NOT etc. – see Bart Kosko [Kosk92] for details. Notice that there is a close relationship between fuzzy logic and many-valued logic, e.g. the $\L_{\aleph_1}$ logic of Lukasiewicz. The latter can be thought of as a fuzzy logic on a unitary set, i.e. a set with one element, with a continuous membership function.

2.3 Continuous & Singular Execution Models

This subsection introduces terminology concerning the underlying mode of operation of the hypothetical system. This is an extension of the formal description of processes given in the appendix, but from an implementation perspective.

Traditional computing is about results. An expression is evaluated symbolically to determine its value. The value is used, and computation moves on to the next task, the expression being discarded if it is no longer needed. This will be referred to as a *singular* execution model. In analogue computing, the expression could be represented in an analogue manner so that its value is produced continuously while the circuit is active. There is no single point of execution, the system as a whole is active. This will be referred to as a *continuous* execution model (cf. concurrent and parallel processing). There are two senses in which the value is produced continuously:

1) Taking time as a continuum, the value is produced continuously over this continuum.
2) The value is produced at discrete points in time, typically every system epoch.

The hypothetical system follows a continuous execution model of the second, discrete kind.

If the expression was a function of time, then in both senses of continuous the value produced need not be a smooth (continuously differentiable) function, i.e. it could be stepped or quantised. For example, suppose the value represents the state of the system, and there are a finite number of designated values corresponding to the recognised states of the system. There are three ways in which the system might change state:

1) Jumping directly from state to state, implying a *stepped* function.
2) Taking small steps toward the next state, implying a *quantised* function.
3) Changing *smoothly* to other states.

In the last two ways, the meaning of the intermediate values between the designated ones is system dependent. Continuous semiotics (described below) changes state in the second, quantised fashion. According to this taxonomy, the following statement is made:

Statement of Continuous Semiotic Execution Model. (S-CSEM)

The hypothetical system follows a quantised, discrete, continuous execution model.

Let $D_X(t)$ stand for any instance of a dissociated disposition active in the system at time t. The time between quantised updates of the graded disposition is defined as Δt_D (which is related to the average maximal base time interval – see the appendix), and n_t is defined as the average number of quantised update steps the average disposition would take to change between typical primary causal-flows (see below) under ideal conditions. Hence, $\bar{t}_{\Delta D} = n_t \, \Delta t_D$, is the average time for a change between typical primary causal-flows (cf. state changes). According to the theorems developed in the appendix, instances of dissociated dispositions can be updated synchronously every Δt_D. Processes that are part of the support mechanism, such as the convergence of an OCOS neural net, must therefore operate at a rate sufficient to satisfy the synchronous update cycle. These processes could be updated synchronously as well, at the faster rate of every $\frac{1}{n_S}\Delta t_D$, where n_S is the number of support updates per disposition update. For a socially interactive individual very approximate time scales are: $\Delta t_D \approx 1$ to 10 milliseconds, $n_t \approx 10$, $n_S \approx 10$, $\bar{t}_S \approx 25$ milliseconds, which is approximately determined by the minimum event resolution required.

2.4 Modes of Operation: Stochastic & Deterministic

This subsection introduces terminology concerned with the mode of operation of the hypothetical system.

To start with, consider the operation of real neural nets. There are two ways of looking at their operation: as a deterministic device wherein every pulse and its spatial-temporal relation to every other pulse is exactly determined; or as a stochastic device wherein the exact relationship of each pulse is not critical – a certain amount of leeway is tolerated – see Brian Gaines [Gai87b].

The stochastic view can be thought of as a mixture of causal interaction as defined in the appendix or statistical processing depending upon the context. In causal interaction, the spatial-temporal relation of events by themselves is of the main importance. For example, if a number of events co-occur this signifies something, or the frequent presence of an input contributes evidence toward a proposition etc. In statistical processing individual input pulses are seldom important, it is some statistic of the inputs that matters most, such as the average rate of inputs – see Grossberg [Gss68].

The main stance taken here is to treat instances of dissociated dispositions as exact functions operating in a deterministic fashion on statistics of their inputs and, accordingly, encoding outputs as stochastic pulse streams. Noise arises from imprecise inter-disposition communication, and interference from cross talk and ghost past states. This noise is soaked up by the function's statistical leniency.

2.5 Causal-Flow Spaces

The following terminology for causal-flows is introduced. First consider the analogy between a state space and a causal-flow space: a particular causal-flow can be thought of as one state out of possibly many. All these possibilities can be arranged according to some similarity metric into a *causal-flow space*:

Statement on Causal-Flow Space. (S-CFS)

A causal-flow space is a heterogeneous multidimensional continuum of causal-flows ordered by similarity. An infinitesimal volume in the space denotes a specific causal-flow.

There is some similarity to Hopfield's analogy between the state space of a neural net and an energy surface – see Hopfield [Hop82]. In this case, states in the causal-flow space corresponds to potential causal energy (tensors?). However, a causal-flow space would be a second order state space in comparison since each infinitesimal volume represents a particular sequence of state vectors.

In what follows, evolving instances of dissociated dispositions have to be dealt with in which the type of the disposition is changing, in this case the causal-flow space is said to be dynamic meaning the type of a basis axis may be changing. While an instance of a dissociated disposition is defined over a period of time, its environment will be changing after each epoch. This means the disposition will almost certainly be absorbed into a new disposition instance that has grown to dominate the current situation. Therefore, it forces attempts at modelling this behaviour to deal with infinitesimals and instantaneous causal-flows. Recall that a dissociated disposition is characterised by a metrically related set of causal-flows, although in a given context it will follow just one flow, corresponding to an infinitesimal volume in the causal-flow space:

Statement on Causal-Flow of an Instantiated Dissociated Disposition. (S-CFIDD)

At a particular instant, an instance of a dissociated disposition is identified with the instantaneous causal-flow it is most similar to in causal-flow space.

This is called its *primary causal-flow*. At a specific point in time a causal-flow space will have a particular basis, that is, a set of variables and associated mechanism that delimits the dimensions and causal structure of the space. The number of primary causal-flows in a space will be called the order of the causal-flow space. This will be different to the dimension of the space – see below.

It is worthwhile elaborating what the above statement means. Thus, an instance of a dissociated disposition is a causally guiding pattern, defined over a period of time, which causally influences the evolution of other instantiated dissociated dispositions. If the period of time was divided into an infinite sequence of epochs, then the structure of the causal pattern would be constructed by linking together in order the causal influence of each epoch in the sequence – see Figure 12.

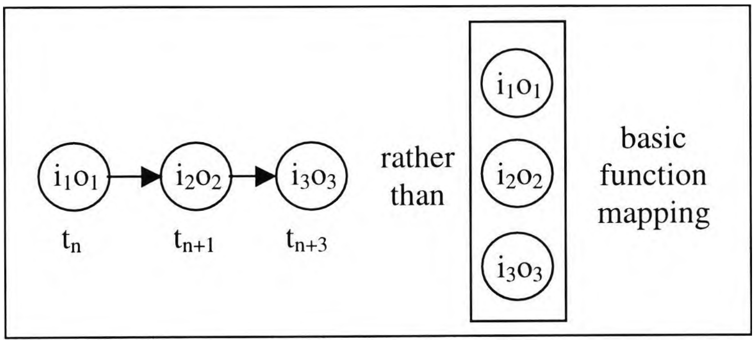


Figure 12. Contrasting temporally ordered & unordered I/O mappings.

This shows the contrast between a temporally linked sequence of input/output mappings, and a function, which can perform each of the I/O maps independently. An instance of a dissociated disposition can be thought of as imposing a temporal order on a set of I/O relations – see the discussion of temporal-representations in the appendix. The disposition is analogous to a template for a (confluent) sequence of I/O relations. This temporal structuring is parametrically embedded in the mechanism. Hence, the presence of an instance of a dissociated disposition imposes a deterministic tendency for the sequence of states to occur (in the instantaneous limit) given an embedded temporally structured environment, i.e. its environment is also similarly temporally constrained structurally.

A useful analogy is ocean waves breaking on a beach. The ocean corresponds to the support medium, which is structured by the shore profile, wind strength, direction, and ocean-bed depth. An instance of a dissociated disposition would be analogous to the evolving pattern of potential energy set up by the waves on a moving area of the ocean surface. Disposition interaction would be analogous to wave interference between ocean areas. Consequently, the major differences between this analogy and the dispositions considered here, is as follows. Firstly, rather than potential energy, it is potential causal energy. Secondly, instances of dissociated dispositions may influence the pattern of potential causality through changes in the medium structure (i.e. parameter changes, cf. coastline erosion). Finally, patterns are multi-dimensional and highly specific, e.g. instances of dissociated dispositions may be interconnected precisely.

3 Features of the Hypothetical System

To delineate the scope of the requirements for a dissociated disposition notation a broad summary of the composition of the hypothetical system is presented.

3.1 Specification of the Support Mechanism & Initial Causal Structure

The hypothetical system is defined by a specification of its subsystems, substructures and their interconnection. Each component is defined by a specification of the support mechanism and a specification of the dissociated dispositions initially instantiated on that support. This involves specifying the following aspects: a) low-level representational mechanisms, e.g. sensory transducers, b) any explicit constraints to be applied over sets of dispositions, such as normalisation, and c) the fixed interconnect. For example, with respect to interconnect, is the structure determined by specifying a particular topology, or by ordering and weighting the parameters of the relevant support functions? In what follows, and without loss of generality, the system wide support structure can be taken as fixed. Any changes in the interconnect can be viewed as a parameterised change in the strength of the connections concerned.

3.2 Specification of Dissociated Disposition Group Operations & Structure

Later it will be shown how the multiplicative structure of a mathematical group can be used to model the causal structure of a related set of instances of dissociated dispositions. This involves specifying the group table to define the group operation, e.g. does $A * B \rightarrow C$ or $A * B \rightarrow D$ etc. Other details include how an operation affects the state and output, e.g. is it temporary or permanent? Is this a property of, or part of, the group element specification? However, there is a further catch, the group table may be implicitly defined and programmable by the evolving instances of the dissociated dispositions!

3.3 Specification of Group Element Functions

The group elements are implemented as parametric functions – described later. For example, element g is the continuous function $\psi_g(X, M, H, G)$, where X are other group inputs, M is the state influence, and H is the support influence. The function $\psi_g(\ldots)$ is itself continuously defined by the group G. As its parameters move, its behaviour will move either toward its characteristic behaviour, or away toward that of other group elements. Hence, g is a group element label determined by a metric as to how close the behaviour is to particular elements, i.e. there is a label function $\phi_g(\ldots)$, that determines the elements of a group, and the group order. What has to be specified are the characteristic functions, one function for each element. For example, in the Iconic Mood Detection example described later, each of the Gaussian feature detectors can be thought of as a characteristic function. In addition, to cater for self-programming, the pattern prototype vector may be a parameter.

Therefore, the causal-structure of an instance of a dissociated disposition is specified by a combination of the group structure, the group element function, and the influence of the support mechanism. In a simple case, where the group element functions are basic logical primitives, such as AND, the causal-structure of the disposition is determined by the group table. For instance, if the input is X = AND, which is defined to mean all inputs are active, the state is M = NOT, which means it will invert inputs, and the state group table entry for these values is: AND $*$ NOT $\rightarrow$ OR, which means the state changes to OR. In addition, if there is the state group table entry: AND $*$ OR $\rightarrow$ NOT, this will then give rise to cyclic behaviour whenever the inputs are continously active. This behaviour could be identified with a disposition that *oscillates the ouput whenever all inputs are active.* In fact, the text in italics is the causal structure of the disposition and *it is* the disposition. However, notice that a table entry may act in other dispositions as well. For instance, as the input activity changes another entry becomes dominant. This is further explained shortly.

It would be convenient to keep the group element functions generic by parameterising structure influences. Similarly, with the supports influence, e.g. topologically weighting. For instance, one group element's topology weighting parameter may be set to respond to a happy mood in the Iconic Mood Detection example, while another responds to a sad mood. The topologic support remains neutral, simply acting as a relay for the causation. Although group operators may operate on heterogeneous inputs and outputs (domain and range), in most cases only homogeneous transforms will be of interest. This implies all data is accounted for by group elements, i.e. there is no data-code divide.

3.4 Notation Requirements for Specifying Dissociated Disposition Based Systems

Take an identity map or matrix and suppose that it is modulated, shaped, structured or constrained by an instance of a dissociated disposition that is based on a set of mathematical group operators. That is, in some respects the operators abstract away from issues concerned with parallelism and flow control, and instead declare the mapping in non-computational terms. There would still be a base system time epoch. The initial specification of a system would be a set of coupled identity matrices with a combination of fixed and variable internal structure. The matrices are themselves interconnected in the necessary manner. An immediate notation requirement is to be able to specify this fixed and variable aspect, and the interconnect. This can be summarised by the requirements:

1) Specification language for the semi-permanent support structure. This corresponds to resilient features.
2) Programming language for the group operations and initial instantiation. This represents the real-time operation.

Is it possible to integrate these two requirements into a coherent whole? To what extent is self-programming needed? Is the category structure really fixed or self-programmed? The stability of a scene suggests instances of dissociated dispositions are structural fixed-points. Another issue is whether dissociated dispositions are dynamically instantiated. What are the constraining relations with the implementation of the support? What are the constraints between the support and dispositions as group operators? What is the relationship to traditional combinator computing and approaches to implementing functional languages?

Concerning the number of parameters needed to describe dissociated dispositions and their group operators, what would be a suitable parametric basis? Should they be treated as having microstructure for reasons of notational convenience? Is it that the group operator self-programming perspective happens at this parametric level of description? There could be different sets of group operators tailored to a modality or function. One way to handle these microstructure variations would be to define standard sets that encapsulate the variations between functions. This suggests only the parametric mechanism need be considered, which could be achieved in a programming language fashion.

4 Selecting a Formal Modelling Technique for Dissociated Dispositions

In this section, a cross-section of mathematically oriented modelling techniques and programming languages are briefly reviewed for their suitability as an operational formalism for capturing the lawful and causal relations within the hypothetical system. This would eventually lead to implementations whereupon, by observing the system's performance, further insights may arise.

Via a naïve principle of reductionism, everything could be reduced and represented by mappings. However, this would ignore what is important, namely the structure of causal-flows. What is of interest are the nomic, reality based set of laws and constraints that structure an instance of a dissociated disposition's causal-flows. Some possibilities to consider as a formalism include:

1) Function maps and state machines. Just specifying the states of a function, or i/o map, does not capture the causal structure, i.e. what state follows next and why. For this reason descriptions in terms of state machines are below the dispositional level.
2) Group theory, group elements and combinators. As will be shown, this is flexible enough to handle a language of thought and problem solving, but is it too flexible?
3) Parameterised functions wherein instance of dissociated dispositions are viewed as function application. However, this may not be flexible enough for the top level of the category structure. The functions would have to be highly parameterised.
4) Contemporary functional programming languages. The procedural and applicative reduction order of function evaluation is not directly compatible with the causal driven mode of operation of instance of dissociated dispositions. This mode would have to be modelled.
5) Temporal and dynamic logics. Similarly, all the logic notations encountered in this area focus on a singular execution model rather than continuous operation.
6) Higher order and reflective notations. Similarly, these notations are based on a singular execution model and the ability of the model to operate on its own structure.

Points 4) to 6) are discussed in more detail shortly. Of the formalisms covered, the combination of group theory and combinators is most strongly affiliated with the nature of dissociated dispositions. In what follows this will form the main notational vehicle to which will be incorporated ideas from statistics, many-valued logic and fuzzy-logic.

4.1 Fetzer's Language for Universal Dispositions

One way to situate dissociated dispositions formally is to extend the language $\boldsymbol{U}$ for universal dispositions presented in Fetzer's probabilistic causal calculus by adding parameterised dispositions and reference class definitions, and an axiom for maintaining the numerical identity of things through time – see Fetzer [Fetz81, p62]. Along these lines, the appendix defines a type of system level property p^n_{ij} as any process for which an operator ω^n_{ij} is definable on the level domain E^n_i – see definition D-SLP. An instance of the property type is then determined by applying the operator to a specific set of elements: $p^n_{ijx} = \omega^n_{ij}\,\varsigma^n_{ij}(C^n_{ijx}, E^n_i)$, where the selector ς^n_{ij} picks out the specific collection of elements, and C^n_{ijx} describes their configuration. This can be parameterised with respect to time – see the appendix definition for a time dependent property value D-TDPVV, and the discussion on continuous versus discrete processes. The ongoing action of the selection operator thereby defines the numerical identity of the property through time.

4.2 Algebraic Languages

Programming languages based on an algebraic approach have been proposed. For example, R.B.Kieburtz and J.Lewis [Lew94] have developed ADL (algebraic design language). This is a variation on functional languages, such as ML and Miranda. Its central feature is that algebraic monoids are introduced by 'signature' declarations, which define the permissible operators and types for the domain. This is similar in principle to class definitions in C++, although it uses the stricter conventions of algebra. ADL does not make use of group operators or group structure, which would be important to a study of dispositions. It also follows a function application, reductive evaluation order.

R.H.Gilman [Gil91] illustrates how group theory can be used to describe context-free languages. A sentence in the language can be described by a product of group elements. Various theorems are proved with respect to operations between subgroups and the generation of sentences. A selection of examples is presented, for example, how the language generated by a stack-automaton may be described by operations on a monoid. It would be interesting to extend this work to a description of causal-flows. The objective would be to clarify the relationship between a confluence of similar flows (i.e. sentences) and how this may be characterised by a particular monoid element. R.L.Grossman [Grs93] shows how a discrete finite state automaton, described algebraically, can switch between continuous control systems. There are useful analogies here between the group structure on dissociated disposition alteration and the underlying (approximately) continuous system evolution. The paper presents useful mathematical tools that could be used for modelling the structure of causal-flows.

Quantum computation is a computational model that is growing in theoretical popularity – see A.Ekert [Eke93]. Quantum computations involve states represented by waves and evolution modelled through operators in a similar fashion to the dispositional mechanism. However, what sets quantum computation apart is that a superposition of computations can be carried out at the same time. This is succinctly expressed by noting that a classical N-bit logic gate has 2^N states, whereas a quantum gate has 2^{2^N}, and can process a superposition of these states in one go. There are useful analogies to be drawn with respect to models of elementary quantum gates that have been proposed and primitives for micro dissociated disposition instances – see A.Barenco et al. [Brn95].

4.3 Feedback, Recursion, Reflection, Self-Reference

To position a disposition language it helps to distinguish these different forms of operational dependency and at what system level they may arise.

Feedback can be defined as a dynamic process in which there is a causal dependency in effect from its outputs to it inputs. It can be direct, indirect, positive, and negative, in phase or out of phase. The mechanism by which the feedback loop is closed may even be via an external environmental chain. It is commonly viewed as part of a continuous execution model. Feedback is common in electrical circuits and is a corner stone of cybernetics where it is used for self-regulation. The hypothetical system uses indirect feedback at the dispositional level and direct feedback at the implementation level.

Recursion can be defined as a process that sequentially invokes itself. A recursive function, as typically defined in functional programming, is a function that calls itself during evaluation – see Simon Peyton Jones [Peyt87]. The recursion may or may not terminate. Recursion may be direct, or indirect wherein a parent function is called by a sub-function. A singular execution model is assumed. Typically, state information is saved for each recursive call, unlike a feedback process where there is only ever one set of state information.

Reflection refers to a processes ability to examine and modify its own state. One way to implement this is via continuations. A continuation is a function that denotes the remainder of the computation. Typically, this is passed as an argument when functions are called – see P.Cointe et al. [Coi96], P.Brisset and O.Ridoux [Bri93]. Related areas are meta-programming, higher-order programming and higher-order logic's – see P.M.Hill [Hil94] and C.Prehofer [Pre95]. Here the program manipulates other programs or sub-modules that may not be part of the currently executing program.

Self-reference is defined here as a process that directly refers to itself at the same system level – see R.M.Smullyan [Smul94]. This is commonly illustrated by one of the semantic paradoxes, such as the Liar paradox, which is a logical proposition that asserts its own falsity, e.g. "This sentence is false." – see Alfred Tarksi [Tar69], Mark Sainsbury [San98, p105] and Smullyan [Smul94, p81]. An important aspect is how the process performs and interprets the self-reference. It is crucial to realise that the mechanism and referent are at different system levels.

These operational models are strongly influenced by the singular execution model and function evaluation. A dispositional system is more limited, it follows a continuous execution model and has a fixed underlying physical structure to which its operational organisation is closely bound. At the dispositional level the system operates in a causal way, with parametric changes in causal structure rather than function manipulation or manipulation of control structures.

4.4 Many Valued Logic's, Fuzzy Logic & Bayesian Belief Networks

Strictly speaking, logic's are formal systems concerned with the preservation of truth – see J.Slaney [Sla90] and E.G.C.Thijsse [Thi91]. The hypothetical system is concerned with the interaction of causal-flows and their influence. The notion of truth is less tenable at the dispositional level since it implies a conventional representational point of view. However, as a basis for the propagation of causation the calculus of logic's is relevant. Therefore, rather than Boolean or Modal logic's, a suitable basis for the dispositional approach would be a pseudo infinite form of Lukasiewicz's many valued logic $\text{Ł}_{\aleph_1}$ combined with the semantics of Fetzer's universal language for dispositions – see Rescher [Resc69], G.Restall [Res92] and the appendix. Here the rule of excluded middle does not apply and values need not be truth designated. For practical reasons, discrete quantities are used to approximate $\text{Ł}_{\aleph_1}$. However, although discrete, this does not make it an Ł_n finitely many-valued logic. The semantics are still those of the infinite valued logic. As mentioned above, Lukasiewicz's many valued logic is a natural choice to use in conjunction with fuzzy logic. Notice that a deterministic mechanism basis is chosen rather than one that relies on probabilities, such as the probabilistic version of Fetzer's calculus, or Bayesian belief networks – see L.C.Gaag [Gaa96]. That is, an evidential approach is espoused. This does not rule out the 'weight' of evidence being related to the probabilistic outcome – see P.E.Green [Gre87].

4.5 Mobile Processes & Dynamic Logic's

The calculus of mobile processes, π-calculus, proposed by R.Milner et al. [Mil92], (see M.Boreale and R.De Nicola [Bor96]), is an extension to λ-calculus that allows links and (references to) processes to be passed as parameters, thus enabling dynamic structures to be modelled. The dynamic transitional state structure of processes, such as the relationship between their states, and hence their flow structure, is not a major concern. This is crucial to modelling dispositions. Unfortunately, there are no facilities for modelling this aspect in the π-calculus.

B.Penther [Pen94] presents a dynamic logic of action. This is based on propositional dynamic logic. It models agents that can perform actions. A distinction is made between static facts and dynamic actions. Again, the focus is on a logic of action from a singular execution model point of view. More importantly is the difference between purposes and actions. Penther gives the example of an action: "to close the door". There is a sense of direction and immediate change; an agent acts. Penther also talks about an agents 'ability' to act. Consider the purpose of a ruler: "for drawing a straight line", this would involve an agent carrying out an activity. Whereas, the act "to draw a straight line", is concerned with the consequences of drawing, the process of drawing itself is treated as an atomic action. Consequently, the difference between action and purpose is that the former is what an agent does, while the latter is a mechanism for carrying out an activity.

4.6 Temporal Logic's

Temporal logic's deal with the influence of time situated events in models. M.A.Orgun and W.Ma [Org94] present an overview of temporal and modal logic programming. L.Vila [Vil94] presents a survey of temporal reasoning in artificial intelligence. The temporal logic's reviewed predominantly focus on enabling the behaviour of a system to be modelled. In this respect, a temporal logic could be used to analyse the hypothetical systems behaviour. For example, logic's that model the behaviour of a process based on its state trace, i.e. sequence of state transitions – see J.C.M.Baeten et al. [Bae94]. The causal-flow of an instantiated dissociated disposition is characterised by a confluence of state traces. In would be interesting to analyse causal-flows using techniques from the field of Symbolic Dynamics – see Lind and Marcus [Lin95].

Alternatively, one possibility is to use a programming logic as a notation for the hypothetical system. For example, A.A.Faustini and W.H.Mitchell's multi-dimensional logic language 'InTense' [Fau89]. However, these programming languages follow a singular execution model. Logic formulas are to be reduced. The primarily interest of this section is in a notation for specifying the hypothetical system. Here, while the impact of time is important, in that processes must proceed at a sufficient rate, it is not explicitly represented. Time is not an important aspect of the dispositional level, although it is a crucial part of the manner in which reality is signified. What is important is the causal dependency between processes (i.e. dispositions) and that the system is active in a continuous execution sense.

4.7 Conceptual Formalisms

The theory of conceptual structures (see J.G.Lopes and M.Wermelinger [Lop94]) and conceptual graphs (see G.W.Mineau [Min94]) is a comprehensive formalisation of the notion of concepts. The field of conceptual theory stems from the pioneering work done by J.F.Sowa [Sow95,Sow92] on conceptual graphs. These are a graphical extension of logic's and combine the existential graphs of Peirce with semantic networks. They allow images as concepts (spatial relations) and are flexible enough to represent or model all current concept formalisms.

Dissociated dispositions are more similar to concepts than the function of a logic gate. Just as a concept is an abstract mental construct, so that only its support structure could be pointed to, so too with dissociated dispositions. However, an instance of a dissociated disposition imposes a definite degree of structure, in that the activity of the support over a period of time characterises a disposition. They have a causal profile. In the system, instances of dissociated dispositions were positioned at the dispositional level. Concepts can be positioned at two levels depending on the context. From an implementation perspective, they would be representations across the implementation level. From a qualitative level, they would be conscious beliefs.

What distinguishes a concept from a dissociated disposition? Concepts, as perceived at the phenomenal level, are a signification of a generalisation or a common event that has occurred. At the dispositional level, they are (more or less) integrated habits of thought and habits of action that are unified by some common element (responding to signs of the same kind, for example). Thus, concepts are derived from networks of integrated instances of dissociated dispositions. In contrast, concepts as propositional structures at the implementation level represent to the dispositional level. They would embody structure, rules, facts, knowledge, whereas instances of dissociated dispositions embody operation, transformations and ongoing causal structure. Treatments of concepts are dominated by the stationary-representational view, which focuses on the function of relational structures rather than their temporal-structure.

5 Implementing Instances of Dissociated Dispositions

This section marks the start of the operational description of dissociated dispositions. To begin with the overall structure is presented and then refined in the following subsections.

A physically extensive mechanism with a logical structure similar to that in Figure 13a is capable of supporting a group of semi-orthogonal and fuzzy instances of dissociated dispositions. The mechanism is shown divided into three phases.

Statement of Support Mechanism for Instantiated Dissociated Dispositions. (S-SMIDD)

An instance of a dissociated disposition is supported by a mechanism that can be logically treated as three integrated phases: reduction, alteration and production.

The reduction phase is concerned with projecting a high dimensional input pattern into a lower dimensional space, i.e. group. The alteration phase is analogous to a decision stage wherein the system changes state based on the current state and the reduced input. Finally, the production phase translates the new state into a high dimensional output pattern. Remember that this operates according to a continuous execution model, so inputs and outputs vary continuously in quantised steps, with disjoint jumps in behaviour being the exception rather than the rule.

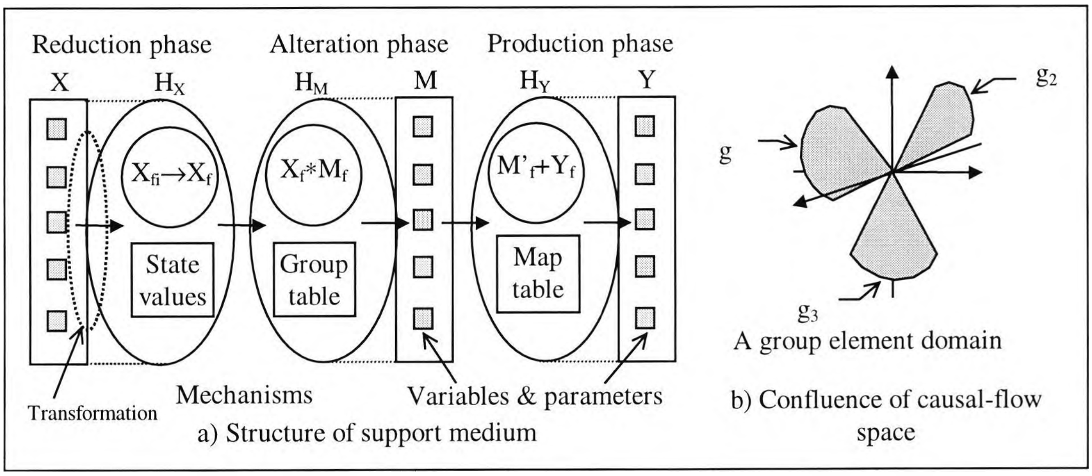


Figure 13. The support structure for instances of dissociated dispositions.
a) Stylised structure of a support medium for instances of dissociated dispositions. b) Example of confluence of causal-flows (cf. states) for a three dimensional real causal-flow space.

5.1 The General Dissociated Disposition Instance Evolution Relation

This subsection introduces disposition evolution in the hypothetical system by way of a general relation equation.

The first thing to do is separate out the fixed hardware aspects from the functional, parametric aspects. Hence, the hardware support will be denoted by a fuzzy operator H_f. The current (instantaneous) state of an instance of a dissociated disposition is denoted by a fuzzy operator M_f. For now, all inputs are denoted by a fuzzy X_f, and outputs as a fuzzy Y_f. It is useful to define a neutral operator I, which when applied to a state operator M_f, leaves it unchanged and produces the current output operator Y_f. Similarly, it is useful to define a state operator I_M, which acts as an identity element to any group operation X_f.

Definition of The General Dissociated Disposition Instance Evolution Relation. (D-GDDIER)

The general dissociated disposition instance evolution relation can be expressed as a fuzzy compound group operation:

$$X_f * M_f \xrightarrow{H_f} M_f' + Y_f .$$

Here M_f' is the next quantised fuzzy state. Operator H_f is more of a structural framework, and M_f kind of fills in parameter slots within this. X_f and Y_f are similar to boundary conditions, or sets of attributes. The relation is a compound one. This and the exact nature of the group elements is described later.

A fuzzy state operator M_f may be operated on by a set of fuzzy operators $\{X_{fi} \mid i = 1..n\}$, to produce a new state M_f' and output Y_f' – see Figure 14a.

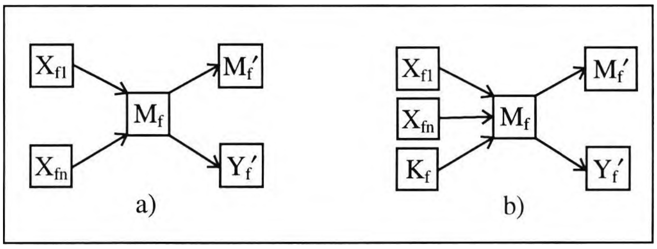


Figure 14. a) Inputs operating on state. b) Influence of knowledge on state.

In Figure 14b, the influence of knowledge on the state is shown. The operator K_f will have been produced from parental disposition instances and the support mechanism. In this case, the support mechanism acts as a source for disposition instances, the mechanisms involve a type of inference engine and knowledge domain selection. The idea that both are embedded in a mechanism and structure implies different knowledge domains are some how surrounding the central disposition instance. Is there a domain for every conceivable semantic object? Or is knowledge retrieved from a global memory on demand? How are group operators for domains created and selected? Do they arise from self-programming via operators evoked from memory following a context dependent selection mechanism? In what follows K_f can be treated as an ordinary input.

5.2 Formally Describing Causal-Flow Spaces

Continuing with the formal characterisation, for ease of exposition the parameters and variables within the reduction, alteration or production phases are divided into those that characterise the causal-flow space and those remaining that are involved with the mechanism. The associated causal-flow space and mechanism state spaces for the three phases are denoted, respectively, by tuples as follows:

$$X \text{ is } P_X = (P_{Xi} \mid i = 1..n_{PX}),\ M \text{ is } P_M = (P_{Mi} \mid i = 1..n_{PM}),\ Y \text{ is } P_Y = (P_{Yi} \mid i = 1..n_{PY}).$$

For the mechanism state spaces:

$$H_X \text{ is } S_X = (S_{Xi} \mid i = 1..n_{SX}),\ H_M \text{ is } S_M = (S_{Mi} \mid i = 1..n_{SM}),\ H_Y \text{ is } S_Y = (S_{Yi} \mid i = 1..n_{SY}).$$

The complete 'state' space for an active mechanism is the union of all of these.

The causal-flow space P_M of M can be characterised by a group G_M, where the domains of its elements correspond to a confluence of paths through the causal-flow space – see for example Figure 13b. Typically, the element domains will be hyper-ellipsoids, or hyper-cones. They are not necessarily disjoint, and boundaries may be fuzzy (in a fuzzy logic sense), e.g. defined by a Gaussian function (cf. Lie Groups). Let the group $G_M = \{m_i \mid i = 1..n_{GM}\}$ for some n_{GM}, and let $\psi_{mi}(P, S)$ denote a group element function, where P is a vector or matrix of causal-flow variable values, and S a vector of mechanism parameter values. In the example shown, this function could be a simple function of a vector inner product: $\psi_{mi}(P, S) =_{df} \varphi_g(\Sigma_j\, p_j\, s_j)$, where $\varphi_g(\ldots)$ might be a Gaussian function such as the one used in the Iconic Face Mood Detector. These functions will typically have a single maximum, which will be called their *primary mode*. Often this will correspond to a unique pattern of input variables, this will be called their *characteristic value*. The causal-flow space P_X can be characterised in a similar spirit to P_M, by the group $G_X = \{x_i \mid i = 1..n_{GX}\}$, and $\psi_{xi}(P, S)$ the corresponding element functions. Similarly for the causal-flow space P_Y, $G_Y = \{y_i \mid i = 1..n_{GY}\}$, with $\psi_{yi}(P, S)$ element functions.

Notice that one of the group members will be a neutral, or identity, depending on context. Also notice that the dimension of the causal-flow space will normally be greater than the order of the group. However, it is desirable to have the primary modes of the group correspond to the primary causal-flows. In what follows this will be assumed.

The mechanisms in the dispositional level support structure share a number of common features, such as the group nature of the causal-flow spaces. To exploit this, the above group notation is now generalised and the input notation for each phase extended to tuples of inputs. Let $G = \{g_i \mid i = 1..n_G\}$ stand for one of the groups G_X, G_M, G_Y, with corresponding causal-flow space $P = (P_i \mid i = 1..n_P)$, and state space $S = (S_i \mid i = 1..n_S)$. Let $X_f = (X_{fi} \mid i = 1..n_X)$, stand for a tuple of input group operators acting on G_f rather than a single operator, where each group is denoted by $X_i = \{ x_{ij} \mid j = 1..n_{Xi}\}$. Let $K = X_i$, $i \in \{1..n_X\}$, denote a specific input 'knowledge' group and K_f a specific operator. For example, K might represent the contextual influence of a knowledge domain. V_f is defined to be the tuple X_f excluding K_f, $V_f = (V_{fi} \mid V_{fi} \in X_f, V_{fi} \neq K_f)$. Let $Y_f = (Y_{fi} \mid i = 1..n_Y)$, stand for a tuple of output group operators produced from G_f. Finally, let $H_f = (H_{fi} \mid i = 1..n_H)$, stand for a tuple of mechanism based group operators acting in a left or right sense (i.e. pre or post operative) on X_f, Y_f and G_f.

5.3 Applying Operations

Consider the operation $A * B \rightarrow C$. This has two interpretations for the result C, it could be:

1) A self modification, e.g. $C \equiv B'$, where A is said to operate on B.
2) A transitory or permanent value, in which case it may need to be buffered or stored in a memory for future use.

The first case corresponds to an alteration action, the present state is altered in some way. The second case corresponds to a type of production action, an effect is produced which is a consequence of the operation. Hence, the terms *alterative action* and *productive action* will be used to distinguish these two cases.

5.4 Transforming Operations

Since the group operations are being used to model the causal structure of a dissociated disposition it is necessary to examine whether causation is preserved following algebraic transforms according to the formal laws, if not then the law cannot be used in transforms:

Associative law: $(x+y)+z = x+(y+z)$, $(xy)z = x(yz)$. Necessary to be a true group.

Commutative law: $x+y = y+x$, $xy = yx$.

Distributive law: $x(y+z) = xy+xz$.

The first consideration is to distinguish between group *operations* $\mu_o(x, y)$ and group *actions* $\mu_a(x, y)$. In computational terms, they have different parameter types. The group operator's arguments must be elements from the same group, as is its result. Whereas the group action's arguments are different types, the first is an element from a set to be operated on, the second is an element from a group that acts on the first. The result is an element in the first set. Therefore, argument types must be considered when applying the above laws.

The second consideration is the atomicity of the operation and its causal significance. Contrast purely atomic operations on primary causal-flows, and quantised operations on intermediate causal-flows. The first kind corresponds to the regular notion of a discrete operation, the second corresponds to a drawn out quantised-continuous operation. Assume that an atomic operation takes one update step to complete, whereas a quantised operation takes n_t steps. Clearly, each quantised step could have some causal significance for dependent processes.

In conclusion, to apply the transformation laws the following must be observed:

1) The causal profile of a dissociated disposition instance is preserved with respect to dependent processes.
2) The relevant group operation or action upholds the law in the first place. For example, group operations need not be commutative.

Notationally, it may be convenient to specify when the order of operations is, or is not, significant (be that for causal or lawful reasons). The previous chapters suggest only the causal profile of the instantiated dissociated disposition as it influences its environment is significant.

5.5 The Reduction Phase

In this and the following sections, the various possibilities for updating the general evolution relation are enumerated.

The main purpose of the reduction phase is to reduce a non-specific high dimensional pattern of activity to a specific group operator. Remember that inputs are statistical quantities. Effectively this is simple statistical pattern recognition. Its other purpose is to compose and transform operators from the inputs into a sequence to be applied to the main dispositional state G in the alteration phase.

The dendritic trees of a neural assembly can be thought of as a set of trees that collect and combine neural activity from various sources. These input trees correspond to influences from impinging instances of dissociated dispositions.

A tuple of inputs can be treated in two ways. Firstly, they can be combined to form one operator, i.e. their combined pattern of activity makes up a representation which then determines an operator. That is, the point defined by the tuple of X_i's can be viewed as denoting a single (fuzzy) operator. Secondly, each input corresponds to an operator that can be applied to the state in turn, where the state is modified after each operation. Alternatively, the operators can be composed into one operator and then applied. In both cases the operators could be fuzzy. Finally, composition involves straightforward sequential combination of operators by either multiplicative or additive operations. Notice that these will be group actions.

In what follows X_f and K_f will be treated as a generic operator, i.e. distinct or fuzzy depending on context, with a defined group composition.

5.6 The Alteration Phase

The purpose of the alteration phase is to alter the course of causation from a disposition's causal-flow perspective, i.e. at the dispositional level. In a reductionist sense the reduction and production phases alter or guide causation, but at a sub-causal-flow level. To focus on what is happening the alteration phase is divided into three simultaneous mappings (for clarity reasons reverting back to the notation for the alteration phase M):

1) State transform μ_{fM}: $X_F \times M_F \rightarrow M_F$. Pseudo-homomorphism.
2) Output production μ_{fY}: $X_F \times M_F \rightarrow Y_F$. Pseudo-heteromorphism.
3) Group transform μ_{fG}: $X_F \times M_F \rightarrow G_M'$. G_M is transformed by μ_{fG}, which is itself a function of X, G,...!

With the understanding that in these mappings X_f acts as the context. The capital 'F' subscript denotes the space of fuzzy operators for the operator concerned. Fuzzy operator space M_F is defined on the tuple of group operators for M, i.e. G_M. The possible ways in which these mappings might act are:

1) Act on current G_f, e.g. $X_f \,.\, (K_f \,.\, G_f)$.
2) Parameter influence on G_f, e.g. $(K_f \,.\, X_f) \,.\, G_f$.

Here G_f is the current fuzzy set of g_i's. In both cases the action may be alterative or productive. Each of the mappings is now considered in turn. In what follows, it must be remembered that the dynamics are quantised continuous, implying that it takes a number of iterations to change between primary modes.

5.6.1 The State Transform, μ_{fM}

The actual group action X_f*M_f defined by the mapping μ_{fM} is context dependent. At a particular instant, a pseudo group table equivalent will be definable. However, the table is implicitly encoded, possibly parametrically, into the mechanism. For example, if the causal-flow space P_M was restricted to the surface of a hyper-ellipsoid and the state functions $\psi_{mi}(P, S)$ where as described above, then the group element functions could be direction operators, incrementally translating the current point in P_M toward a specific primary mode:

$$\psi_{xi}(P, S, D) =_{df} \varphi_d(d_i), \qquad X_f*M_f \equiv (\varphi_b(\varphi_g(\Sigma_j p_j s_j) * \varphi_d(d_i)) \mid i=1..n_{GM}).$$

Here d_i is one of a set D of direction vectors, which are not necessarily orthogonal, and $\varphi_b(...)$ is a boundary function which projects or transforms the result to the nearest point in the elements domain.

In the above example, the primary modes could be thought of as the elements in a cyclic group C_m defined by generators: **Z**/m, m>1 – see Cohn [Cohn93, p56]. The order of the group, in this case the number of elements, equals the number of primary modes. The notation **Z**/m means that a number n, such that $n \equiv i \pmod m$, is a member of the i^{th} element, i.e. $n = i + m*d$. This could be interpreted as a set of overlapping spirals on the surface of a sphere, with the centre of each spiral positioned on one of the primary modes.

5.6.2 The Group Transform, μ_{fG}

The group transform corresponds to self-programming, which corresponds to modifying the parameters of the mechanisms that support the causal-flow spaces for each phase. This is how learning would take place. Self-programming is central to ontological revision, which is concerned with expanding and updating the systems knowledge – see the review by N.Foo [Foo95].

5.6.3 Generality of Self-Programming

There are two restrictions on the generality of self-programming. Firstly, the input domain X and output range Y are ultimately dependent on the physical interconnect structure. Secondly, the order of a group, i.e. the number of primary causal-flows, is related to the number of unique element functions, which are each bound to a physical realisation or the total number of useful modes supported by the medium.

The group G could be defined by one "pseudo-implicit" group table provided the original operators for each transform in the phase formed disjoint subgroups, i.e. equivalence classes. Self-programming would then amount to two possibilities, modifying entries in the group table, or modifying parameters to group elements. Changing the apparent order of G is accomplished by either initially including elements with an idle causal-flow in the table, or parametrically – see the Iconic Mood Detection example below, cf. Quantum numbers in P.W.Atkins [Atk83]. These elements are claimed when required. Repartitioning G into different subgroups is more involved. It requires modifying the state confluence of the appropriate subgroups and input operators.

Hence, self-programming ranges from gradual parameter changes to extreme arbitrary group transformations. The former is likely to be the more biologically plausible scenario. This would imply that the gross structure and number of disposition primary causal-flows was pre-set or mechanistically constrained following evolutionary requirements on modelling reality. This raises a question. Suppose a qualia space is defined on an aspect of a group. Would the quality space eventually expand to fill all the operators and mappings of the group? This bears upon Dennett's example of professional wine-tasters whose powers of taste discrimination develops with experience.

5.6.4 Implementing μ_{fG}

The group elements demark a particular causal-flow state confluence. This may have a compact functional representation that is more meaningful that an arbitrary map specified via sets of tuples. The group element functions and their parameter constants have to be specified. Due to hardware constraints either there will be a parameterised set of element functions to choose from, or the permissible functions will be part of the structure's specification. Grouping elements into subgroups and catering for knowledge domain inputs is also a requirement. Are there knowledge domain specific element function primitives? For self-programming these features would have to be parametrically accessible.

One issue that needs to be addressed is how self-programming depends on the knowledge domain and context switching. For example, does a single learning mechanism control a number of domains, or is each domain self-controlled? Are different causal-flow spaces switched in depending on context, or are they merely deformed? The question is, how general does self-programming need to be? Are there structures that at one instant may be supporting visual dispositions, and a short time later supporting acoustic dispositions, or does a support always carry a certain modality? The latter would simplify the demands placed on self-programming.

5.7 The Production Phase

This phase is determined by the transform μ_{Yi}. The influence of the output phase depends on the supports physical connectivity. There are two possibilities:

1) Duplicating one output to children.
2) n different output segments sent to n children.

Both cases can be handled by the support mechanism when the output is mapped to the connections, although this needs to be notationally specified.

5.8 Constraints of the Support Structure

As well as providing the support for the disposition causal-flow spaces and their activity, the support structure has two other functions:

1) Providing the interconnect structure.
2) Applying fast, sub-dispositional level constraints to disposition operation.

The first function requires a network style specification language to handle general interconnects. However, it may be possible to abstract and separate out the interconnect requirements so that systems can be denoted by a more conventional hierarchical modular notation. The second function is expressed via the operators: H_{fX}, H_{fM}, and H_{fY}. These act as pre and post filter like transformations in a particular phase, e.g. pattern detectors for masking out irrelevant inputs in the reduction phase, and normalising constraints that are applied across a set of instances of dissociated dispositions or mechanisms in the production phase. To specify this succinctly requires the arithmetic and flow control features of languages such as Miranda or C.

6 Operationally Modelling Dissociated Dispositions

This section briefly considers issues concerned with the practicalities of modelling the operation of instances of dissociated dispositions. This is in preparation for the applied example that follows. In this study, a wave mechanics oriented modelling approach for instances of dissociated dispositions was followed – see Pain [Pain93].

However, alternative approaches may be possible provided they capture the requirements mentioned above and below.

Determining an approach by which to model the operation of instances of dissociated dispositions is constrained in essentially two ways: firstly, by the implications of upholding the Requirement for Continuous Coding Constraint (R-CCC), and secondly, by the need for self-programming. The first constraint has two manifestations: a supervenient (statistical) mode of operation coupled with a form of wave mechanics. The second constraint further refines the class of applicable mechanisms. A useful analogy, which captures these considerations, is to treat instances of dissociated dispositions as interacting noisy waves. Self-programming amounts to partial self-determination during wave interference. With this in mind, and the points made in the previous sections, the following requirements for disposition programming can be discerned:

1) **Wave nature of evolution of instances of dissociated dispositions**. From what has been said, instances of dissociated dispositions are supported by a medium that has field like properties, e.g. a densely interconnected neural net. The disposition instances correspond to patterns of activity across the state evolution of this medium. Without loss of generality, they can be described by a time and state dependent wave function, e.g. $\psi_D(x, t) = \psi(x)e^{-iEt/h}$ – see Atkins [Atk83].
2) **Wave propagation**. Dissociated dispositions are characterised by two types of component waves: a) standing waves corresponding to dynamic fixed-points, and b) evolving waves reflecting disposition evolution. Constancy in stationary circumstances implies a kind of stochastic standing wave that is physically localised.
3) **Wave interactions**. Continuous coding suggests linear superposition as an interaction mechanism. It also implies the receiving process obeys this mechanism. This requirement on the receiver acts as a bottom up reason for there being a distinct dispositional level. That is, superposition implies the receiver state is identified with a dynamic mechanism rather than with a static set of parameters at the lower level – cf. Connectionism.
4) **Wave influence**. The influence of a wave can vary over two dimensions: a) phase shift, and b) amplitude. This can be summed up as the relative degree of wave overlap. A wave may influence an arbitrary number of other waves.

5) **Adaptation**. If the waves are characterised by a supporting basis, then adaptation may require a changing basis.

How does the wave basis determine causal-flow structure? How does the group structure determine evolution? These questions are explored shortly.

6.1 Macro & Micro Dynamics

It is difficult to think of complex causal-flows. However, for all the cases encountered during the investigations, they can be built up compositionally from simpler causal-flows. The terms *macro* and *micro* are used to refer to this compositionality. A macro-dissociated disposition is composed of sub dispositions, whereas micro dispositions are not – Fetzer would call these molecular and atomic dispositions, respectively. This section explores the relationship between these two types of dissociated dispositions.

The continuous coding constraint, as well as suggesting superposition, places some considerations on the manner in which instances of the dissociated dispositions are distributed in the support medium. Firstly, a disposition instance extends over the set of primitives, which make up the medium, in a fuzzy-wave like fashion. One possibility is that a dissociated disposition instance corresponds to one such set, the other is that it is the average of a causal-flow taken over a region of atomic dispositions – see Figure 15.

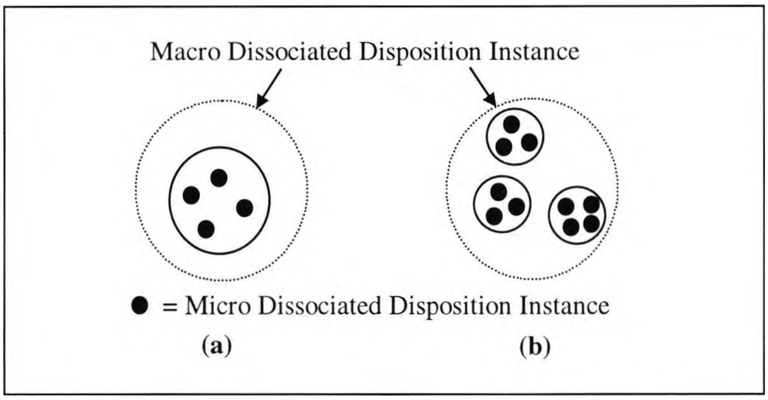


Figure 15. Macro & Micro dissociated disposition instances.
a) The macro disposition instance is defined on a set of micro disposition instances which may be either all the same or different. b) Similarly, a macro disposition instance defined on a collection of sets.

To elaborate, according to wave mechanics the evolution of the system can be modelled by an operation, **U**, on the wave function: $\psi' = \mathbf{U}\,\psi$. Therefore, a region of micro disposition instances can be modelled as a set of topologically coupled equations:

$$\begin{aligned}\Delta d_{-i} &= U_{-i}\Psi(x_{-i}),\\ &\vdots\\ \Delta d_0 &= U_0\Psi(x_0),\\ &\vdots\\ \Delta d_i &= U_i\Psi(x_i).\end{aligned}$$

Notice that the set may be mutually constrained by the superposition requirement. In one scenario, to determine the instantaneous value of a disposition instance would involve taking the expectation value over an appropriate region of the medium via a suitable integral approximation ($dD = \langle\{d_i | i \in R\}\rangle$). It would be interesting to derive a differential calculus for the evolution of dissociated disposition instances.

The causal-flow of a dissociated disposition instance extends in time – it has duration. The flow may, or may not be periodic. The propensity of the disposition can be defined as the inclination of its instantaneous activity. Thus, the activity of a disposition instance will vary according to either a) its purpose, which is its normal propensity when in an identity context, or b) parametric influences from impinging disposition instances. Both these are under the continuous coding constraint. This leads to the distinction between the instantaneous state of the system (a static, instantiable structure), and the instantaneous causal-flow of disposition instance (a differential infinitesimal).

6.2 Self-Programming

The evolution operator, **U**, can be parameterised in terms of state variables and function determining variables. Self-programming then amounts to the ability to modify the function determining variables via the normal inputs to the operator. In general the parameter space can be transformed into a partitioned space reflecting the nature of the parameters, i.e. the tuple of parameters (**w x y z**), where **w** corresponds to a vector of parameters that determine the operator function, **x** is a vector of 'conventional' inputs, **z** is the 'conventional' state vector, and **y** the vector of outputs. Representing the operator as a matrix the system evolution can be expressed as a matrix equation:

$$(w\ x\ y\ z)\ U = (w'\ x\ y'\ z'),$$

where the dashed variables represent the updated vector. This should be compared with the general dissociated disposition instance evolution relation.

Ultimately, the extent to which a dissociated disposition instance can be self-programmed is limited by the need for the underlying mechanism to have some fixed structure. For example, a Universal Turing Machine requires a programme wherein the programme symbols are fixed *a priori* – cf. innate structure. Therefore, the parameter **w** becomes a function selector from the class of functions supported by the medium. For reasons that will become clear shortly, **w** is referred to as the group selector. Consequently, the specification of U involves a function determining structural specification, and a disposition evolution specification, which corresponds to a specific instance of parameters over the structural specification. As will be seen in the following example, such specifications are awkward to deal with directly at the network level. A disposition programming language may alleviate this difficulty.

7 Dissociated Disposition Oriented Example

This section presents an applied example to illustrate a dissociated disposition oriented programming approach. Of note in this section is the generic Gaussian alterator. This embodies the fundamental mechanism by which the dispositional form of a dissociated disposition is coded.

7.1 Continuous Semiotics

From the perspective of the phenomenal level, it appears as though some facets of mind consists of, and evolves in terms of, discrete signs. However, this is a consequence of the entity categorisation process occurring at the dispositional level, which effectively quantises the phenomenal level. The dispositional level was characterised as the ongoing interaction of causal-flows. While the requirement for continuous coding leads to continuous dynamics, the need to categorise entities occurring in reality leads to distinct causal structures and flows. In particular the object disposition, in conjunction with the system structure necessary for signifying reality, gives rise to the experience of objectness in the category structure at the phenomenal level. Thus, discrete signs arise from a continuous process, and the term "continuous semiotics" will be used to emphasise this operational basis of semiotics.

When viewed in terms of signs, the phenomenal level changes state smoothly rather than from one discrete sign to the next – see Figure 16. Hence, the influence of one state dies out as another takes over.

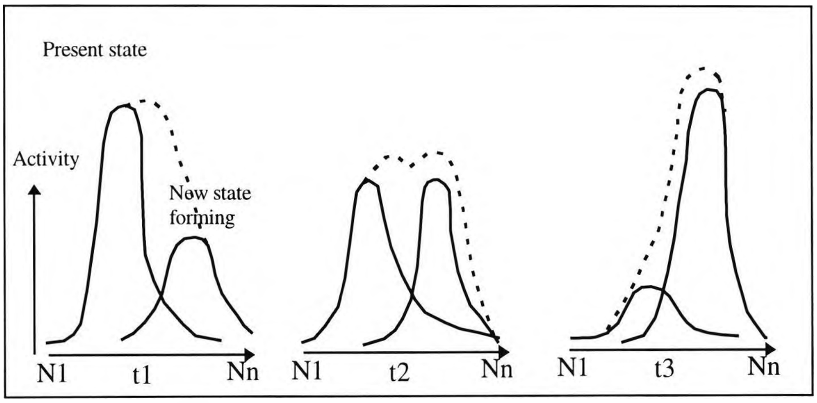


Figure 16. Continuous semiotic changes portrayed as wave superposition.

A state is recognised purely by its distinction from the background ebb and flow. To use an analogy from quantum mechanics, electrons exist as electromagnetic waves. However, when they are observed, or their behaviour characterised at the particle level, the wave is said to collapse to a single state. Similarly, cognition may proceed in a wave like manner. It is only when a definite sign is needed, e.g. for object categorisation, talking, etc., that a fuse of activity is mapped to a series of objects or words.

7.2 Supervenient Statistics & Determinism

The following deals with statistics measured over the interval 0 to 1. Sometimes a statistic might be a probability – see A.M.Mood [Mood74]. In other cases it might be the confidence in a belief, or a type of fuzzy measure – see G.J.Klir [Klir88] and B.Kosko [Kosk92]. Stressing the difference between a probability measure and the latter two: the referent of a probability may or may not be satisfied at a particular moment while the measure's value does not change. For example, the probability might be based on the frequency of the referent being satisfied, e.g. the frequency of success in drinking water from a fountain based upon a mental model that includes a water fountain located at a specific place, etc. The confidence and fuzzy measures express the degree of uncertainty in an observation due to a lack of knowledge (cf. Rough-sets) or because a number of alternatives can legitimately coexist at that time – see Zdzislaw Pawlak [Paw94], Hung Nguyen and Elbert Walker [Ngy96]. The illustrative continuous semiotics example presented below deals with confidence measures.

Statistics are by definition superveniently efficacious. To borrow an example from William Seager [Seag91, p170], "the efficacy of economic properties, such as 'having and inflation rate of 10 percent' is supervenient efficacy." This point is highlighted for two reasons. Firstly, continuous semiotics deals with statistics, and these are relatively independent of the physical nature of the subvening processes. Seager also notes that "there can be both 'bottom-up' and 'top-down' causation between entities residing at different levels of description." Secondly, by appealing to statistical measures, solutions have to be designed in a specific way. However, the reward is inherent noise tolerance and generalising abilities, but for some problems these same properties are a hindrance.

Continuous semiotics is strongly deterministic, even though it is based on statistical pseudo confidence measures. This is in contrast to Boltzmann machines (see Terence Sejnowski [Sej86c]), or Probabilistic Logic Nodes (see Aleksander [Alk87b]), which are probabilistic and hence only weakly deterministic. Similarly, it is unlike stochastic computing (see B.R.Gaines [Gai87b, Gai78]), which is moderately deterministic, or non-deterministic automata in general. The version of continuous semiotics explored here resembles a simplified variation on deterministic Bayesian networks (see J.Pearl [Pear88] and S-S.Chen [Chn87]), but more in the flavour of the probabilistic nets proposed by D.F.Specht [Spe90]. Incidentally, because feedback in continuous semiotics is controlled by design, full wave mechanical treatments are seldom needed, such as those given by M.S.Cohen and W.H.Julian [Cos87].

7.3 Continuos Semiotic Approach to Iconic Mood Detection

The problem is to detect the mood revealed by the image of a simple face made up from just under a dozen line sections. The mood detector is iconic in the sense that the facial features are coded by a set of dispositions whose strength is proportional to the angular shape of the feature – see below. This is the simplest possible kind of iconic resemblance. Typically, to qualify as an iconic resemblance would involve signifying the type of the relation of resemblance itself, i.e. it's qualitative content. In this first example implementation, a continuous semiotics approach is taken from the implementation level – so this is not a true semiotic process since it is not part of a three-levelled system wide triadic semiotic process. The second example builds on this and considers the dynamics of the dispositional level. Again, only by subsequently integrating this within the systems signified reality with the necessary semiotic relations would a semiotic process be produced.

It is assumed that the hard work of detecting a line and matching it with its correct feature has already been done. Hence, the network is presented with a vector of eleven feature values corresponding to the angle the line section makes with the vertical for the respective feature – see Figure 17. The net comprises three layers. A feature layer of eleven nodes, which in this case simply holds the current line section angles coded as a frequency from $0(=0^{o})$ to $127\ (=360^{o})$. An expression layer of five nodes, which is fully connected to the feature layer. The expression nodes respond most vigorously to mellow, happy, angry, sad and squint expressions, respectively. Finally, an emotion layer of six nodes, which are selectively connected to nodes in the expression and emotion layers. The first four nodes in this layer detect dynamic expressions such as a laugh, frown, cry and wink. The other two nodes detect cyclic dynamic expressions such as crying and winking. In addition, connections are modelled as though the net was distributed across an asynchronous communications network. Different facial expressions are presented to the net by varying the feature angles smoothly from one expression to the next in real time.

The network specification language NeSeL, was used as a convenient programming language for modelling this problem – see D.Whobrey [Who89b]. As an implementable language it supplies an operational semantics to network specifications. A complete specification of this net in NeSeL is given in the appendix.

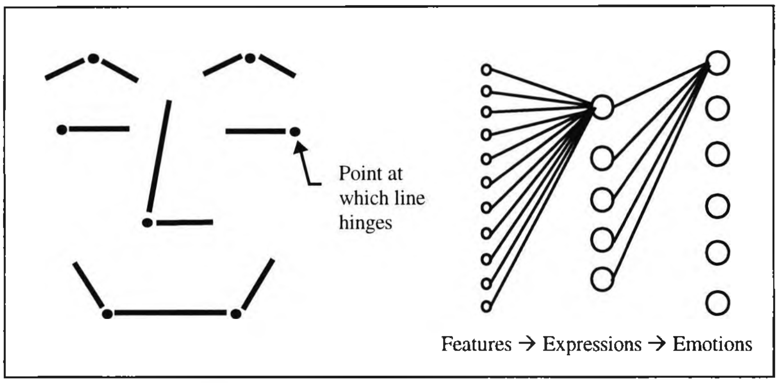


Figure 17. Mr Moody. A simple line section representation of a face.

Consider first an expression node. Its feature inputs (x_i, $n=11$) are coded as frequencies in the range $\{0, R\}$. The memories m_i hold the ideal frequencies for the expression being detected. Here are the node equations:

$$\text{confidence} = \prod_{i}^{n} \varphi_f\left(1 - \frac{|x_i - m_i|}{R}\right),$$

$$\text{output} = 1 + (R-1) \times \text{confidence}.$$

As an input frequency approaches the ideal the argument to the sigmoidal function $\varphi_f(\ldots)$ approaches one. The parameters to $\varphi_f(\ldots)$ are chosen so that it pulls arguments close to one even closer, and quenches all others. Looking upon its output as a confidence measure, an overall confidence measure is formed by multiplying these individual measures together. This approaches one as the input approaches the memorised expression. Finally, this is converted to an output frequency.

Emotion nodes have to detect dynamic changes in the expression and emotion output frequencies. In this example, they simply look for a net frequency change within a set time period of a specified magnitude and direction. The magnitude change must be within set limits. A Gaussian function is used to detect magnitudes within this range. Here are the node equations:

$$s_i(t+1) = \alpha s_i(t) + \beta(x_i(t) - x_i(t-1)), \qquad \frac{ds_i}{dt} = C\frac{dx_i}{dt},$$

$$\text{confidence} = \prod_i^n \varphi_f(\varphi_g(s_i(t))),$$

$$\text{output} = 1 + (\text{R}-1) \times \text{confidence},$$

where α is a decay constant, β a gradient, and $s_i(t)$ is the current net frequency change, which is bounded to within [0,1]. The output of the Gaussian function $\varphi_g(\ldots)$ is normalised to be in [0,1], responding most vigorously when the net change is close to the ideal:

$$\varphi_g(x) = e^{-\frac{1}{2}(\frac{x-m}{\sigma})^2}, \qquad \varphi_f(x) = \frac{1}{1+e^{-k(c+x)}}.$$

The standard deviation σ is related to the mean by $\sigma=\phi m$, where the value of ϕ and m depends on the communications delay – see Table 3. The five basic expressions used to test the net are show in Figure 18.

Parameter	No Comms	With Comms
α	0.80	0.90
ϕ	0.50	0.70
m	0.75	0.75
α'	0.93	0.95
ϕ'	0.90	0.90
m'	0.85	0.85
k_x	100	
c_x	-0.08	
k_m	25	
c_m	-0.70	
Δt	9	14

Table 3. Parameters for Face Mood Detector net.

The primed parameters refer to the cyclic dynamic emotion detectors. The x and m subscripts refer to expression and emotion nodes respectively. The Δt figures are the average number of simulation cycles between expressions. The *With Comms* column is for connections with a relative path length normally distributed in {1,10}.

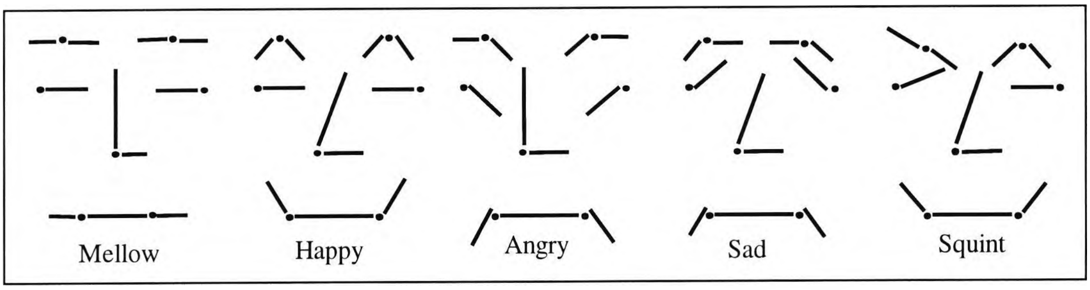


Figure 18. The five expressions used to test the Face Mood Detector net.
Inputs to the net were smoothly varied between these prototypes in real time.

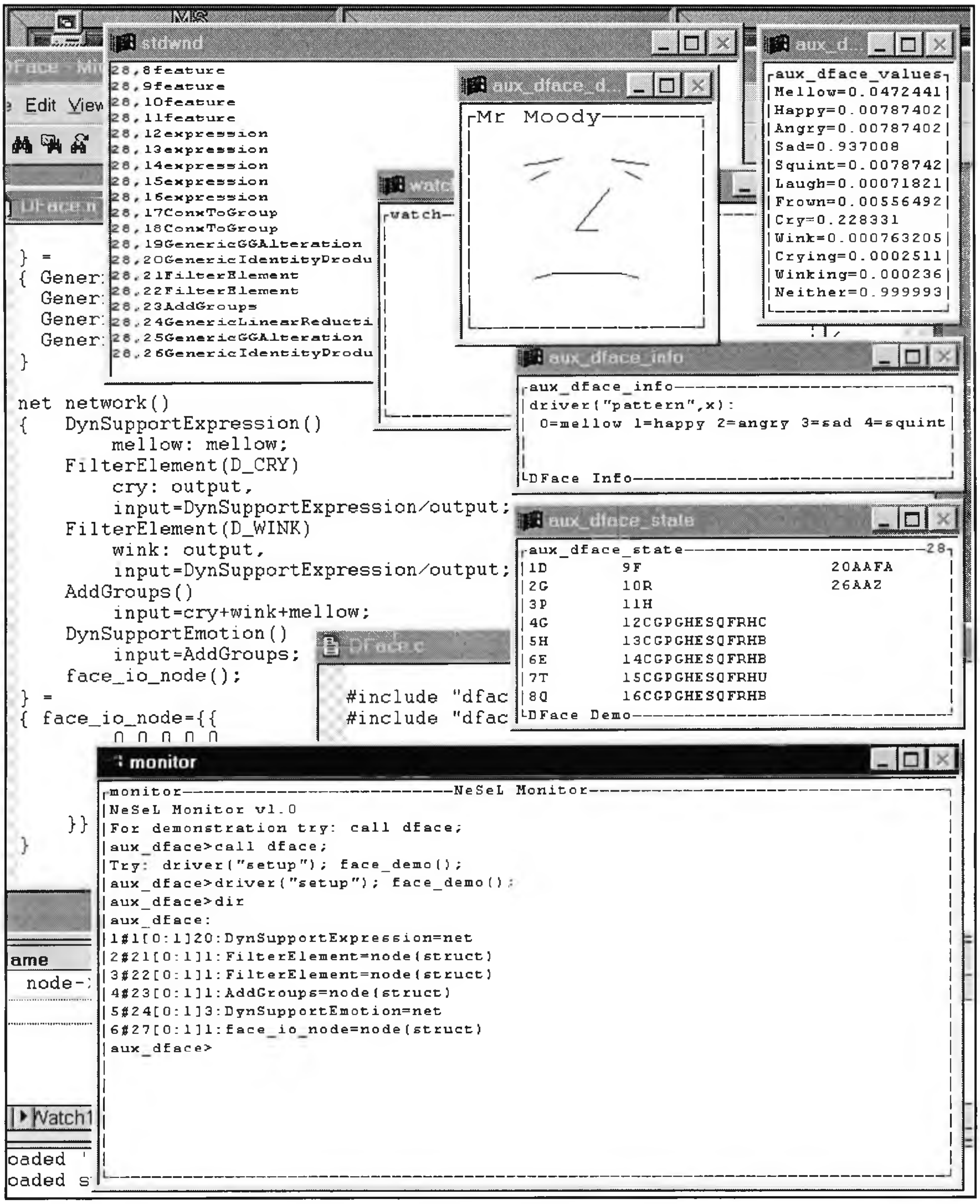
Drawing some conclusions from this net, notice how its output varies smoothly with the input. It does not jump from detecting a crying state to a winking one. Instead, the confidence in the first dies away as confidence in the other rises. Adding varying asynchronous communication delays only required some fine-tuning of the decay and deviation parameters. As the delay went up, from 9 to 14 inter expression cycles, so the rate of decay had to be decreased to compensate. By working with statistics of the inputs, noise and asynchronous input arrival have been absorbed naturally as a consequence of the continuous semiotics design approach rather than being explicitly catered for.

7.4 Dispositional Approach to Iconic Mood Detection

To progress these investigations into dispositional modelling it helps to consider the Iconic Mood Detector reengineered in a disposition oriented manner. To sum up, the task is to identify certain expressions and emotions based on a set of dynamically changing features. The goal here is to understand the nature of the evolution of dissociated disposition instances, so the detector is presented with an input vector of feature angles, rather than an image. For the sake of illustration, the following four categories of moods are artificially delineated:

1) Features, f = {set of twelve angles corresponding to line segments on a face},
2) Expressions, g = {set of five static expressions: mellow, happy, angry, sad, squint},
3) Static emotions, W^1 = {set of four static emotions: laugh, frown, cry, wink},
4) Dynamic emotions, W^2 = {set of three dynamic emotions: crying, winking, neither}.

The terminology is intended for illustration reasons only, so talk of disposition causal-flows as in, "the causal structure of laughter", are just helpful labels. In reality, a more complex network of disposition instances would be required to capture the causal structure of laughter. The goal of this exercise is to demonstrate the mechanism of dissociated disposition instance evolution. The next step would be to devise a disposition language and so enable realistic models to be constructed and investigated. Therefore, in what follows, it helps to replace references to expressions (etc.) by neutral labels, such as L1, L2. At no point is it implied that a particular group element or basis function is equivalent to an expression or emotion. As stated above, a more complex network of micro disposition instances would be required, and this mechanism would have to be integrated into a system wide, triadic semiotic process to be considered a true semiotic process – as explained in the next chapter.



```
28,8feature
28,9feature
28,10feature
28,11feature
28,12expression
28,13expression
28,14expression
28,15expression
28,16expression
28,17ConxToGroup
28,18ConxToGroup
28,19GenericGGAlteration
28,20GenericIdentityDrodu
28,21FilterElement
28,22FilterElement
28,23AddGroups
28,24GenericLinearReducti
28,25GenericGGAlteration
28,26GenericIdentityDrodu
```

```
} =
{ Gener:
  Gener:
  Gener:
  Gener:
}

net network()
{   DynSupportExpression()
        mellow: mellow;
    FilterElement(D_CRY)
        cry: output,
        input=DynSupportExpression/output;
    FilterElement(D_WINK)
        wink: output,
        input=DynSupportExpression/output;
    AddGroups()
        input=cry+wink+mellow;
    DynSupportEmotion()
        input=AddGroups;
    face_io_node();
} =
{ face_io_node={{
        0 0 0 0 0

    }}
}
```

```
#include "dfac
#include "dfac
```

```
aux_dface_values
Mellow=0.0472441
Happy=0.00787402
Angry=0.00787402
Sad=0.937008
Squint=0.0078742
Laugh=0.00071821
Frown=0.00556492
Cry=0.228331
Wink=0.000763205
Crying=0.0002511
Winking=0.000236
Neither=0.999993
```

```
aux_dface_info
driver("pattern",x):
 0=mellow 1=happy 2=angry 3=sad 4=squint


DFace Info
```

```
aux_dface_state                              28
1D        9F                       20AAFA
2G        10R                      26AAZ
3P        11H
4G        12CGPGHESQFRHC
5H        13CGPGHESQFRHB
6E        14CGPGHESQFRHB
7T        15CGPGHESQFRHU
8Q        16CGPGHESQFRHB
DFace Demo
```

```
monitor--------------------------NeSeL Monitor
NeSeL Monitor v1.0
For demonstration try: call dface;
aux_dface>call dface;
Try: driver("setup"); face_demo();
aux_dface>driver("setup"); face_demo();
aux_dface>dir
aux_dface:
1#1[0:1]20:DynSupportExpression=net
2#21[0:1]1:FilterElement=node(struct)
3#22[0:1]1:FilterElement=node(struct)
4#23[0:1]1:AddGroups=node(struct)
5#24[0:1]3:DynSupportEmotion=net
6#27[0:1]1:face_io_node=node(struct)
aux_dface>
```

```
ame
node-
```

```
oaded
oaded s
```

Figure 19. Screen shot of NeSeL environment & the Iconic Face Mood Detector.

7.4.1 Reformulated Iconic Face Mood Detector

The aim is to reformulate the Iconic Face Mood Detector network in terms of instances of dissociated disposition reduction-alteration-production mechanisms. One way to do this is to treat features as inputs to an expression disposition support mechanism, which gives rise to instances of dispositions for the static emotions. These then impinge on a second mechanism that supports disposition instances for the dynamic emotions. The steps involved are now explored in more detail. The mechanisms where implemented using NeSeL and simulated as though on top of an underlying noisy communications network – see Figure 19 and the appendix for details. Satisfactory performance results where obtained as for the first implementation of the example. The main phases in the two disposition instance support mechanisms are now discussed in more detail.

7.4.2 Expression Reduction

The task here is to reduce feature inputs into a continuous fuzzy group representation for the expression category. The mechanism employed closely follows that described in the first implementation of the example.

7.4.3 Expression Alteration

A requirement for the alteration phase is that the character of the alteration has a group structure. It is the group structure that determines the direction of the state alteration, and the support basis that determines the nature of the change, which is problem dependent. A simple vector basis is used for the support. Groups defined by Gaussian function elements are used for describing the interaction of expressions and emotions.

The necessary group actions for the alteration phase can be determined via the following method. The cells in the pseudo group action table can be determined by considering the consequence of the requisite interaction. Labelling the expressions as group elements $g_i \in g$ (i.e. mellow, happy etc. as described above), and the emotions by group elements $W^1_j \in W^1$ (laugh, frown etc.). Then the group action of g_i on W^1_j results in the group element W^1_k . For example, (angry × laugh → frown). Thus, a simplified version of the group action table is as follows:

		W^1_1 laugh	W^1_2 frown	W^1_3 Cry	W^1_4 wink
g_1	Mellow	–	–	–	+
g_2	Happy	+	–	–	–
g_3	Angry	–	+	–	–
g_4	Sad	–	–	+	–
g_5	Squint	–	–	–	+

The plus and minus symbols in the table cells are a further simplification. They indicate that a group action either strengthens or weakens the current state. This is a requirement from the continuous coding constraint, otherwise the system could jump between states. Notice that the cell matrix is almost diagonal and orthogonal. Although this isn't a requirement, it can make implementation easier. In this case, the order of the g_i is arbitrary, so the table can be rearranged to be near diagonal. The final observation is that the table can be made square since the continuous coding requirement also implies that the cardinality of the set of g_i is an approximation to a continuum.

7.4.4 Generic Gaussian Alterator

This section presents an evolution equation for instances of dissociated dispositions that embodies the form of the disposition. This is a central aspect of any system based on the instantiation of dissociated dispositions.

In the Iconic Face Mood Detector a generic alteration mechanism can be devised for group action tables that can be expressed in diagonal form. This mechanism can be parametrically adjusted to accommodate groups of arbitrary cardinality. The parameters themselves can even be set through self-programming. The mechanism is based on using a Gaussian pulse function to selectively strengthen the appropriate state group elements, and weaken the influence of others. It is defined as follows:

$$\text{pulse Gaussian } \varphi_p(a,b) = e^{-\frac{1}{2}\left(\frac{a-b}{\sigma}\right)^2}, \text{ normal Gaussian } \varphi_g(x) = e^{-\frac{1}{2}\left(\frac{a-m}{\sigma}\right)^2}.$$

The first parameter is a function of the cell row, and the extra parameter is a function of the cell column. Thus, the evolution equation for the fuzzy input group $g_f = \{g_i \mid i = 1..n\}$ and fuzzy state group $W_f^1 = \{W_j^1 \mid j = 1..p\}$ is:

$$W_j^1(t+1) = \varphi_f(\alpha W_j^1(t) + \beta \tfrac{1}{N} \sum_{i=1}^{n} g_i(t) \varphi_p(\tfrac{i-1}{n-1}, \tfrac{j-1}{p-1})),$$

where α is a decay rate constant, and β is a state influence rate constant. $\varphi_f(x)$ is a sigmoidal boundary limiting function as above. The term $\frac{1}{N}$ is a normalising factor:

$$N = \sum_{j=1}^{p} \sum_{i=1}^{n} g_i(t) \varphi_p(\tfrac{i-1}{n-1}, \tfrac{j-1}{p-1}).$$

This is the discrete approximation to the continuous evolution equation:

$$\Delta W(y) = \alpha(y) \tfrac{1}{N} \int_0^1 g(y) \varphi_p(x,y) dx, \quad N = \int_0^1 \alpha(y) \int_0^1 g(y) \varphi_p(x,y) dx dy.$$

Parameter values where as follows: for the sigmoidal function $c = -0.7$, $g = 12.0$. The rate constants where within the intervals $\alpha \in [0.8,1.0]$ and $\beta \in [0.9,0.95]$ depending on the extent of network communications. The most critical parameter was the Gaussian pulse width, which is set by choosing an appropriate value for σ. This was determined analytically by solving the pulse function, such that the pulse width at half height is proportional to 1/p, which gives, $\sigma = 1/(p2\sqrt{2\ln 2})$.

7.4.5 Expression Production

In this case the production phase performed a simple Identity mapping of the alteration's fuzzy state group state. Normally the alteration's input would be mapped to a set of output patterns. Notice that the superposition and continuous coding restrictions still apply.

7.4.6 Emotion: Reduction, Alteration, Production

The task for the dynamic emotion disposition instances is to react to longer-term emotions. These emotions were contrived in order to help illustrate and explore the dispositional mechanism. Apart from a couple of auxiliary functions for filtering and adding groups, the mechanism was very similar to the expression disposition instance support mechanism. The main difference was the group action table, which was as follows:

		W^2_1 crying	W^2_2 winking	W^2_3 neither
W^2_3	Cry	+	–	–
W^2_4	wink	–	+	–
g_1	Mellow	–	–	+

See Figure 20 for the process diagram. Full details on this example are given in the appendix.

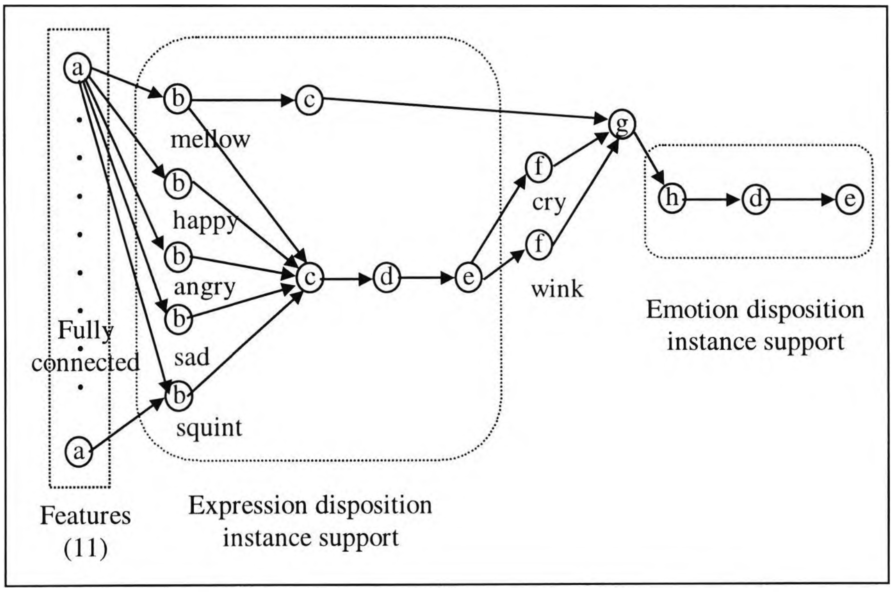


Figure 20. Simplified process network for Iconic Face Mood Detector.

Note this is simulated as though distributed across the communications network of a multiprocessor. Node key: a = feature, b = expression, c = ConxToGroup, d = GenericGGAlteration, e = GenericIdentityProduction, f = FilterElement, g = AddGroups, h = GenericLinearReduction – see NeSeL listing in appendix for details.

8 Operational Modelling Limitations & Issues

In the above example, the generalisation step taken in the design of the alteration phase, to produce the Generic Gaussian Alterator, suggests that eventually a set of alteration primitives could be devised. Such primitives were already becoming apparent in the reduction and production phases. There, operators were required for taking linear combinations of fuzzy group activities (e.g. waves), and operators were required for selecting and filtering elements.

Perhaps the most significant point to make is that in the previous example the approach used to model the dispositions was an indirect method. Wave equations were used to describe the activity of the disposition instances. Evolution was then modelled through an operator. The difficulty lies in modelling changes in group cardinalities. *This amounts to adding terms and state variables when continuous processes are approximated by sampling over a discrete set of points.* Trying to model dissociated disposition instances directly in terms of their observed statistical behaviour suffers from a level of indirectness. The parameters so determined reflect the observed behaviour, rather than the actual state parameters. It would seem that the only alternative is to model the micro processes themselves whose statistical behaviour gives rise to the dissociated disposition instances, i.e. the implementation level. However, it is felt that with further research and insights, the complicating cases in which continuous to discrete transformation problems occur can be isolated and dealt with via suitable operators.

In some ways, the difficulties arising from continuous versus discrete representations and its impact on notational complexity is due to the underlying specification language and target computer architecture. This is true for most conventional computer languages, such as NeSeL, where the underlying symbolic oriented control mechanism even permeates the nature of the abstract machine of the language. A disposition programming language would seek to avert this problem.

8.1 A Convenient Operational Notation for Dissociated Dispositions

From the above discussion, one of the main problems facing a notation for dissociated dispositions is that current notations adopt a singular execution model and a stationary-representation perspective. At worst, this means a virtual machine has to be defined on top of the base notation. An interesting exercise would be to implement a virtual machine and notation in a functional language such as Miranda. Continuuism also affects the semantic interpretation of temporal events and actions.

An important aspect of the disposition notation is the structure-operator division. In one sense, the structure specification is redundant since this may be captured by the operators. For example, the structure is treated as the manifold on which operators parade. A system of homogeneous columns could be assumed, or that all nodes have a small local fan-in and fan-out, or symmetrical connectivity. Placeholders would then be required in each field for each node, and some alignment notation between fields. An alternative is to make nodes homogeneous and train functions in a similar spirit to MLP neural nets. A programming environment based on a diagrammatic portrayal of the structure would be useful.

9 Summary

Notation and principles from group theory were found to be an appropriate means of describing the operation of instances of dissociated dispositions. Operationally, these disposition instances follow a continuous execution model, as opposed to a singular one. This proved to be the major restriction on finding a suitable programming language for dissociated dispositions amongst the applicable contemporary languages.

A system based on the dispositional approach has structural and operational components. This complicates notations. One solution is to hide as much structure as possible through operators and modularization. An instance of a dissociated disposition was shown to be based on a support mechanism comprising three phases: reduction, alteration and production.

A practical example was developed that featured a generic alteration mechanism governing the dispositional form of a dissociated disposition instance. This mechanism is fundamental to any system based on dissociated dispositions since it codes the form of the disposition, i.e. the nature of its influence on fellow dissociated disposition instances. The interaction of dissociated disposition instances can be closely modelled by a combination of wave mechanics and deterministic fuzzy logic based on a many-valued logic, such as that of Lukasiewicz, rather than a probabilistic model. However, wave mechanics was used indirectly to model these dispositions. This runs into notational problems when approximating the statistical properties of continuous processes by discretely sampled ones. A continuous change in the former amounts to adding or removing terms to the latter. It is hoped that further research may lead to a formulation that averts this problem.

Chapter 6

Logical Requirements for Replicating Semiotic Processes

1 Overview

When concern lies with the mathematical form of a system's dynamics rather than properties of the system, such as the learning and emergent capabilities typically attributed to neural nets, then in this mathematical sense, biological minds are a type of numeric system and digital computers a type of symbolic system (the terms *numeric* and *symbolic* are explained later). For many decades these two classes of systems have been investigated, in the case of numeric systems to determine just how mind arises in certain systems, while with symbolic systems the investigations have yet to determine whether they can support mind at all.

The goal of this chapter is to determine whether semiotic processes can be replicated on a digital machine and under what conditions this might be possible. This involves two aspects: that the system operates semiotically, meaning in has the right kind of processes for the right reasons, and establishing an appropriate equivalence between numeric systems (the systems behind minds) and symbolic systems (the systems behind computers) and their ability to support the semiotic processes concerned. This, in conjunction with the results from the prior chapters is used to determine whether the semiotic properties of mind can be replicated under these limited conditions and so help determine whether mild AI can succeed.

1.1 Logical Development

The introduction chapter drew attention to the importance of the ground relation in semiotic processes and highlighted Fetzer's analysis of the analogical argument for strong AI that suggested digital machines lacked such relations when viewed from a computational perspective. The first half of this chapter explores Fetzer's suggestion in some detail where the focus shifts to examining the semantic grounding of semiotic processes and the logical form of the operational structure this in turn implies about the semiotic processes.

The second half analyses and compares the structure of the two classes of system and states under what conditions a symbolic system might be considered weakly equivalent to a numeric system, this would correspond to replication. The primary conditions relate to the ontological levels in the system, the perspective of the observer, and the causal structure of the system. If the semiotic properties of mind can be reproduced in a symbolic system under these conditions, then mild AI is possible. A brief calculation is presented in order to estimate the future performance capacity of digital machines and whether they would have sufficient capacity to support a replication of mind.

2 Grounding Semiotic Processes

Charles Morris [Mor38] distinguished three branches in the field of semiotics: semantics, which concerns the meaning of signs; syntactics, concerning the structural relation between signs; and pragmatics, concerning the ways in which signs are used and interpreted. The previous chapters looked at the operational and structural properties of semiotic processes. In these terms the meaning of a sign was attributed to its dispositional tendencies in the system. This section explores in more detail the relationships that must exist within a system if a process is to be called semiotic. The starting point for this exploration is the semiotic ground relation since it features prominently in Fetzer's analysis of the analogical argument for strong AI.

2.1 The Ground Relation

As mentioned in the introduction chapter, Peirce called a *ground* the nature of the semantic relation between an entity (the dynamical object) and its representation in the system. The ground conveys the *respects* by which the entity is represented to the interpretant. It semantically relates the sign to its entity. Peirce distinguished three types of pure ground in his second trichotomy: iconic, indexical and symbolic – see the introduction chapter. Peirce distinguished three kinds of icons: images, diagrams and metaphors, suggesting an image resembles an entity in terms of simple qualities, and diagrams represent an entity via analogous relations between their respective parts – see Hausman [Hau93, p89]. The face mood detector was iconic since it represented the angles of facial features by the strength of a dispositional tendency.

Often the ground is said to express the meaning of the sign and is equated with its content. However, Daniel Chandler [Chan95] warns that this is problematic since it suggests “that meaning can be ‘extracted’ without an active process of interpretation and that form is not in itself meaningful.” In support of this, chapter four suggested a system wide structure is necessary for generating meaning within a system. Theories of content attempt to provide an explanation for the production of meaning in systems that doesn’t succumb to the interpreter-regress problem and satisfactorily elucidates any semantic primitives – see Fodor [Fod90] and Umberto Eco [Eco76]. In contrast, this thesis is concerned with the nomic form of semiotic processes, from an operational perspective, to the extent that this helps determine whether a mild version of AI is possible.

2.1.1 Exploratory Example: Shape Signification

In the analysis of the ground relation it is helpful to consider a simple example involving the signification of primitive shapes. This entails a semantic process that produces signs whose capabilities reflect the geometric nature of the shape to the mind of the system. Unlike the iconic mood detector example presented in chapter five, which was concerned with the operational nature of the dispositions underlying semiotic processes, the shape example helps in conceptualising the various relations within the system. In addition, according to Fetzer’s semiotic classification of mentality, the system will be a mind of Type I when it has the capacity to utilise icons, which is potentially manifested in its capacity to make a mistake by misidentifying a shape “by virtue of taking a resemblance relation of one kind for a resemblance relation of another” – see Fetzer [Fetz90, p40].

Consequently, in the shape signification example the goal would be to construct a creature that can identify, in a semiotic sense, simple shapes such as a circle, triangle and square. Although, iconic mentality consists in the ability to recognise (specific sorts of) resemblance relations, which is not necessarily restricted to these well-defined shapes. So, for example, an alternate goal would be recognising different instances of colours (aroma, sounds etc.) as instances of the same colour. These shapes would be presented to the creature individually as coloured geometric figures on a white background. The intention is that each shape would invoke in the creature a set of dispositions, e.g. objectness, spatial extension, regional continuousness, straightness, curvyness, symmetry, persistence, colour etc. The following sections discuss the nature of the ground relation bearing this example in mind. In particular, an iconic ground relation is examined with respect to its function and relationship to the other aspects of semiotic processes within the system.

2.2 Peirce's Triadic Sign Relation

Chapter one introduced Peirce's terminology for the aspects of semiotic processes and included Fetzer's version of the sign triad. The figure below shows the common form of Peirce's triad according to his original terminology and a more contemporary version. In what follows Peirce's original terminology will be used. To start with various triad diagrams are presented and then discussed in more detail as the section progresses.

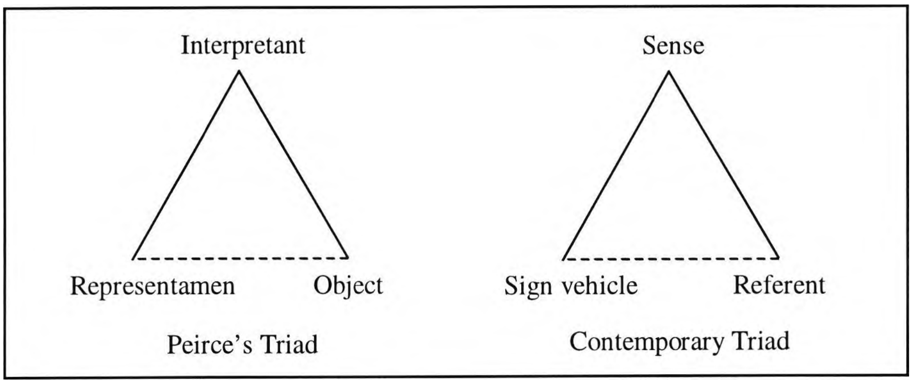


Figure 21. Simplified diagrams of semiotic relationships.
The first shows the key aspects of Peirce's conception of the triadic sign relation – see Eco [Eco76, p59]. The second shows a contemporary version using "more familiar" terms – see Chandler [Chan95].

The above diagrams are a simplification of the sign triad and also lean toward a subjective characterisation stemming from the way signification appears from a first person perspective. For the present purposes, Carl Hausman [Haus93, p71] presents a more useful characterisation, by reading Pierce more closely, that brings the ground relation into the picture:

> "A sign, or *representamen*, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object*. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen" – Peirce (2.228).
>
> A sign or representamen is "connected with three things, the ground, the object, and the interpretant" – Peirce (2.229).

The diagram on the left of the figure below shows how Hausman portrays this as a triadic relation. The diagram on the right is derived from what Peirce called the references of the sign:

> "1st, it is a sign *to* some thought which interprets it; 2nd, it is a sign *for* some object to which in that thought it is equivalent; 3rd, it is a sign, *in* some respect or quality, which brings it into connection with its object" (5.283).

The causation relation corresponds to the first reference, the interpretant corresponds to the second, and the ground the third.

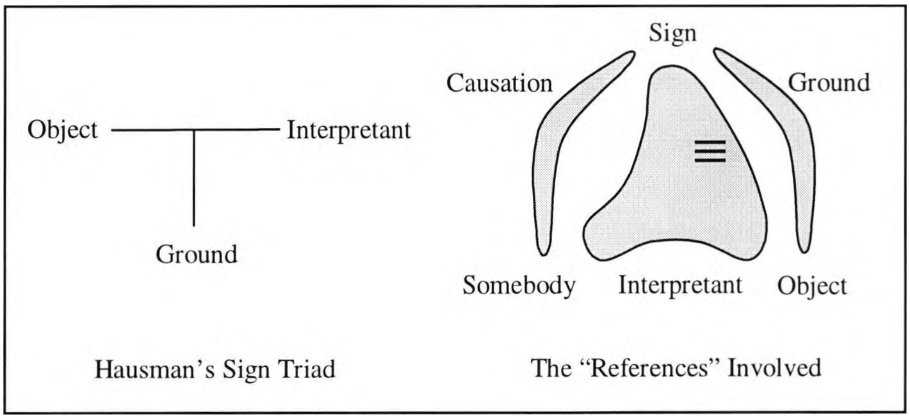


Figure 22. Triadic sign relation showing ground relation.
The first shows the ground version of the triadic sign relation according to Hausman – see [Haus93, p72]. The second shows the references involved according to Peirce – see Peirce 5.283.

To help understand the form of semiotic processes the following sections present augmented diagrams of the sign triad that includes annotations for the types of the relations and the ontological level of the various entities involved.

2.3 Fetzer's Interpretation of the Triadic Sign Relation

This section analyses Fetzer's Human Being and Digital Machine Relations from his analysis of the analogical argument for strong AI introduced in chapter one. This starts by clarifying what is suggested to be problematic about the ground relation in digital machines, and then refines the Human Being Relation. The next section builds on this and examines what is required for ground relations to be instilled in digital machines.

Chapter one presented Fetzer's diagram for the Human Being and Digital Machine Relations – reproduced below. Noticeable in comparing the two relations is the absence of the ground relation in the case of digital machines.

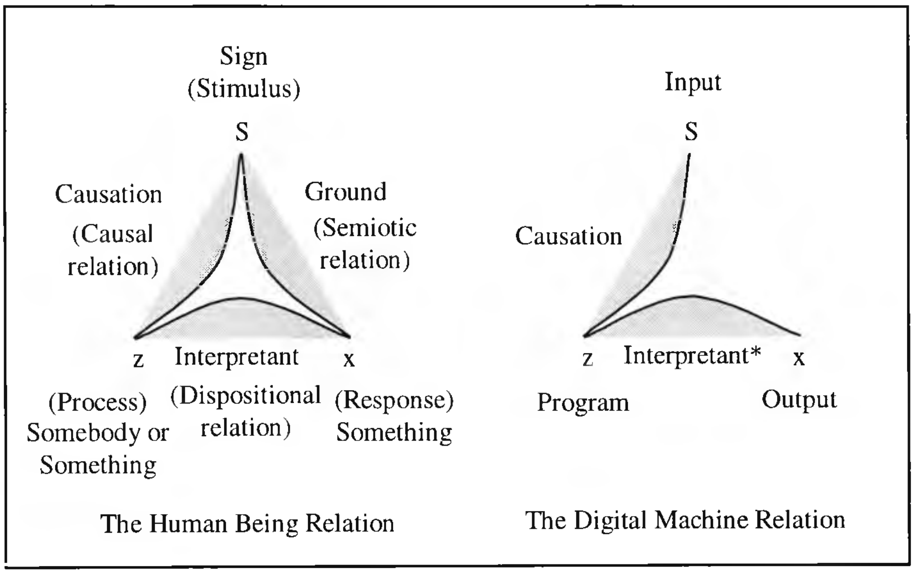


Figure 23. Fetzer's Human Being & Digital Machine Relations – see Fetzer [Fetz90, p277].

To recall, Fetzer [Fetz91, p278] suggests "there may be causal connections between a cause C (call it "stimulus") and an effect E (call it "response"), but unless that causal connection obtains *because* that sign is an icon or an index or a symbol in relation to that effect (…), it cannot be a *semiotic* connection." In other words, two conditions must be met:

1) There has to be a ground relation within the process, necessarily, as an inherent aspect of the operation of the process, e.g., the process must operate in terms of semiotic relations and associated mechanisms, necessarily.
2) The ground relation must have been invoked because such a ground relation was detectable in the environment, e.g. by virtue of a relation of a cause or effect in the case of indexical signs.

The first condition is saying the system must have an identifiable semiotic structure to which it intrinsically owes its proper functioning, in contrast to treating the system as-if it was semiotic. This parallels Dennett's intentional stance and Searle's as-if intentionality – see the section on Mind as Intentionality in chapter two. The second condition suggests signification is a nomic process in that the sign relations produced are implicitly determined by the dynamical laws driving the semiotic process – although this doesn't prevent the system making a mistake by misidentifying something.

Together these conditions highlight that in comparing systems, in terms of their semiotic abilities, the internal structure becomes important – as suggested in chapter four. When compared irrespective of their internal structure all digital machines would appear to lack ground relations. This draws attention to the analogy made in Fetzer's diagram between processes and programs (Premise 3 of the analogical argument). Thus, this opens up one avenue to pursue, namely, the relation between the dispositional structure set up by a digital machine executing a certain program and the dispositional structure underlying semiotic processes. This is based on the premise that a semiotic process is operationally characterised by its causal structure. Before pursing this matter in the next section, the relationships within Fetzer's Human Being Relation are elaborated in terms of Peirce's "References".

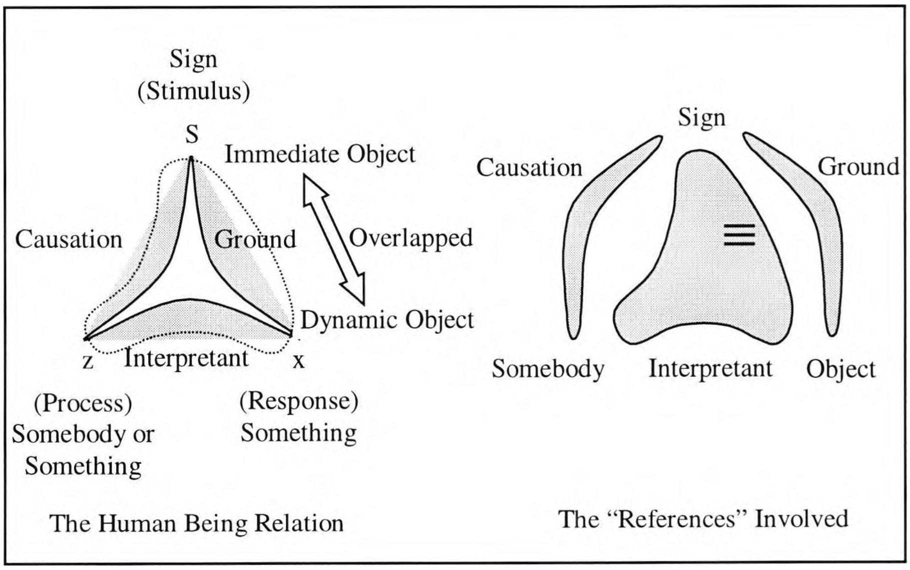


Figure 24. Comparing Fetzer's Human Being Relation with Peirce's "References".

Initially Fetzer's labelling of the object apex in the sign triad, as the "Response", appears problematical in regard to the direction of causation. If the edges of the triangle denote the partial relations involved in the triadic relation, such as the interpretant, and a relational direction is imposed on these edges, then they would be expected to emanate away from the object apex. The intention, Fetzer suggests, is that this apex portrays what the sign stands for – see Fetzer [Fetz91, p277]. When viewed in this way it helps to treat the interpretant relation in the manner of Peirce's "References" where the interpretant effectively makes the dynamical and immediate object appear as equivalent to the sign user.

The Human Being Relation focuses on the abstract relations between the entities as signified at the phenomenal level. In contrast, the Digital Machine Relation simply portrays the causal relations between inputs, programs and outputs, which all can be situated at the same ontological level, in this case the implementation level. Hence, in light of the conditions required for the ground relation, to determine whether digital machines can have ground relations requires determining if the necessary abstract relations can exist when taking into account the ontological levels involved and the causal structure of the system. This is the task for the next section.

2.4 Ontological Level Interpretation of the Triadic Sign Relation

The objective of this section is to supplement Fetzer's Human Being Relation with details on the relations between the ontological levels involved in the hope that this will suggest how a digital machine might be programmed to have semiotic processes.

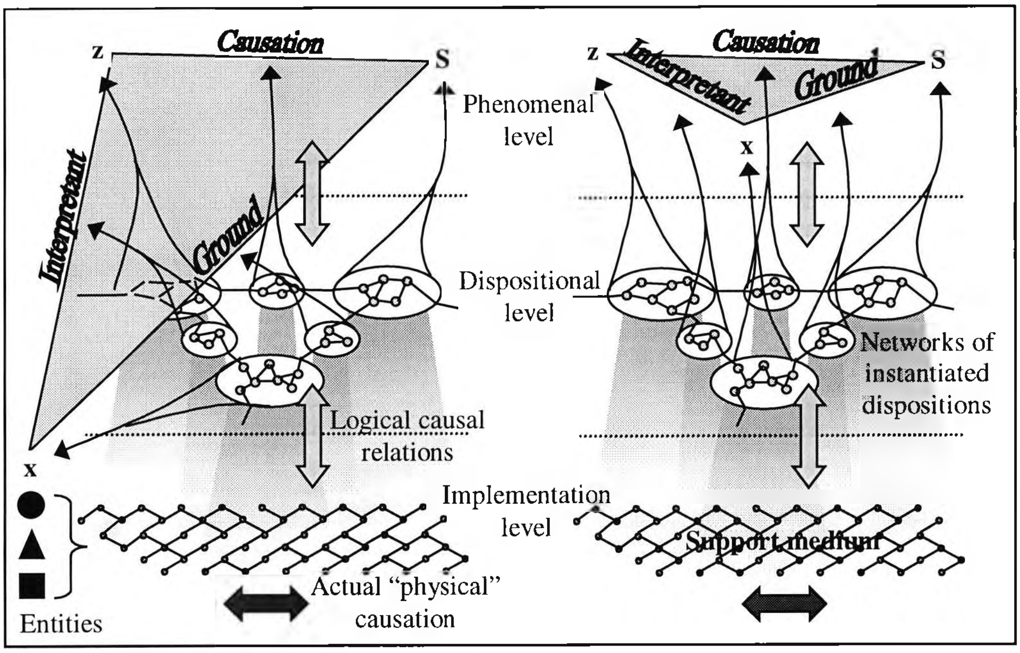


Figure 25. Ontological level interpretation of Fetzer's Human Being Relation.
The dotted lines represent level boundaries. Left diagram has a physical entity as dynamical object, right diagram has a sign (thought) as dynamical object.

The figure above shows the conventional Peircian sign triad overlaid on the three ontological levels of the system. A number of observations can be made.

Perhaps the first observation is the question raised concerning the objective purity of the sign relation, since it can be interpreted from a first or third person perspective. The subjective (first person perspective) view of the triadic relation is instilled upon the sign user by the "presentational" perspective the system imposes on the user's beliefs about the sign and the manner in which they use it. In other words, all mental signs have some properties in common to do with their "signness", such as their "triadicness". In contrast, an objective view could be taken from an operational perspective that emphasises the nature of the sign triad relations across ontological levels. Hence, simply because signs appear to the user to be part of a triadic relation it doesn't follow that the underlying mechanism is triadic. In the diagram, networks of instantiated dissociated dispositions are shown as contributing to the phenomenal experience of the sign's triadicness.

The second observation concerns the difference between the actual causation in the system and the logical causal relations – although this distinction is somewhat arbitrary depending upon what the base level is taken to be. For example, the actual causation in the system arises from the causal powers of the support medium at the implementational level. A logical causal relation is then a logical mapping between two sets of implementational events, one of which is causally dependent on the other. Hence, "Causation" corresponds to a logical causal relation in that it is the occurrent instantiation of the dispositions that produce the sign that influences the sign user. In turn, these instances of dispositions correspond to the causal powers of the underlying patterns of activity in the support medium – see chapter five.

With these two observations in mind, the next point concerns the impact of the relative ontological level on the nature of the relationships in the sign triad. The diagram on the left shows the Peircian sign triad positioned at the phenomenal level and descending to the physical level (here equated with the implementation level) when the dynamical object is a physical object as opposed to a mental entity as in the case of the diagram on the right. This highlights the fact that the three partial relations (causation, interpretant and ground) are intended to be generic relations in the diagram. For instance, the ground relation on the left has an extensional aspect (e.g. it could be an indexical ground), while that on the right is intensional. In the case of the interpretant relation, it is the immediate object that embodies the dispositional influence of the dynamical object on the sign user. The experience of the extension to the dynamical object, if any, is in turn a consequence of a signification process which chapter four suggested involved signs, and therefore instances of dispositions, "portraying" the entity's orientation and location. The notion of "portraying" leads to the position of the sign user in the diagram, which is discussed next.

Positioning the sign user (z) at the phenomenal level has a number of implications. Firstly, there are the instantiated dissociated dispositions at the dispositional level that are influenced by prior dispositional processes underlying the sign processes. Secondly, there is the system wide sign user, the mind, to which consciousness is attributed when it has the ability to use signs, and is not incapacitated from using that ability. Thus, “portraying” implies a host of further signs that produce an experienced “presentational” relation to, in turn, a signification of a sign user – see the discussion on the outline of the hypothetical system structure and the requirement for system structuring for causal accessibility in chapter four. As suggested there, at the dispositional level there is no thing to be “represented” too. In addition, at this level there is only one causal direction, the direction of the instantiated disposition’s influence. This avoids the interpreter regress problem, and highlights why computational explanations for the ground relation fall prey to it. Specifically, the causal direction in the process of representation, at least in the computational sense, is in the wrong direction, i.e. it implies an interpretation of the representation leading to a regress – see the statement on the direction of instantiated dissociated disposition flows in chapter four. From this it follows that neither is the operation of the dispositional level computational in nature, i.e. about producing results. Instead, the occurrent causal structure is of importance. This suggests “signifies” and “causal consequences” are perhaps better terms for talking about the ground relation in contrast to the “representational” and “computational consequences” terminology used by Eckardt – see Eckardt [Ecka93, p296] and Fetzer [Fetz98a].

Examining the ground relation in more detail, consider an image, an icon that resembles its dynamical object in terms of its simple qualities, such as colour. Peirce distinguished between a quality, such as red, and a condition of quality, such as redness, where this pure abstraction is the ground of the embodied quality and is what is experienced – see Hausman [Hau93, p104, p125]. Chapter four suggested that the sensation of colour, as a property of the phenomenal level, is produced by an interacting network of occurrently instantiated dispositions in combination with their orientation and location within the overall causal structure of the system – see Figure 26. To the system, at the dispositional level, colour became an instantiated disposition signifying a mood indicator, such as red signifies a mood common to anger, danger, and warmth etc. Hence, in this case, the ground relation refers to the presence of an occurrent network of specific types of instantiated dispositions situated in a system wide causal network as part of the signification process.

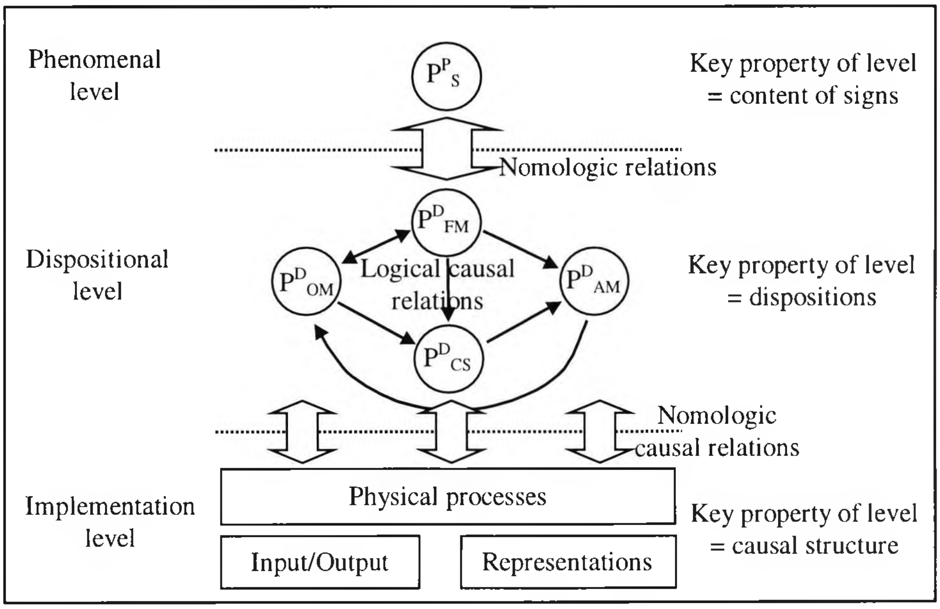


Figure 26. The system wide dispositional structure underlying the ground relation.
P^P_S = property conveying phenomenal content of sign, P^D_{OM} = network of instantiated disposition properties for singularisation mechanism, P^D_{FM} likewise for fixing mechanism, P^D_{CS} category structure, and P^D_{AM} attentional mechanism – see chapter four.

This interpretation of the ground relation conforms to the conditions placed upon the relation by Fetzer as discussed above. Consequently, this suggests the task of instilling semiotic relations into a digital machine reduces to showing how the necessary occurrent dispositional structure can be set up in the machine, e.g. formulating the network of dispositions behind iconic ground relations. This would benefit from a dispositional programming language. To achieve this in theory requires showing under what conditions it might be possible for a symbolic system to replicate the relevant dispositional structure of the idealised numeric system upon which a model of mind might be based. This is the topic of the remaining sections in this chapter.

3 Characterising Conventional Kinds of Systems

This section reviews definitions for formal systems, digital machines, and the two main systems being analysed. Later this will be used to answer such questions as whether the simulation of mind is also replication.

3.1 Conventional Definition of Formal Systems & Models

A *system* is informally said to be some bounded *thing* that has internal structure. For example, a system (of kind K) might be defined as an instantiation of a property (or of an arrangement of properties – see the appendix for a formal discussion along these lines) of kind K, which will bring with it all of its permanent properties, including causal potentialities. A *formal system* is an uninterpreted symbolic system characterised by a set of descriptive sentences in a language, and a set of rules or axioms that govern what deductions are made possible from the sentences. Notice that the definition mentions symbols but it does not mention time – see B.Cohen et al. [Cohe86]. Formal systems are abstract objects. They are related to *real systems* via an interpretation, which relates the formal symbols to their real counterparts. A real system that satisfies the sentences and rules of a formal system is said to be a *model* of the formal system. More than one model may satisfy a formal system. Models need not be real systems, they are often abstract structures.

More formally, a formal system is defined as a tuple: $\mathcal{F} = \langle L, C \rangle$, where L is the complete set of sentences generated by a language, and C is the consequent closure, i.e. the set of operational rules. The formal specification S of a system is often given as a proper subset of sentences drawn from a general specification language $\mathcal{L}$. In addition, a formal specification can be referred to as a theory T, which demarcates a class of models that satisfy the specification – see Machover [Mach96].

A model is denoted by an $\mathcal{L}$– structure, which is defined as: $\mathcal{M} = \langle L, U, R, F \rangle$, where L is the syntactic language, U refers to the underlying domain of discourse, and R and F map symbols in L to the corresponding relational and functional operations in U. In other words, a structure attaches meaning to the symbols. A structure is called an $\mathcal{L}$- interpretation when the domain of discourse is also a structure. For the purposes of reasoning about models and theories, an $\mathcal{L}$- interpretation can be augmented to an $\mathcal{L}$– valuation, which adds a mapping that assigns semantic values to variables. Hence, a theory T is a set of sentences draw from a language $\mathcal{L}$. An $\mathcal{L}$-structure M is said to be a model of T if it provides a consistent valuation for the sentences of T. This is denoted by, $M \models T$ – see W.Hodges [Hodg93] for further details.

3.2 Conventional Definition of Digital Machines

For the present purposes a definition of a digital machine based on a simplified definition of the standard Turing machine will suffice. A Turing machine is conveniently defined in two stages by firstly defining a finite-state machine (FSM). A FSM consists of external input and output alphabets, a set of internal states and a next-state function. A Turing machine consists of a control unit based on a FSM, a given initial state and designated halting states, and a tape alphabet for a read-write tape, where a subset of the FSM states result in operations on the tape, and a subset of the FSM inputs and outputs are read and written as symbols to the tape respectively.

An idealised digital machine, analogous to a conventional computer, is a Turing machine in which the tape is replaced by a finite random access memory. A digital computer is a machine for automating the manipulation of models – see Savage [Svg98] and D.Wood [Wood87]. Turing [Tur50] emphasised that digital computers should be considered as discrete-state machines, but that, "Strictly speaking there are no such machines. Everything really moves continuously. But there are many kinds of machine which can profitably be *thought of* as being discrete-state machines."

3.3 Conventional Definition of Symbolic & Numeric Systems

Newell and Simon [Nwl76] put forth the idea of the physical symbol system, and how such systems are instances of a universal machine that could be equated with a Turing machine. Here, a symbol is defined to be that which can be used to *designate* an arbitrary expression. Further, if the expression designates a process, the system is able to *interpret* the expression and so perform the process. An interpreter manipulates the symbols and, hence, they are not directly causally efficacious. Digital computers can be used to represent, to the user, symbols abstractly as arbitrary patterns of zeros and ones in their memories. Thus, digital computers are often called *symbolic systems.*

Traditionally, the term *dynamic system* was used to refer to natural processes, such as the motion of the planets and electronic circuit behaviour. Consequently, the term *numeric system* was introduced to refer to this narrower class of systems in which *variables* rather than a symbolic configuration define the state of the system. The state variables of a numeric system evolve simultaneously and continuously in time, unlike symbol systems – see Norton [Nort95]. By this is meant that a variable's value changes numerically rather than symbolically. Secondly, the rate of change is normally predetermined by physical factors, for example: what force is being applied, potential differences etc., whereas the rate at which a symbolic system changes state depends solely on the speed of the digital computer.

A numeric system is a model of a dynamic system in which the state is defined over a set of variables whose evolution in time is described by a system of differential equations. However, there may be some dynamic systems that exist but cannot be described in this way. The behaviour of some numeric systems may be approximated by discrete time dynamics and difference equations – for details see Norton [Nort95].

In what follows, the numeric and symbolic systems will be implicitly understood to be dynamic systems i.e. their state is changing in time, as opposed to static systems in which the properties of interest effectively remain constant over the time period considered.

4 Comparing Numeric & Symbolic Systems

The objective now is to highlight the relevant differences and relationships between numeric and symbolic systems.

4.1 Working Definition for Numeric and Symbolic Systems

Consider general definitions for these two types of systems using the terminology developed in the appendix:

Define a numeric system S_N as consisting of two levels l_N^0 and l_N^1. The domain of l_N^1 is a set of elements on which is defined a set of variable properties $P=\{p_i \mid i=1..n_p\}$. Its laws are the axioms A_N^1, and its configuration C_N^1 determines the structure between the elements. For convenience the variables are organised as a tuple, $\langle p_i \mid i=1..n_p \rangle$, $p_i \in [a_i,b_i]$, although their relative order does not matter, and system evolution amounts to property changes modelled as numerical variations. The elements and evolution laws of l_N^1 are constructed from the base level l_N^0.

Define a symbolic system S_S as consisting of two levels l_S^0 and l_S^1. The domain of l_S^1 consists of a set of symbols $S=\{s_i \mid i=1..n_s\}$, its axioms A_S^1 are the manipulation rules, and its configuration C_S^1 determines the structure between the symbols. For the present purposes the symbols are organised as a tuple, $\langle m_i \mid i=1..n_s \rangle$, $m_i \in \{s_j \mid j=1..n_s\}$, and manipulation amounts to permuting the order of the symbols in the tuple. The mechanism for manipulating the symbols resides at the lower base level l_S^0.

The definitions are deliberately similar in order to focus on what is common and distinct, which the following subsections examine in more detail.

4.2 Comparing the Domains & State Spaces

The variables in the example numeric system correspond to functions defined on a transient set of the domain elements. The state space can be represented as a continuous n_p-dimensional hypercube, where state trajectories may lie in the interior of the cube and tend to form continuous paths.

The symbols in the example symbolic system correspond to domain elements. The state space is most easily represented as a tuple of the n_s symbols. Alternatively, the state space can be represented as a polyhedron where the vertices correspond to a permutation of the symbols. State trajectories would be disjoint transitions from and to arbitrary vertices where neither the interior nor the edges of the polyhedron is traversed. A more popular approach involves coding the symbols as numbers and representing the state as a discrete n_s-dimensional hypercube. Again, state trajectories would be disjoint jumps, but this time they can include interior cube points.

While it may be concluded that as the number of symbols increases the state space of a symbolic system approaches a continuous space, and so approximates a numeric state space, this does not carry over to the dynamics of a symbolic system. State space trajectories remain a disjoint series of points rather than forming continuous paths. This is because the continuity of the paths is determined by the nature of the laws driving the dynamics. To sum up, symbolic systems have discrete state spaces where variations are discrete functions, while numeric systems have continuous state spaces where variations are continuous functions – see the discussion on system processes in the appendix.

4.3 Comparing the Nature of the Axioms

For a system to be productive its axioms must be coherent in some sense. In numeric systems the axioms take the form of constraints that are continuously in effect. For symbolic systems the choice of axioms is more arbitrary and they are sequentially applied in the form of discrete actions, e.g. rules such as “If A=B then C.” They operate directly on the domain elements by changing their relative ordering.

Hence, the nature of the axioms determines the continuity or discreteness of a system. Continuous, and therefore numeric systems, are governed by axioms in the form of continuous constraints. Discrete, and therefore symbolic systems, are governed by axioms in the form of conditional actions. However, this distinction is an idealised one. For example, some numeric systems, such as the electrical circuits considered in the appendix, are based on the dynamics of finite objects, which means that state changes are only statistically continuous.

4.4 Comparing Causal Structure & Dynamics

The state space representation does not reveal much about the causal structure of a system. Instead, it is necessary to look at the network of causal dependencies as detailed in the definition of causal structure (D-CS) given in the appendix. For a numeric system the network can be formulated from the property dependencies. Variables refer to properties of processes, which have direct causal power, whereas a symbol is an indexical, i.e. it is a designated constant. Thus, for symbolic systems there is a degree of indirectness. The causal structure is encapsulated by the axioms in the form of the symbol manipulation unit, which was defined as belonging to the base level. Consequently, the symbols have no direct causal power or causal structure. They are merely tested or exchanged by the manipulation unit.

Concerning system dynamics, in one atomic time interval all variables in the numeric system can simultaneously vary, for the symbolic system a permutation of the symbols is performed. This can be represented as: $\delta S_N = \langle \delta p_i \mid i=1..n_p \rangle$, where $p_i(t+\delta t) = p_i(t) \pm \varepsilon_i$, i.e. $\delta p^n_{ijx}(t) \approx \nabla \psi^n_{ijx}(t)dt$, from the definition of a property variation (D-PV) given in the appendix. For the symbolic system: $\delta S_S = \langle \delta m_i \mid i=1..n_s \rangle$, $m_i(t+\delta t) = T(m_i(t),S)$, where T is a discrete transition function setting the tuple position m_i occupied by one symbol to a new symbol determined by the context. Notice that it is the content of the tuple position that changes, not the symbol; a symbol is a constant indexical. It's tempting to describe T as a mapping, but this would be looking at it from the base level.

5 Replicating Systems

The goal of this section is to establish under what conditions numeric and symbolic systems are theoretically equivalent, such that one is a replication of the other in specific respects.

5.1 Simulation, Emulation, Replication

This section reviews three broad terms that have been used to characterise the nature of system equivalence.

Simulation is typically used with its mathematical meaning and in general refers to the construction of a mathematical model for something in order to study its properties in terms of the model. There are two methodological approaches to simulation: qualitative versus quantitative, the former models the internal processes of the larger model, while the latter merely computes the consequences of the processes after some time interval. For example, a mouse could be modelled entering and being trapped in a mousetrap, a detailed qualitative model might include equations for the movement of the mouse and trap, a quantitative model might just compute a probability that the mouse is trapped given initial positions for mouse and trap – see B.Kuipers [Kup86]. A computer is often used to iteratively evaluate the model equations. One could construct a mathematical model of a mousetrap and simulate it on a computer, but it could only catch simulated mice – see Dennett [Denn78, p191].

Emulation is broadly defined as the faithful imitation of something. Normally its use would be qualified, e.g. a functional emulation of a wooden mousetrap might be made from metal and use a different mechanism. The objective is to emulate the function of catching real mice. In electronics, programmable logic array chips are often used as emulators for other chips. The emulator is programmed to perform the same logical function. Fetzer [Fetz90, p17] defines emulation as "effecting the right functions by means of the same – or similar – processes implemented within the same medium." Insisting that mind could only be reproduced through emulation would imply that consciousness could only be explained in terms of neurons and our knowledge of the mechanism would stop there. To emulate human minds would mean constructing biological creatures. This would be supporting a material chauvinism thesis toward physical realisation.

Replication is a weaker type of imitation than emulation; Fetzer defines it as "effecting the right functions by means of the very same – or similar – processes." Replication would imply consciousness only requires a certain X (e.g. causal pattern) to be instantiated, which is not an exclusive property of the medium (i.e. neurons) – X could be instantiated in other mediums. It might turn out that to achieve X requires the functionality of neural analogues. Searle [Sear84] states that a computer simulation could never duplicate mind. Here, duplication is understood to mean reproduction of the properties of mind, in particular qualitative consciousness. By definition, a successful emulation or replication would duplicate the properties being imitated.

5.2 Theoretical Conditions for System Equivalence

To determine whether mild AI can succeed requires establishing under what conditions numeric and symbolic systems are theoretically equivalent, such that one is a replication of the other in specific respects. The appendix presents a formal characterisation of systems in terms of the ontological levels they contain and uses this to develop a theory for process equivalence. The implications of this are summarised here.

In summary, an ontological level is individuated by the laws that are in force, and the initial configuration of the elements in the level. The formal definitions suggest that a number of conditions apply to comparing systems. These centre on system level equivalence, dynamical equivalence and structural equivalence. Common to these is the perspective of the observer, the fidelity of their measurements and the properties of interest. A distinction was made between strong and weak equivalence, where strong equivalence demands all sublevels are equivalent, while weak equivalence only applies to the levels concerned. Weak equivalence essentially requires an appropriate isomorphism exist between the levels. This was summed up in the theorem for weak structural continuous-discrete equivalence (T-WSCDE), which stated when a discrete process could be considered equivalent to a continuous process.

The characterisation of systems built up in the previous sections and the appendix suggests that a suitable digital machine might be able to approximately reproduce the causal structure of a continuous process, for a given base time interval, in a weakly equivalent sense – see theorem T-WSCDE. In other words, it might be possible to replicate the causal structure of a numeric system via a qualitative simulation implemented on a symbolic system. This means for a certain class of numeric systems, if causal structure is all that matters, then the causal structure of the numeric system might be subject to replication by a symbolic system. Now, according to chapters three and four, since causal structure is all that matters for reproducing mind, it follows that, to this extent at least, mild AI looks feasible. The next section considers whether it is practically possible to replicate mind on digital machines.

6 Producing a Mind in Future Computers

The potential ability of a digital machine to implement the hypothetical system (as discussed in chapter four), and so replicate a human mind based on the dispositional approach presented in the last chapter, can be estimated with a few calculations.

With regard to digital machines based on integrated circuits, Moore's law can be used to estimate their performance over the next few years. According to Moore's law the capacity of a memory chip doubles every two years – see G.Moore [Moo65]. This is related to the number of transistors on a chip and is also proportional to the number of instructions a processor can perform. Recently, Moore revised his law suggesting the rate will decline in light of the looming physical limitations of integrated circuits.

In 1997, Moore estimated that there are unlikely to be no more than five generations before the physical limits are reached in 2017. In 1997, the Intel Pentium Processor (233MHz) had a performance rating of approximately 400 MIPs (millions of instructions per second), contained approximately 5 million transistors, and could address 4 giga bytes of memory. By 2017, the projected processor performance would be 12,800 MIPs, contain 160 million transistors, and be capable of addressing 128 giga bytes of memory.

The following table gives a breakdown of the factors that determine the processor performance required to replicate a human mind:

	Factor	Range
A	# neurons in human brain	10^9 to 10^{12}
B	# connections per neuron	10^3 to 10^4
C	% neuron redundancy	10 to 50
D	% neurons active	1 to 10
E	# neurons per DD	0.1 to 10
F	% neurons using flows	50 to 100
G	# processor instructions to replicate DD	10×B to 100×B
H	# DD updates per second	100 to 1000

Table 4. Factor estimates for replicating a human mind via dissociated dispositions (DD).

The estimate for the number of neurons in a human brain comes from the DARPA neural network study – see [Darp88]. To incorporate the possibility that the brain may use sub-neural processes, such as quantum effects supported by the micro-tubule structure of neurons (see Dimitri Nanopoulos [Nan95]), the factor E has a lower value of 0.1 suggesting one neuron could support ten dissociated dispositions. The performance estimate is then:

Processor performance (# MIPS) = $\frac{A}{E} \times C \times D \times F \times G \times H / 10^6$ MIPs.

This gives a lower and upper bound of:

MIPs ≈ [$5/10^5$ to 50] × # neurons MIPs = [5×10^4 to 5×10^{13}] MIPs.

A conservative estimate (taking the lower bounds only for A, B and H) gives:

Conservative estimate: 5×10^8 MIPs.

Of equal importance is the processor memory required. The memory required to hold state information on each dissociated disposition is estimated as follows:

Processor memory per DD: # bytes ≈ J × B, where J ≈ 2 bytes.

Lower and upper bound for human mind (scaling upper bound by ten to allow for sub-neuron structure of micro-tubules):

[2k to 20k] × # neurons bytes = [2×10^4 to 2×10^8] giga bytes.

Based on these estimates the lower bound implies single processor digital machines would not have the performance power to replicate a human mind within the next twenty years, but perhaps within the next thirty years if the physical limits to Moore's law are overcome. The conservative and upper bounds would require computer performance to double twenty and thirty-seven times respectively, which is well beyond the maximum performance attainable at the physical limits. Notice that these estimates do not take into account the degradation of performance due to cache thrashing, which is a strong possibly since the processor cycles through its memory on each update cycle.

Another way to increase performance is through parallel processing, such as clusters of shared memory multiprocessors (SMPs). In this configuration, it's plausible that by 2017, a cluster of 10 SMPs each containing 100 processors rated at 10,000 MIPs would have a peak performance of 10^7 MIPs, which is close to the conservative estimate. To attain the upper bound performance through parallelism would require the equivalent of connecting together five million of these parallel machines.

7 Summary

A characterisation of systems in terms of ontological levels was developed and formally presented in the appendix. This served two purposes: to clarify the relationships involved in semiotic processes, and to compare numeric and symbolic systems, the two types of systems typically used to model brains and computers, respectively. Clarifying the relationships involved in semiotic processes focused on Peirce's triadic sign relation and the importance of the ground relation to Fetzer's counter argument to the analogical argument for strong AI. The comparison of numeric and symbolic systems showed that when levels are equated, they are fundamentally different kinds of systems.

Numeric systems are characterised by laws in the form of continuous constraints and continuous property variations. Symbolic systems are characterised by laws in the form of discrete actions leading to disjoint structural changes in the order of the symbols. However, two points have to be considered. Firstly, mild AI by definition does not subscribe to the thesis of material chauvinism. Secondly, by taking into account the perspective of the observer and the level of comparison, it might be possible to replicate the causal structure of a numeric system via a qualitative simulation implemented on a symbolic system.

The ability of digital machines to practically implement dispositional systems for the purposes of replicating a human mind was estimated. Upper and lower bounds and a conservative estimate were calculated for the required processing power. According to Moore's Law computer power doubles every two years. Based on this, a parallel digital machine capable of satisfying the conservative estimate could be built within the next twenty years. However, beyond this date, physical limitations prevent further performance improvements and mark an end to the applicability of Moore's Law. Consequently, to satisfy the upper bound would require the equivalent of connecting together five million of the parallel machines – not a very encouraging prospect.

Chapter 7

Conclusion

1 Overview

The thesis has investigated the operational requirements for replicating a species of mind on a digital machine as part of an exploration into the prospects and promise of mild AI. Central to this was a definition of mind as a semiotic system. This led to an exploration into the operational requirements for implementing a semiotic system.

Two of the more important requirements suggested were as follows. Firstly, preserving the causal-flow structure of a mental process is central to any replication of mind. Secondly, the system is causally structured through its need to signify reality by categorising its aspects as operational entities (i.e. networks of signs) from which decisions can be made, and which in turn are causally accessible to other operational entities.

The operational requirements led to the suggestion that to implement mind, irrespective of any underlying hardware, requires a system with at least three ontological levels, where an ontological level is individuated by the laws that are in force. A formal characterisation of systems in terms of ontological levels was presented.

Focusing on the nomic form of systems, brains and computers were mathematically categorised as types of numeric and symbolic systems, respectively. The formal characterisation of the ontological levels in a system set out conditions in which certain symbolic systems can be said to replicate certain numeric systems. This suggested circumstances in which a computer would be capable of replicating a species of mind, and thus supporting the promise of mild AI.

1.1 Logical Development

The main points made in the previous chapters are drawn together and discussed.

The field of AI is revisited in light of the exploration into the prospects of mild AI. Given that mild AI may show promise, the implications and impact of this on the analogical argument for strong AI is discussed. Major issues face the future of mild AI, such as determining the requirements for, and authenticating, an AI person. The prospects of mild AI are discussed.

Peirce's triadic semiotic process is categorised in terms of ontological levels. This places the sign, subject and immediate object at the phenomenal level, the interpretant at the dispositional level, and the dynamical object at the implementation level. A Peircian sign is the qualitative counterpart of an entity signified for operational purposes.

Fetzer's laws for human beings are supplemented by suggesting brains could be treated as compound systems containing a neurological system and a lawfully dependent dispositional system.

Finally, the version of mild AI supported here takes the operational view that mind is a property of an occurrent system wide process. Therefore, this version of mild AI is not susceptible to arguments levelled against the analogical argument for strong AI since it doesn't equate minds with programs (Premise 3).

The main contributions to knowledge are summarised. Areas for further research are itemised.

2 Strong and Mild AI Revisited

In his early writings, Searle [Sear84] argued that the level of comparison between minds and programs that was dictated by strong AI was not sustainable since the computer operated solely according to syntactical means, whereas the brain involves more than syntax. Consequently, Searle (p36) suggested that machines can not think when the question is posed as "Is instantiating or implementing the right computer program with the right inputs and outputs, sufficient for, or constitutive of, thinking?" Searle's more recent writings are more liberal toward computers – see Searle [Sear98]. Fetzer argued along similar lines that minds are semiotic systems while computers are symbolic systems, and these are different kinds of causal systems. Both counter arguments to strong AI depend on the notion of a suitable level of comparison and an unspecified means of implementing non-syntactical processes and semiotic systems.

While these arguments certainly appear to hold at a direct level of comparison between mind and programs, their applicability to other levels of comparison is less certain. This leads to the question, Does a computer-based simulation of mind also replicate it? Searle [Sear84, p37] suggests otherwise, no matter how powerful its ability to simulate, the computer is unable to replicate (Searle uses the word duplicate), “a computer simulation of a fire is [un]likely to burn the house down.” “No computer program by itself is sufficient to give a system a mind.”

Consequently, to determine if these arguments applied to mild AI the previous chapters conducted an exploration into the ontological levels involved (in contrast to epistemological or pragmatic levels) and the relationship between simulation and replication.

2.1 Requirements for an AI Person

Creating an artificial person would require endowing it with the capabilities of mind discussed in chapter two, e.g. reasoning, imagination and sensation. Would the creature require eyes, ears, arms and legs, a language of thought, a biological or silicon based body? Chapter two noted that a range of creatures populates the animal kingdom whose abilities originally matched the skills they needed to survive in their environment. Some animals lack language but, nevertheless, show signs of perceptual intelligence. This suggests a range of AI creatures with various abilities could be developed. Such developments are already taking place and have given rise to two sub-fields of AI, namely Artificial Life and Intelligent Agents – see Franklin [Frnk95].

The majority of the current practical AI developments are based on digital hardware, but, as was discussed in chapter six, whether this hardware will support mind is still questionable. However, the analysis of the four categories of AI suggested that for mild AI to be accepted as having succeeded it would have to implement a full blown artificial person, which includes consciousness. This extreme requirement arose largely as a result of the lack of knowledge concerning the relationship between mind, cognition and consciousness. For example, while it is quite acceptable in cognitive science to study features of mind in a modular way, this does not carry over to attributing success to mild AI when it replicates one of these modules. Thus, for mild AI to succeed on digital hardware it would have to show how mind could arise from such hardware.

2.2 Testing the Authenticity of AI

How would we know that the AI person is not a mere robot lacking consciousness? The Turing Test for intelligence (see Truing [Tur50]) only tests the conversational abilities of a subject, it says nothing about thought processes or consciousness. Indeed, as Searle [Sear84] points out in the Chinese room argument, a simulation of intelligence could pass the test.

Ultimately, testing mild AI involves verifying whether the system is conscious and has qualitative experiences. This brings up the problem of other minds, which asks how can we know that others have qualitative experiences? Russell [Russ48] concluded that at best we could only infer by analogy that others have such experiences. By a process of inference to the best explanation, current scientific practice is to tentatively accept the theory that fits the data and provides the most viable explanation – see Fetzer [Fetz96]. To come to more solid conclusions would involve having a detailed theory for qualitative experiences and verifying that the system implements a model of the theory. This approach is endorsed in the more recent work of Searle [Sear98]: “the requirement that science be objective does not prevent us from developing an epistemically objective science in a domain that is ontologically subjective.” Therefore, to test the validity or equivalence of a mild AI mechanism in respect of its biological counterpart requires having a scientific theory of mind.

2.3 The Prospects of Mild AI

The field of AI was characterised in terms of its definitional dimensions. This focussed on Searle’s distinction between strong AI, “that the mind is to the brain, as the program is to the computer hardware,” and weak AI, “the view that the computer is a useful tool in doing simulations of the mind.” However, the applicability of this distinction depends on the foundational assumptions subscribed to.

Classical AI is based on symbolic systems, more recently these have been displaced in favour of numeric systems. A comparative analysis of these systems suggests a version of AI between the strong and weak extremes. That is, a mild form of AI, according to which a suitable computer is capable of possessing species of mentality that may differ from and be weaker than ordinary human mentality, but qualify as "mentality" nonetheless.

The success of mild AI only depends on it being shown that one of these systems could have a mind, and so much speculation surrounds their relative merits. To address this, systems were analysed in terms of ontological levels, and the two systems compared. The analysis shows that when levels are equated, numeric and symbolic systems are fundamentally different kinds of systems.

Numeric systems are characterised by laws in the form of continuous constraints and continuous property variations. Symbolic systems are characterised by laws in the form of discrete actions leading to disjoint structural changes in the order of the symbols. However, two points have to be considered. Firstly, mild AI by definition does not subscribe to the thesis of material chauvinism. Secondly, by taking into account the perspective of the observer and the level of comparison, it is possible to replicate the causal structure of a numeric system via a qualitative simulation implemented on a symbolic system.

Consequently, the success of mild AI depends on whether qualitative consciousness is solely dependent on a process having the right causal structure, and that it is not necessarily continuous. If this is so, the prospects for mild AI look encouraging. Chapters three and four suggested that replicating the right causal structure in a system is sufficient for producing a semiotic system with qualitative consciousness.

3 Semiotic Systems & Ontological Levels

From an operational perspective, the requirements developed for a semiotic system in chapters three and four, led to the conclusion that to generate mind and consciousness requires a system containing a hierarchy with at least three ontological levels in the sense of chapter three. These were identified as an implementation level, a dispositional level and a phenomenal level. This has certain consequences. Firstly, it suggests Peirce's semiotic process may promisingly be mapped onto the operational view. Secondly, it suggests ways of supplementing Fetzer's laws of human beings and digital machines. Finally, it implies that certain digital machines might be able to replicate certain semiotic systems.

3.1 An Operational View of Peirce's Semiotic Process

Peirce characterised the semiotic process in terms of a triadic relation between the sign, object and interpretant, which was analysed for its logical structure by focusing on what was revealed to an observer from a first person perspective. The requirement for properties of levels highlighted that the nature of a phenomenon depends on the ontological level and the observer's perspective. Hence, the analysis of the semiotic process can be extended by distinguishing between the phenomenal aspects and the underlying operational processes, and showing how they are situated according to the system's ontological levels.

Consequently, Peirce's trichotomies can be analysed further by contrasting the phenomenal and operational. The first trichotomy deals with the properties of signs. Hence, the qualitative properties of a sign are a manifestation of the operational structure and accessibility of the underlying dispositional causal-flows when embedded in a specific context – see the requirement for system structuring for causal accessibility. Similarly, the second trichotomy, which concerns the semiotic ground of the sign, reflects an indexical, iconic or symbolic relation being modelled in the signified perception of reality by an underlying network of instantiated dispositions. Finally, the third trichotomy, which refers to the effect of the interpretant on the system, operationally corresponds to the domain of influence of these underlying disposition instances.

Situating the components of the semiotic process in terms of the system's ontological levels can be explained by reference to the earlier analysis of the interpreter regress problem. One reason the regress problem arises is that the interpreter and representation interpreted are situated at the same ontological level, in this case, the phenomenal level. However, the sensed direction of the sign-subject relation is a phenomenal level property. The subject being a representational perspective view of reality constructed / signified by the system – the famous Cartesian theatre. This point is summarised in the statement on (the) direction of dissociated disposition flows (S-DDDF), which is linked to the statement of irreducible binding (S-IB), and the statement of the first person perspective view of consciousness (S-FPPVC). Operationally, at the dispositional level, there is no stationary-representation to a subject, there is just the impinging causal influence of a network of causal-flows. It is the net effect of this activity that produces a signified perception of reality with a representational perspective positioned as the subject to which our signified folk-psychology tells us is a conventional (stationary) sort of representation – see the section on irreducible binds in chapter three.

Therefore, in Peirce's sign relation, the sign, subject and immediate object (representation) are properties of the phenomenal level. Operationally, signs arise as a manifestation of the systems requirement to signify reality in order to distinguish and categorise entities in the process of making decisions. Thus, a Peircian sign is the qualitative counterpart of an entity modelled for operational purposes – see the section defining "entity" in chapter three, and those following in chapter four. The interpretant is a property of the dispositional level, and finally, the dynamical object (i.e. the entity) can be thought of as a property of the implementation level – if this is equated with the physical reality. However, Peirce allowed the sign from one semiotic process to act as the object of a subsequent semiotic process. This leads to two possibilities operationally. Firstly, the instantiated dissociated dispositions underlying one sign (i.e. signified object) can influence another object's instantiated dissociated dispositions. Secondly, an extension on this, the influence may be indirect through its effect on the attention and decision mechanisms and their subsequent effect on the object – see start of chapter three for meaning of "dissociated disposition" and the end of chapter four.

3.2 Adding Ontological Levels to Fetzer's Laws

The analysis suggests an ontological level mediates between brain and mind. Consequently, Fetzer's laws (see Fetzer [Fetz90, p284]), for example HL6, can be refined by treating the base system for humans, i.e. brains B*, as a lawfully compound system that obeys:

(HL6') $(x)(t)(B^*xt \Rightarrow D^*xt)$

(HL6'') $(x)(t)(D^*xt \Rightarrow M^*xt)$

from which it follows:

(HL6) $(x)(t)[(B^*xt \Rightarrow M^*xt)$,

which asserts that, given for all x and all t, if x were a brain of kind B* at t, that produces a kind of mind M*, then x would contain a (neurological) system of kind B* and a dispositional system of kind D*, that are lawfully dependent. The subjunctive arrows (…⇒__) are justified on the ontological as opposed to logical grounds of corresponding permanent property relations between B* and D* and between D* and M*. The laws were stated with respect to brains and minds, "as predispositions to acquire semiotic dispositions", rather than transient brain-states and transient mind-states, since the spatial-temporal limit problem makes relating a static brain-state with a mind-state problematical. The other laws can be supplemented in a similar manner.

3.3 Symbolic Replication of Semiotic Systems

Theorem T-WSCDE, presented in the appendix, effectively concluded with three conditions that a numeric system must satisfy if it is to be replicated by a symbolic system. Firstly, the numeric system must be solely characterised by its causal structure. Secondly, it must be *possibly* (defined in appendix) discontinuous with respect to time. Finally, observable properties must be equivalent to an observer from a given perspective and for a given resolution.

Chapter four suggested that a system of instantiated dissociated dispositions is sufficient for producing a semiotic system. Instances of dissociated dispositions were characterised as causal-flows and a system of instantiated dissociated dispositions as a topological structure amongst causal-flows. This satisfies the first condition.

Although consciousness appears to the beholder as a continuous experience, two factors suggest the underlying process need not be continuous. These are, firstly, the fact that every facet of reality has to be signified, including the experience of continuity – see the section on binding qualia into entities in chapter three. Secondly, dissociated disposition dynamics may permissibly be based on statistical properties of the underlying medium, such as averages – see the section on redressing the spatial-temporal limit problem in chapter three. This holds so long as any granularity introduced into the dynamics is below the resolution of the observer. In this case, this is relative to the first person perspective view of consciousness (S-FPPVC). However, the resolving power of the conscious creature is a property of the phenomenal level. Consequently, it will not be capable of observing any dispositional level discontinuities unless they are signified explicitly by the dispositional level. These two factors collectively satisfy the remaining two conditions.

Finally, with regard to the third premise in the analogical argument for strong AI (see chapter one), programs can be categorised as aspects of the implementation level.

Mind arises as an occurrent system wide activity, while the experience of mind is a phenomenal level property. Consequently, mild AI does not equate programs with minds, and so does not conform to the analogical argument, or succumb to arguments levelled against it.

4 Summary of Main Contributions to Knowledge

1) Distinguishes the category of mild AI from strong AI and weak AI, and explores its prospects and promise.
2) Presents a formal characterisation of systems in terms of the ontological levels they contain, where an ontological level is individuated by the laws that are in effect and the configuration of its elements. Ontological levels are to be distinguished from organisation levels.
3) Suggests that, starting from a physical implementation level, three ontological levels are required in order to implement a semiotic system, irrespective of the underlying physical realisation. These are an implementation, dispositional and phenomenal level.
4) Presents a set of operational requirements for implementing semiotic systems under the terms of the dispositional approach and independent of a particular physical realisation.
5) Suggests two important requirements for producing a conscious semiotic system are the causal-structure of processes, and the causal accessibility of the contents of consciousness.
6) Formally states conditions under which a symbolic system may replicate the ontological levels of numeric systems.
7) Presents a description of instantiated dissociated dispositions based on the notion of causal-flows. These characterise the variable dispositional potential of the processes within certain kinds of causal systems. This lends itself naturally to a treatment in terms of group theory and wave mechanics.
8) Identifies the nature of a dispositional programming methodology and language for operationally modelling the semiotic systems and the phenomenal aspects of consciousness.
9) Presents a practical example of dispositional oriented operation that has some functional advantages over conventional symbolic manipulation when causal structure is of importance.

5 Further Research

The following areas currently remain open and require further research:

1) Formalising the operational requirements. A useful research project would be to pick up from where chapter five left off, and formalise the operational requirements in a deductive fashion and examine which are necessary or sufficient for implementing semiotic systems.
2) A dispositional programming language is needed to help take the next step of investigating networks of instantiated dissociated dispositions and what topologies give rise to particular qualia in a given model context. This would enable further systems to be investigated and described in terms of instances of dissociated dispositions rather than symbolic manipulation. For example, an interesting research project would be to formulate the pure ground relations in terms of networks of dissociated dispositions. Another project would be to translate a semantic network into a dispositional network.
3) One indirect approach to modelling dissociated dispositions is to map the group theory of dissociated dispositions to a functional language, such as ML. Problems to be dealt with are how best to specify the network structure and capture the evolving nature of the causal-flows underlying the instantiated dissociated dispositions. A good place to start would be a hybrid of Fetzer's language for universal dispositions with Lukasiewicz's multi-valued logic and the operational equations presented in chapter five.
4) To reduce processing costs, better ways are needed for replicating causal structure and dispositional dynamics without having to replicate the full dynamics via the implementation level.
5) Cataloguing and characterising the relationship between a causal-flow space and perceptual space. For example, what causal-flow structure gives rise to the psychological colour red? One approach to discerning the causal-flow structure of modalities would be to comparatively analyse common features of the natural environment, e.g. what causal-flows are common to all natural things that are coloured red?
6) Basic group theory was used to describe dissociated dispositions. This needs to be firmed up. What are the group theory implications for self-referent groups, i.e. group operations that modify the size and element content of the group itself?

Again, how would this be described in a dispositional programming language? In addition, what is the best way to program a system in which the type of the axes basis not only evolve and change, but also change across the domain? Also of concern is that the brain is an evolving system: the dissociated dispositional structure is forever changing. The thesis suggests that this only happens in specific locations in a parameterised way. What are these?

7) Devising a practical test case. To test and explore dissociated dispositions a suitable test creature needs to be characterised for modelling purposes. The simple examples explored in the thesis are too simple to be useful in this regard. However, the number of instances of dissociated dispositions required by a problem rises sharply with conscious resolution for even the most basic cases.

6 Epilogue

It is difficult to say how much the neural net structure of the brain has influenced these investigations into the operational requirements for replicating mind in machines. Therefore, it is, or is not, surprising to see that the investigation suggests neural nets are able to support semiotic systems and instances of dissociated dispositions. Alternatively, one could argue that instances of dissociated dispositions could only be supported by a mechanism similar in structure to neural nets. Consequently, evolution, constrained in this way, was naturally driven toward embodying dissociated dispositions in neural nets by evolving neural nets to suit dissociated dispositions! In any case, the thesis presents a neutral account of the operational requirements for implementing semiotic systems within the confines of the dispositional approach from which more general and detailed accounts could spring forth. No longer are qualia irreducible mysteries.

In the final analysis, the bedrock to philosophy of mind and cognitive science will be a marriage between philosophical analyses and precise mathematical formalisms for general consciousness. The thesis suggests that these notations and conventions will enable artificial systems to be infused with sensory experiences and qualitative consciousness.

Although this thesis cannot claim to have proved that species of non-human mentality can be implemented in a machine, it strongly suggests that if such an outcome is possible at all, it looks as though the direction to pursue is that of mild AI, whose prospects appear to be highly promising.

Appendix

1 $Ł_{\aleph 1}$ The Nondenumerably Infinite System of Lukasiewicz

Truth Rules:

1) $/\neg p/ = 1 - /p/$

2) $/p \wedge q/ = \min\,[/p/, /q/]$

3) $/p \vee q/ = \max\,[/p/, /q/]$

4) $/p \rightarrow q/ = \min[1, 1 - /p/ + /q/]$

5) $/p \leftrightarrow q/ = /(p \rightarrow q) \wedge (q \rightarrow p)/$

Truth values: [0,1] i.e. the continuum. Only 1 designated.

Recast using max only instead of min:

$$\min[p, q] = 1 - \max\,[\neg p, \neg q] = \neg \max\,[\neg p, \neg q].$$

2) $/p \wedge q/ = \neg \max\,[/\neg p/, /\neg q/]$

3) $/p \rightarrow q/ = \neg \max\,[0, /p/ - /q/]$

4) $/p \leftrightarrow q/ = \neg \max\,[\,/p/ - /q/, /q/ - /p/]$

Note that: $/p \wedge q \wedge r/ = \min\,[/p/, /q/, /r/]$.

For further details see Rescher [Resc69].

2 Basic Group Theory

This description is a summary of that given by Cohn [Cohn93, p42+]. A monoid is a set S with a neutral element e and a binary operation $\mu\colon S \times S \rightarrow S$. If $z = \mu(x,y)$ is the result of applying μ to the elements $x,y \subset S$, with $z \subset S$, then:

$$\mu(x, \mu(y,z)) = \mu(\mu(x,y),z) \quad \text{for all } x,y,z \in S.$$

$$\mu(e,x) = \mu(x,e) = x \quad \text{for all } x \in S.$$

If the binary operation is defined as addition, the neutral element is denoted as 0, if the operation is defined as multiplication, the neutral is denoted by 1.

An element x is said to be invertible if there exists an element y such that: $xy = yx = 1$, where $x,y \subset S$. The inverse of x is denoted by x^{-1}. If $xy = yx$, then x and y are said to commute. A group is a monoid in which every element is invertible.

One way to specify the binary operation is via a group table. For example, in the group of three elements: 1, a, b, under muplication, one of the possible group tables is:

x \ y	1	a	b
1	1	a	b
A	a	b	1
B	b	1	a

Let G be a group distinct from the set S. In an analogous manner to the definition of a monoid, a group *action* on S is defined by a binary operation μ: $S \times G \rightarrow S$, satisfying:

$\mu(p, gh) = \mu(\mu(p,g),h)$ for all $p \in S$, $g, h \in G$.

$\mu(p,1) = p$ for all $p \in S$.

Note that G could act on itself, $S \equiv G$.

An *operator* ω: $G \rightarrow G$, is defined as a mapping which operates on a group: let Ω be a set of operators that act on G such that:

$(xy)\omega = (x\omega)(y\omega)$ for all $x,y \in G$, $\omega \in \Omega$.

Sometimes the elements of a group may be called operators when they are applied to a specific domain.

Z is the set of all integers, negative, positive and zero.

3 Formalising the Ontological Levels of Systems

The motivation for this section is to formally characterise the structure of a system in terms of ontological levels, properties and laws, in a manner that will help in understanding the fundamental difference between numeric and symbolic systems.

Emmeche et al. [Emm97] presents the beginnings of an ontology for levels. They distinguish between primary levels and sublevels, suggesting that the primary levels include the physical, biological, psychological and sociological levels, and that within the biological level, sublevels include the cell level, the organism, the population, the species and the community levels. Michael Polanyi's [Pola68] introduction of boundary conditions is highlighted for constraining the behaviour of the higher level system. Here the structural organisation of the lower level elements in combination with the laws of the level determines the properties of the higher level. For example, S.M.Ali and R.M.Zimmer [Ali98] have presented a formal framework for describing levels of emergence in systems based on cellular automata and meta-rules. Here a state transition rule at one level becomes a state at the higher level.

Nis Baas [Baas94] defines a multi-level emergent structure as a hyperstructure of order N given by: $S^N = \Xi(S^{N-1}_{i_{N-1}}, Obs^{N-1}, Int^{N-1}, S^{N-2}_{i_{N-2}}, \ldots)$, where Ξ (Baas uses the letter R) is an abstract construction process over a number of lower level structures ($S^{N-n}_{i_{N-n}}$, $i \in$ some index set), according to interactions Int^{N-n} between these structures. Obs^{N-n} corresponds to a set of observational mechanisms that enables properties of the level to be measured. P is said to be an emergent property of level N if $P \in Obs^N(S^N) \wedge P \notin Obs^N(S^{N-n})$, in other words, if P cannot be observed given details only on lower level structures.

The following sections formally refines Emmeche et al.'s ontology of levels and extends Baas's emergence framework [Baas96] in a systems context. This is necessary since their framework is property centric in that the observability of a property is used as a criterion for establishing emergence. In addition, the details of the construction process and its relation to the hyperstructures are undefined. The following sections present a formal description of the ontological levels, laws and properties of a system.

3.1 Structure of Systems

The first task is to define the universe from which the systems will be constructed. For the present purposes this is done by characterising the universe in terms of the laws that are in effect, the elements it contains, and the structure on the elements induced by the laws. The assumptions introduced by this characterisation are discussed after the definition. Later it will become clear that a duality exists between descriptions of the universe and systems. Thus, the universe will be treated as a special kind of base system i.e. it's laws are proper axioms rather than derivable from a lower level – see below for details. Therefore, adopting the model theory framework of Hodges [Hodg93], a system is defined in piecemeal via the following five definitions:

Definition of System Axioms. (D-SA)

In a system S_i the set of laws that are in effect are collectively referred to as the axioms of the system, and will be denoted by the Hintikka set $A_i = \{a_{ij} \mid j \in 1..n_a\}$, where a_{ij} is a axiom, and n_a the number of laws.

A distinction could be made between laws of nature and less fundamental derivative principles, but for consistency these will both be referred to as axioms of the system. Laws can be thought of as externally given, they are what intrinsically drive the system dynamics. Principles arise at higher system levels, they are consequences or a reformulation of the underlying laws and their structuring of the system. Requiring the axioms form a Hintikka set simply asserts some common sense meta-constraints, for example if axiom a_{ij} is the law $F=ma$, then being a Hintikka set guarantees that $F \neq ma$ is not also one of the axioms.

Definition of System Configuration. (D-SC)

The configuration of the elements for a system S_i is specifiable as a set of sentences C_i, in a formal language L, under an interpretation I_i.

Definition of System Domain. (D-SD)

The set of elements $E_i = \{e_{ij} \mid j \in 1..n_e\}$, where n_e is the number of elements, upon which a system S_i is constructed is called its domain.

The formal language L and the interpretation I_i are introduced as a convenient method for theoretically discussing the structure of the system. For the systems dealt with here L is incidental to the main thread and it will suffice to take L as a first order logic. For convenience, L-structures are understood to be L-interpretations for which an L-valuation exists where appropriate.

Definition of Class of Systems. (D-CS)

The L-structure $K_i = \langle E_i, F_i, R_i \rangle$ axiomatised by A_i, defines a class of systems, where E_i, F_i and R_i are sets of symbolic names for the elements, operations and relations of K_i.

The axiomisation of K_i by A_i means that all true sentences in K_i are those that are true in every model of A_i, in other words every true sentence is consistent with the axioms.

Definition of a System. (D-S)

A system is defined by the L-structure $S_i = \langle E_i, F_i, R_i \rangle_{Ci} \subset K_i$, where the subscript C_i indicates a set of sentences from L specifying the state of E_i, where $C_i \models A_i$.

The set C_i describes the configuration of the elements (in terms of the F_i and R_i) into a particular instance of a system, and the requirement $C_i \models A_i$ constrains all specifiable configurations of the elements to obey the axioms. In effect, a set of axioms and a configuration of elements determine a system.

This definition of a system is based on the assumption that there is a set of axioms which somehow structure a collection of elements and determine the system evolution. It also presupposes a set of elements exists in some form. The next step is to clarify the nature of the axioms and their relationship to elements. Before that some terminology and consequences of the above definitions are presented. Firstly, $\varnothing$ represents the empty set, and the relative complement $A \setminus B = \{ x \in A \mid x \notin B \}$.

Definition of Higher Order System. (D-HOS)

Given a base system S_i^0, *a higher order system* S_i^n *is a structure:* $S_i^n = \Xi_i^n(S_i^{n-1})_{C_i^n}$, *where* Ξ_i^n *is a construction process on* S_i^{n-1}, *such that* $(A_i^n \setminus A_i^{n-1}) \neq \varnothing$, *and the subscript* C_i^n *denotes the configuration of the elements necessary for* S_i^n, $n \geq 0$.

The notation C_i^n for the configuration of the elements follows that as in D-SC & D-S.

Definition of System Basis. (D-SB)

The system S_i^{n-1} *from which* S_i^n *is constructed is called the basis for* S_i^n.

The notation and construction process Ξ_i^n is similar to that presented by Baas, although the term higher-order is used instead of emergent. In addition, now the emphasis is on a difference in axiom sets as demarcating higher-order systems and, in this case, the emergence of the system is taken to be deducible. For the present purposes definition D-HOS has been made less general than that of Baas since the construction process operates solely on the supporting system rather than a range of lower-order systems. The nature of the construction process Ξ_i^n for generating higher order systems is discussed in chapters three and onwards.

On a technical note, that the axiom sets of a system and its support are not equivalent is a sufficient rather than a necessary requirement for a higher-order system. Since these are systems of different orders, the axiom sets will be a consequence of different domains and so, strictly speaking, incomparable and therefore not equivalent by definition. However, what is important here is not that the axiom sets are, under some mapping, isomorphic, but rather that the actual processes that give rise to one set are not also directly responsible for those in the other set. The 'directly' clause permits the simulation of a system and the system simulated to have isomorphic axiom sets with the simulation still counting as a higher order system. One concern not addressed here is what determines the axioms of the base level?

To help distinguish between a system and the systems from which it is constructed the following alias is used:

Definition of System Level. (D-SL)

The set of systems from which a system S_i is constructed is denoted by the set $L_i^n = \{ l_i^n \mid n \in 0..(n_l - 1) \}$, *where* $l_i^n \equiv S_i^n$ *is called an ontological level of the system of order n.*

Hence, a set of elements and a set of axioms define an ontological level. A level may simultaneously support more than one higher order system.

Definition of System Hierarchy. (D-SH)

In a system S_i, a hierarchy H is a set of objects upon which a partial ordering, $\subseteq$, *is defined.*

For example, the objects could be levels, where the ordering is over the order of the level, or groups of elements where the ordering is over their functional organisation. It then follows that a particular system will have been constructed from a hierarchy of levels relative to a base system.

3.2 First and Third Person Perspectives and Explanations

This section formally distinguishes explanations from first and third person perspectives. Firstly, from the definition of levels, the following generalisation is made:

Definition of Self-Level-Referent Sentence. (D-SLRS)

A self-level-referent sentence r_{ij}^x *of a level* l_i^x *in a system S_i is a sentence that refers to the elements* E_i^x *of* S_i^x.

Here self-reference can be interpreted as a weak form of cross-reference as typified in combinator fixed-point theory – see Smullyan [Smul94].

Definition of First and Third Person Perspectives. (D-FTPP)

Given a system S_i *with a set of levels* L_i^n*, a first person perspective is an interpretation via* I_i^x *of any self-level-referent sentence* r_{ij}^x *of* l_i^x *in* S_i*, where x designates a single level c. A third person perspective for a system* S_i *is an interpretation via* I_i^x *of any sentence* s_{ij}^x*, for any level x, where the sentence* s_{ij}^x *refers to any other system* S_k*,* $i \neq k$*.*

Notice that this definition does not require the system to be a person or to have consciousness. From this, the notion of an *observer* arises as a third person perspective.

Theorem for Weak System Level Equivalence. (T-WSEL)

For an observer O_h^k*, from a first person perspective, in any two systems* S_i *and* S_j*, system levels* l_i^m *and* l_j^n *are equivalent if they have equivalent domains* $E_i^m \equiv E_j^n$*, axiom sets* $A_i^m \equiv A_j^n$*, and configuration sets* $C_i^m \equiv C_j^n$*, where* $(k>m, h=i) \wedge (k>n, h=j)$*.*

Proof: this follows immediately from definitions D-S, D-HOS and D-SL. Equivalence between sets is defined by an appropriate isomorphism. The reason for stating this theorem is that it suggests under what grounds two levels in two distinct systems may be treated as equivalent. However, the equivalence is only weak since it is relative to an observer defined on the system, i.e. it does not uniquely determine the means by which each level is generated e.g. m≠n. Hence, a strong version for an external observer would require all lower levels to be equivalent as well. A goal of chapters three and four is to argue that under certain conditions only weak equivalence is necessary in order for certain systems to replicate mind.

3.3 Properties of System Levels

The following introduces definitions related to properties of systems.

Definition of a System Level Property. (D-SLP)

In a system S_i*, a type of property* p^n_{ij} *of a level* l^n_i*, is any process for which an operator* ω^n_{ij} *is definable on the level domain* E^n_i*, where* $P^n_i = \{ p^n_{ij} \mid j \in 1..n_p\}$ *the set of permissible property types. An instance of the property is denoted by* $p^n_{ijx} = \omega^n_{ij}\, \varsigma^n_{ij}(C^n_{ijx}, E^n_i)$*, where the selector* ς^n_{ij} *picks out a collection of elements:* $E^n_{ijx} = \varsigma^n_{ij}(C^n_{ijx}, E^n_i) \subset E^n_i$ *for a given context* C^n_{ijx}*.*

The set of instantiated properties for a level, P^n_{ix}, is defined later. This is quite a general definition of a property and follows that of Baas. The nature of the operator depends on the context, e.g. if this was a quantum mechanical setting, the operators would have to be Hermitian if the property was to be observable. The definition covers both primary and secondary properties or attributes of systems and also encompasses properties based on statistics of the system behaviour. The set E^n_{ijx} may change as time progresses, for example the selector ς^n_{ij} may signify the current elements in a finite volume defined on the domain. The notation C^n_{ijx} will be used to refer to the configuration of a part of the system, and conforms to $C^n_{ijx} \models C^n_i$.

Definition of Property Equivalence. (D-PE)

In a system S_i*, two types of properties* p^m_{ij}*,* p^n_{ik} *are equivalent* $(p^m_{ij} \equiv p^n_{ik})$ *if there exists a pair of linear and affine transformations,* θ_L *and* θ_A*, such that:* $\theta_L \omega^m_{ij}\, \varsigma^m_{ij}\, \theta_A(C^m_{ij}, E^m_i) = \omega^n_{ik}\, \varsigma^n_{ik}(C^n_{ik}, E^n_i)$*.*

Definition of Connected Properties. (D-CP)

In a system S_i*, properties* p^n_{ijx} *and* p^n_{ijx} *of a level* l^n_i*, are said to be directly connected* $(p^n_{ijx} \xleftrightarrow{0} p^n_{iky})$ *if they have domain elements in common:* $E^n_{ijx} \cap E^n_{iky} \neq \varnothing$*, and indirectly connected* $(p^n_{ijx} \xleftrightarrow{1} p^n_{iky})$ *if there is a mutual connecting property:* $\exists z.(E^n_{ijx} \cap E^n_{ihz} \neq \varnothing) \wedge (E^n_{ihz} \cap E^n_{iky} \neq \varnothing)$*. Otherwise they are unconnected.*

Properties that are either directly or indirectly connected are said to be locally connected.

Definition of Index Function. (D-IF)

Given a set of variables $X = \{x_i \mid i=1..n_v\}$ *and a function* $f(x)$ *where* $x \subseteq X$. *The index function* $\gamma(f, x)$ *returns the set of variable indices referred to by* $f(x)$: $\gamma(f, x) = \{i \mid x_i \in x\}$. *For convenience,* $\gamma(x)= \gamma(I, x) = \{i \mid x_i \in x\}$, *the index set for* x.

Definition of Axiom & Axiom Property Domain. (D-AAPD)

In a system S_i *at a level* l_i^n, *an axiom* $a_{ij}^n \in A_i^n$ *is defined as an equation over a set of connected properties:* $a_{ij}^n =_{df} \alpha_{ij}^n(A_{ij}^n) = 0$, *where the set of connected properties is called the axiom property domain:* $A_{ij}^n = \{p_{ik}^n \mid k \in \gamma(a_{ij}^n)\}$, *and* $\gamma(a_{ij}^n)$ *is the set of property indices referred to in the definition of the axiom. The set* A_{ijx}^n *refers to a particular set of property instances for an instantiated axiom given a context* C_{ijx}^n.

The motivation for requiring the properties in an axiom property domain to be connected is to handle a potential referencing problem. For example, the law F=ma can only be applied if the force or acceleration refers to the same mass. Treating these as properties, the definition of their corresponding operators would refer to a common domain element. A more restrictive definition for connected properties may be defined by requiring that the common elements in the intersection sets support only valid properties.

4 Defining Interaction and Causation in Systems

The task for this section is to analyse the operational nature of the interactions in numeric and symbolic systems in order to characterise their differences in the following sections. This involves understanding what causal processes are in operation. Causation is the principle that an effect has a cause, while causality in the philosophical sense refers to explaining why something is the case – see P.Gasper [Gasp91] and T.Crane [Cra98]. A causal explanation “shows how the possibilities were restricted so that things had to be that way, so that had one known the explanation they would have been justified in expecting things to be that way” – see R.L.Gregory [Greg89, p128]. The thesis is concerned with both causation and causality for explaining under what conditions mild AI might be possible.

This section starts be reviewing causation in neural systems using concepts from analogue electronics. Generalising leads to a formulation, based on differentials, for characterising the dynamic causal structure in numeric systems. However, this does not uniquely determine causal structure, so conditions are also given for determining the causal structure equivalence of dynamic systems. From this, conditions can be established under which systems can be considered dynamically equivalent.

4.1 Causation in Neural Systems

Mind arises from brains, and brains, in theory, can be modelled as neural networks – see S.Grossberg [Gss88b] and R.M.Golden [Gold96]. In turn, neural systems are a type of numeric system whose evolution is determined by the interaction of state variables. Y.Iwasaki and H.A.Simon [Iwa94] have formalised a method for determining the causal ordering (i.e. dependency structure) among the variables in numeric systems. J.de Kleer and J.S.Brown [Klee84] have looked at causality as the propagation of disturbances in a constraint network. Much has also been written with respect to modelling processes (e.g. Milner's mobile processes [Mil92]), time and logic's (for a review see L.Bolc and A.Szalas [Bolc95]), and reasoning about change – for reviews see Orgun and Ma [Org94], and Vila [Vil94]. However, there is still more to be done in comparing the causal structure of numeric and symbolic systems.

Reviewing the neurobiology of neural systems, shows that the dominant operational processes discovered so far are based on the flow of charge carriers, notably electrons and ions such as those of potassium, sodium and chlorine – see G.M.Shepherd [Shep83], A.C.Scott [Scot77] and Grossberg [Gss95]. More recently, the possibility that the brain uses quantum mechanical effects has attracted much attention – see S.R.Hameroff et al. [Hame98]. In any case, the following discussion can be extended to accommodate quantum effects as well. Consequently, the following hypothesis is made:

Hypothesis for Numeric Systems Explication of Consciousness. (H-NSEC)

The mechanism underlying consciousness can be explained in numeric systems terminology.

This could have been cast as an axiom of the thesis. Its purpose is simply to highlight the assumption being made here that numeric systems terminology is sufficient for explaining the mechanism behind consciousness. The hypothesis is a refinement on the dynamical hypothesis as stated by van Gelder [Geld95, p5]: "Natural cognitive systems are dynamical systems, and are best understood from the perspective of dynamics." Notice that it doesn't make any necessary claims in the manner of Newell and Simon's physical symbol system hypothesis for the necessary and sufficient requirements for intelligent behaviour [Nwl76]. Chapters three and four present the operational requirements for qualitative consciousness based on this hypothesis. Therefore, it is important to have an understanding of the causal interactions occurring in neural networks and to formulate this in numeric systems terms.

The analysis starts by examining the nature of the interactions in an idealised neuron embedded in a neural system. Carver Mead [Mead88] presents a useful comparison in which neural systems are successfully reengineered using analogue electronics, with particular reference to transistors and integrated circuits. Consequently, for the present purposes it suffices to treat neural systems as conventional analogue electronic circuits. The interactions in these circuits are typified by voltage variations occurring at points (nodes) in the circuit when there is a change in the distribution of charge carriers.

The idealised neuron (see J.J.Hopfield [Hop82]) produces an output that is a bounded, weighted sum of its inputs. Mead [Mead88, p105] models signal aggregation in a variant of this neuron using differential transconductance amplifiers, which convert a voltage difference to a current. Here a set of input voltages $\{V_i \mid i=1..n\}$ impinges on the neuron, which produces a bounded output V_o. The bounds are a physical constraint due to the finite response range of the neuron. Within its working range, the output can be found by applying Kirchhoff's conservation of charge law:

$$\sum_{i=1}^{n} G_i (V_i - V_o) = 0, \qquad \therefore V_o = \left(\sum_{i=1}^{n} G_i V_i\right) / \left(\sum_{i=1}^{n} G_i\right),$$

where G_i is the conductance or weighting factor of an input.

To examine the nature of the interactions, the change in V_o due to a change in a single input V_j can be expressed as:

$$\delta V_o = A + B\,\delta V_j, \qquad A = C\sum_{i \neq j} G_i V_i, \quad B = CG_j, \quad C = 1/\sum_{i=1}^{n} G_i$$

This corresponds to a small change in the rate of inflow of charge to the neuron from the input, which results in a change in potential across the conductance. Charge accumulates at the neuron soma from all the inputs, and so the inputs essentially interact in a superpositional manner.

Although this was a simple example, the causal interactions in larger systems can be treated in a like manner. The next section generalises this treatment and suggests a mathematical construct for talking about the causal interactions in numeric systems.

4.2 Causally Infinitesimal Variations

In numeric systems causal interaction amounts to a transient property changing due to the influence of other processes. While this can be described in cause and effect terms, the influence causing the effect is typically an on going process, in which case it is said to drive or force the changing property – see Mead [Mead88] and Pain [Pain93]. Often these systems can be modelled as sets of coupled differential equations – see B.R.Frieden [Fri98], Gelder [Geld95] and S.Skogestad and I.Postlethwaite [Skog96]. One natural way of simulating these systems is by transforming the equations to difference equations and employing finite difference methods (FDM) – see B.Massey [Mass98]. For example, discretisation is a common practice in electrical engineering for approximating the dynamics of numeric systems – see M.H.Hayes [Haye96]. The success of the discretisation transformation depends on how well the resulting approximation preserves the effect of the causal interactions. The faithfulness of this approximation and the nature of the interactions can be analysed with the aid of some calculus.

4.2.1 Input-Output Variations

The derivation for the dependence between input and output variations in the case of the idealised neuron is an example of a more general formulation. To begin with the output of a device y, is expressed as a function of its inputs **x**, state parameters **s**, and time t: $y = f(\mathbf{x},\mathbf{s},t)$. A small change is considered in one of the independent variables, for example a change of δt in the scalar t, keeping the other vector variables **x** and **s** fixed:

$$\delta y = f(\mathbf{x},\mathbf{s}, t + \delta t) - f(\mathbf{x},\mathbf{s},t).$$

As the change in t approaches zero, $\lim_{\delta t \to 0} \frac{\delta y}{\delta t} = \frac{dy}{dt} = f'(\mathbf{x},\mathbf{s},t)$, i.e. the derivative of f with respect to t. This yields an approximation for δy as: $\delta y \approx f'(\mathbf{x},\mathbf{s},t)\delta t$, provided δt is small, and as $\delta t \to 0$, $\delta y \to dy$. For a simultaneous change in **x**, **s** and t:

$$\delta y \approx dy = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial s}ds + \frac{\partial f}{\partial t}dt\,.$$

If the parameters to f are combined into a vector, $h = (\mathbf{x},\mathbf{s},t)$, the change in y can be expressed in terms of the directional derivative of f with respect to h: $\delta y \approx \nabla f(h)\, dh$, where $\nabla f(h) = \sum_i h_i \frac{\partial f}{\partial h_i}$ is the differential operator for the gradient of f(h), and $\{h_i\}$ a basis for h. This means the change in the value of a variable can be treated as being proportional to the gradient of the system within its locality. In the following, these infinitesimal changes are called causally infinitesimal variations (CIVs). Applying this to the idealised neuron in the above example, $V_o = C\sum_{i=1}^{n} G_i V_i$, gives: $\delta V_o \approx C\sum_{i=1}^{n} G_i dV_i$, i.e. the change in V_o is proportional to the weighted sum of the changes in the input voltages.

4.2.2 Lawfully Aggregating Infinitesimals

For the circuits considered here, the view of causation that has arisen is one in which electrical properties do not interact through a series of disjoint events, but are forever in a state of continuous change. To understand how properties interact they were analysed in terms of small changes. In the limit, as continuity is approached, the continuous dynamics are recovered. Conceptually, at this limit the causal interactions are modelled by aggregating differentials according to the appropriate law of the system, e.g. Kirchoff's $\sum dV_i = 0$. Up to this limit it is a matter of aggregating infinitesimals (i.e. CIVs), e.g. $\sum \delta V_i$. For some circuits this may give a good approximation to the continuous dynamics and even be justified on the grounds that the interactions in electronic circuits are granular due to the quantum interaction of the charge particles – see Frieden [Fri98, p49]. Hence, the law of aggregation that is in effect determines the form of an interaction. This is now clarified by the following definitions.

Definition of Causal Interaction. (D-CI)

In a system S_i *at a level* l_i^n, *a set of property instances* P_{ijx}^n *is said to interact if it equals an instance of an axiom property domain* A_{ikx}^n, ($P_{ijx}^n \equiv A_{ikx}^n$). *This is interpreted as a relational dependency* H_{ijx}^n *between the property instances* P_{ijx}^n. *The set of all instantiated relations for the level is:* $H_{ix}^n = \{ H_{ijx}^n \mid j \in \gamma(P_i^n), x \mid C_{ijx}^n \models C_i^n \}$. *The set of all instantiated properties for the level is:* $P_{ix}^n = \{ p_{ijx}^n \mid j \in \gamma(P_i^n), x \mid C_{ijx}^n \models C_i^n \}$.

Notice that the set of (uninstantiated) relations for the level is: $H_i^n \equiv A_i^n$, i.e. a relational interpretation of the set of axioms for the level. The clause $C_{ijx}^n \models C_i^n$, in conjunction with the property index set $\gamma(P_i^n)$, restricts the property subsets to only those that are currently instantiated in the system.

Definition of a Constrained Property. (D-CP)

In a system S_i at a level l_i^n, a property p_{ijx}^n is a constrained property of a given set of properties P_{ijx}^n, $(p_{ijx}^n \prec P_{ijx}^n)$, if there exists an axiom $a_{ij}^n \in A_i^n$ such that $(A_{ijx}^n \subseteq P_{ijx}^n) \wedge (p_{ijx}^n \in A_{ijx}^n) \wedge (\{p_{ijx}^n\} \neq A_{ijx}^n)$.

The definition of a system level property expressed the property in terms of an operator on a selector for the domain elements. In what follows it is useful to have an expression for the property in terms of its connected properties. For neural networks of the kind considered here, the following assumption is made: given a consistent current system state, the next value of a property is approximately determined only by its locally connected properties. This requires a couple of intermediate definitions.

Definition of Property Neighbourhood Set. (D-PNS)

In a system S_i at a level l_i^n, the set of constrained properties connected to a property instance p_{ijx}^n is called the property neighbourhood set:

$$P_{Nijx}^n = \{\textstyle\bigcup_{h,y} P_{ihy}^n, \forall h, y \mid (\exists k . P_{ihy}^n = A_{iky}^n) \wedge (\exists g . p_{igz}^n \in P_{ihy}^n \wedge p_{igz}^n \xleftrightarrow{0 \vee 1} p_{ijx}^n)\} .$$

By definition each property instance p_{igz}^n is constrained by at least one axiom a_{iky}^n. Hence, the following related definition:

Definition of Property Constraint Set. (D-PCS)

In a system S_i at a level l_i^n, the set of instantiated axioms connected to a property instance p_{ijx}^n is called the property constraint set:

$$P_{Cijx}^n = \{a_{ihy}^n, \forall h, y \mid (\exists k . P_{ihy}^n = A_{iky}^n) \wedge (\exists g . p_{igz}^n \in P_{ihy}^n \wedge p_{igz}^n \xleftrightarrow{0 \vee 1} p_{ijx}^n)\} .$$

In general the property constraint set P_{Cijx}^n corresponds to a set of simultaneous equations in terms of the properties P_{Nijx}^n. For simple cases it may be possible to rearrange the equations to derive a function for the property value in terms of the other properties, such as the idealised neuron example in which an expression for the voltage was derived from Kirchoff's law:

Definition of Property Function. (D-PF)

The symbolic manipulation of a property constraint set P^n_{Cijx} *to yield a function for the value of a property* p^n_{ijx} *is denoted by the property function* ϕ^n_{ijx}*, such that* $p^n_{ijx} = \phi^n_{ijx}(P^n_{Cijx})$.

This leads to a definition for the variation of a property as a function of variations in the properties it interacts with:

Definition of Property Variation. (D-PV)

In a system S_i *at a level* l^n_i*, the approximate local variation in a property instance* p^n_{ijx} *is:* $\delta p^n_{ijx} \approx \nabla \phi^n_{ijx}(P^n_{Cijx}) dP^n_{Nijx}$*, where* $p^n_{ijx} = \omega^n_{ij}\, \varsigma^n_{ij}(C^n_{ijx}, E^n_i)$*. The set of property variations for the level is:* $W^n_{ix} = \{\, \delta p^n_{ijx} \mid j, x \in \gamma(P^n_{ix}) \,\}$.

Definition of Property Basis Set. (D-PBS)

For a system S_i*, given a property and a property function on a property constraint set such that* $p^n_{ijx} = \psi^n_{ijx}(P^n_{Cijx})$*, the set of basis properties for* p^n_{ijx} *is* $P^n_{Biix} = \{ p^n_{ijx} \mid j, x \in \gamma(P^n_{Cijx}) \}$*, where* $P^n_{Bijx} \subseteq P^n_{Cijx}$ *and* $\Diamond(p^n_{ijx} \notin P^n_{Bijx})$.

The final part of the conceptualisation involves defining a network of infinitesimals in correspondence to the causal interactions at a particular system level:

Definition of Causal Structure. (D-CS)

The causal structure of a system S_i *at a level* l^n_i *can be represented by the network* $G^n_i = \langle H^n_i, W^n_{ix}, H^n_{ix} \rangle$*, where* H^n_i *is a set of hyperedges defined by the axioms of the level,* W^n_{ix} *is a set of vertices defined by the set of variations for each property instance, and* H^n_{ix} *is the set of instantiated relations defined on* H^n_i *and* W^n_{ix}.

See the definition of causal interaction (D-CI), and G.Schmidt and T.Ströhlein [Schm93, p91]. Notice that the vertices are property variations rather than property values themselves. A more powerful analysis would be to express the structure of networks in terms of simplexes and use homology groups to classify the topology – see M.Nakahara [Naka90] and H.Flanders [Flan89].

5 Defining System Processes

This section builds on the prior two sections by characterising system processes.

5.1 Definitions of Processes

As a precursor, property values can be parameterised with respect to time:

Definition of Time Dependent Property Value & Variation. (D-TDPVV)

The value of a property p^n_{ijx} *can be expressed as a function* ψ^n_{ijx} *of time t:* $p^n_{ijx} = \psi^n_{ijx}(t) = \phi^n_{ijx}(\Psi^n_{Cijx}(t))$, *where* $P^n_{Cijx} = \Psi^n_{Cijx}(t)$. *The time dependent property variation is then:* $\delta p^n_{ijx}(t) \approx \nabla \phi^n_{ijx}(\Psi^n_{Cijx}(t))dt$.

For the present purposes a natural definition of a process is in terms of a causal structure:

Definition of a Process. (D-P)

For a system S_i *at a level* l^n_i *a process* $\pi^n_{ijx}(t)$ *is a causal structure* G^n_{ijx} *defined over a time period T=[a,b],* $0 \leq a < b$: $\pi^n_{ijx}(t) \equiv G^n_{ijx}(t)$, *where* $t \in T$. *The subscripts refer to a particular instance (x) of a type of sub-process (j) within the system.*

Here the causal structure has been explicitly parameterised with respect to time. The next step is to characterise the difference between continuous and discrete processes. This is based on the standard definition of a continuous function.

A real function $y = f(t)$ is continuous at a point a if and only if it is defined at $t = a$ and $\lim_{t \to a} f(t) = f(a)$. That is, if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(x)-f(a)| < \varepsilon$ for all x such that $|x-a| < \delta$ – see Swokowski [Swok79].

From definition D-TDPVV, the variation in a property can be written as: $\delta p^n_{ijx}(t) \approx \nabla \psi^n_{ijx}(t)dt$. Recall the definition of the derivative of a function: $f'(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\left[f(t+\Delta t) - f(t)\right]$, for all t such that the limit exists. Hence, in the limit $\Delta t \to dt$, and the property variation is, as expected: $\delta p^n_{ijx}(t) \approx \lim_{\Delta t \to 0} \psi^n_{ijx}(t+\Delta t) - \psi^n_{ijx}(t)$. The definition of a continuous function implies that: $|\psi^n_{ijx}(t+\Delta t) - \psi^n_{ijx}(t)| < \varepsilon$ for all $\Delta t < \delta$, since $\Delta t \geq 0$. This effectively means the variation of a property value with time is itself a continuous function if $\Delta t \to 0$ is permissible. Notice that while the variation $\delta p^n_{ijx} \to 0$ as $\Delta t \to 0$, it does not mean that the system state freezes, but that the number of infinitesimal variations occurring, per unit of time, approaches infinity. This leads to the following succinct definitions:

Definition of a Continuous Process. (D-CP)

A continuous process $\overline{\pi}(t)$ *is a process for which the time interval* δt *between property changes is possibly reduced to zero:* $\Diamond(\delta t \to 0)$.

The nature of the system determines whether the time interval can be reduced to zero.

Definition of a Discrete Process. (D-DP)

A discrete process $\ddot{\pi}(t)$ *is a process for which the time interval* δt *between property changes is greater than zero:* $\delta t > 0$.

In other words, the variation in a property value is not a continuous function.

5.2 Base Time Interval and Variable Sized Infinitesimals

The causally infinitesimal variations can be viewed as indivisible but variable sized changes in the property being modelled. This implies a maximal base time interval may exist between variations at which the continuous dynamics of a system can be approximated by discrete dynamics:

Definition of Weak Relative Measurable Dynamical Equivalence. (D-WRMDE)

Given a system S_i *containing a level* l_i^m *that has a process* $\overline{\pi}^m_{ipx}(t)$ *undergoing continuous dynamics, and a system* S_j *containing a level* l_j^n *that has a process* $\ddot{\pi}^n_{jqy}(t)$ *undergoing discrete dynamics. The two processes are relative measurable dynamically equivalent* $(\exists\eta.\,\overline{\pi}^m_{ipx}(t) \equiv \ddot{\pi}^n_{jqy}(t+\eta))$ *to an observer* O^k_h *from a first person perspective if corresponding properties of both systems are indistinguishable when measured by a higher level operation* M^k_h, $(k>m,\ h=i) \wedge (k>n,\ h=j)$.

The equivalence is said to be relative with respect to the system and level of the observer, and the resolution and accuracy of the measuring operation M. Resolution in this context can be understood as the number of measurements that can be performed per unit of time. If a measuring operation has a high resolution and high accuracy it is said to have a high fidelity. The measurement might be based on a statistic of the properties, in which case the processes are said to be statistically equivalent under some measure of statistical indistinguishability. This is a weak equivalence since it does not require the lower levels to be equivalent. From this it immediately follows that there is a maximal base time interval for the discrete process:

Definition of Local Maximal Base Time Interval. (D-LMBTI)

Given a continuous dynamics process $\overline{\pi}^{m}_{ipx}(t)$ *and a discrete dynamics process* $\ddot{\pi}^{n}_{jqy}(t)$ *that are relatively dynamically equivalent,* $\overline{\pi}^{m}_{ipx}(t) \equiv \ddot{\pi}^{n}_{jqy}(t)$ *under an operation* M^{k}_{h}. *There exists a maximal base time interval* $\tau^{n}_{j} < \varepsilon$, *such that if* δ^{n}_{j} *is the base time interval of the process* $\ddot{\pi}^{n}_{jqy}(t)$, *equivalence only holds when* $0 \leq \delta^{n}_{j} \leq \tau^{n}_{j}$.

The maximal base time interval is said to be a local maximum with respect to the continuous process being considered and the fidelity of the measuring operation. The upper bound ε on τ^{n}_{j} depends on the fidelity. The base time interval refers to the length of time between events in the process – see below. It is also a measure of the temporal sparseness or granularity of the process. When all the processes are considered in a level and statistical effects taken into account, such as electron flow and thermal drift, the maximal base time interval has to be replaced by an approximation to the average value for the level $\langle \tau^{n} \rangle$.

Normally in a dynamic system, infinitesimals corresponding to property variations would be expressed in the limit as a differential with respect to a time differential. However, in this analysis the objective is to characterise interactions and their causal structure. For this, the infinitesimals are conceptualised relative to their local maximal base time interval. In other words, for convenience all infinitesimals are treated as being maximal. This simply averts having to adopt the smallest maximal base time interval in a level and scaling all others with respect to it, which would require specifying how many of one infinitesimal variation it takes to effect a dependent infinitesimal – see Skogestad and Postlethwaite [Skok96].

5.3 Dynamic Structure of Causal Networks

The final part of this characterisation of system structure considers dynamic changes in the process structure and how this relates to system equivalence.

The definition of a constrained property did not distinguish the direction of dependency, i.e. which properties drive the changes. In some cases, the dependency direction may change as a function of the property values, for instance the direction of current flow at a junction. Hence, the independence of a property is relative to the state, axiom and property set under consideration:

Definition of Relative Dependence and Independence of Properties. (D-RDIP)

A relative dependent property $\widehat{p}^n_{ijx}(t)$ *with respect to an instantiated axiom* A^n_{ikx}*, is a constrained property whose current property value is determined by a subset of the axiom domain properties i.e.* $\exists A^n_{ikx}.(A^n_{ikx} \in P^n_{Cijx} \wedge A^n_{ikx} \cap P^n_{Bijx} \neq \varnothing)$*. Conversely, a relative independent property* $\widetilde{p}^n_{ijx}(t)$ *is a property whose current property value is constant with respect to an instantiated axiom:* $\exists A^n_{ikx}.(A^n_{ikx} \in P^n_{Cijx} \wedge A^n_{ikx} \cap P^n_{Bijx} = \varnothing)$*, i.e.* $\exists A^n_{ikx}.(\phi^n_{ijx}(P^n_{Cijx}) = K, \forall \delta p^n_{ihx}(t) \mid p^n_{ihx}(t) \in A^n_{ikx})$*. For an axiom instance* A^n_{ikr}*, the set of dependent properties is:* $\widehat{A}^n_{ikx} = \{\widehat{p}^n_{ijx}(t) \mid \widehat{p}^n_{ijx}(t) \in A^n_{ikx}\}$*, the set of independent properties is:* $\widetilde{A}^n_{ikx} = \{\widetilde{p}^n_{ijx}(t) \mid \widetilde{p}^n_{ijx}(t) \in A^n_{ikx}\}$.

Thus, a property value is either dependent or independent of a particular axiom instance. The constancy of a property is determined by the nature of the system, for example, the mass of an object in F=ma. Sometimes constancy may arise following an irreversible prior stage, such as the aggregation of charge at a capacitor. This suggests the following definition, which is useful when discussing the dynamic structure of causal networks:

Definition of Relative Causal Aggregation Ratio. (D-RCAR)

The relative causal aggregation ratio λ^n_{ikx} *refers to the current ratio of the number of relative independent properties to the number of relative dependent properties for a given instance of an axiom:* $\lambda^n_{ikx}(t) = |\widetilde{A}^n_{ikx}| / |\widehat{A}^n_{ikx}|$.

The ratio expresses the degree of causal determination for the hyperedges in a process. Thus,

$\lambda^n_{ikx}(t) = 1$: Equal number of dependent and independent properties.

$\lambda^n_{ikx}(t) < 1$: Highly dependent system (tightly coupled).

$\lambda^n_{ikx}(t) > 1$: Highly independent system (loosely coupled).

$\lambda^n_{ikx}(t) \rightarrow 0$: Completely dependent system.

$\lambda^n_{ikx}(t) \rightarrow \infty$: No interaction, completely independent systems.

Changes in the ratio for a particular hyperedge correspond to changes in the causal structure of the process:

Definition of Approximately Non-Causal Property. (D-ANCP)

Given a process $\pi^n_{ijx}(t)$ *over a period T, a property is an approximately non-causal property,* $p^n_{ijx}(T) \equiv \dot{p}^n_{ijx}(T)$, *if, for some* ε*, a change of* $(| \delta p^n_{ijx}(t) | < \varepsilon, \forall t \in T)$ *results in no other property changes* $(\delta p^n_{ihx\varepsilon}(t) = 0, \forall p^n_{ihx\varepsilon} \in P^n_{Nijx}, \forall t \in T)$, *where the additional subscript* ε *denotes this hypothetical context.*

In other words, an approximately non-causal property has no effective causal influence on its connected properties over the given time interval. This alters the aggregation ratio for any instance of an axiom covering the property.

Definition of Variation in Causal Aggregation Ratio. (D-VCAR)

For a system S_i *at a level* l^n_i *and an instance of an axiom* A^n_{ikx}*, the causal aggregation ratio varies over a period T,* $\delta\lambda^n_{ikx}(T)$ *if:*

$$\exists p^n_{ijx} . (p^n_{ijx} \in A^n_{ikx} \wedge p^n_{ijx}(T_1) \equiv \dot{p}^n_{ijx}(T_1) \wedge p^n_{ijx}(T_2) \not\equiv \dot{p}^n_{ijx}(T_2)),$$

where $T_1 \subset T$, $T_2 \subset T$, $T_1 \neq T_2$.

Definition of Causal Aggregation Tuple. (D-CAT)

The current causal structure G^n_i *of a system* S_i *at a level* l^n_i *at a time t, can be represented by its corresponding tuple of causal aggregation ratios:* $\Lambda^n_i(t) = \langle \lambda^n_{ijx}(t) \mid \forall A^n_{ijx} \in H^n_{ix} \rangle$.

This leads to the condition for a system to have a dynamic or static structure:

Definition of Dynamic & Static Causal Structure. (D-DSCS)

The causal structure G^n_i *of a system* S_i *at a level* l^n_i *over a period T, is dynamic,* $\vec{G}^n_i(T)$*, if there exists* $t_1 \in T$, $t_2 \in T$, $t_1 \neq t_2$, *such that* $\Lambda^n_i(t_1) \neq \Lambda^n_i(t_2)$. *Conversely, the system has a static structure,* $\overleftrightarrow{G}^n_i(T)$*, if for all* $t_1 \in T$, $t_2 \in T$, $\Lambda^n_i(t_1) = \Lambda^n_i(t_2)$.

The next step is to clarify the conditions under which the causal structure may be simplified:

Definition of Collapsible Aggregations. (D-CA)

In a system S_i *at a level* l_i^n*, a subgraph* g_{ij}^n *within a causal structure* G_i^n *may be substituted, via an operation* χ_i^n*, by a single hyperedge equivalent provided it preserves all designated properties.*

This definition hinges on which properties have been designated as significant and implies causal rates of interaction are maintained. Consequently, a number of subgraphs may be recursively substituted up to some degree. Substitution is analogous to algebraic simplification. A corollary is that a degree of causal indirectness, such as a subprocess, may exist between the observed interaction of properties.

Definition of Weak Structural Equivalence. (D-WSE)

For an observer O_h^k*, from a first person perspective, in any two systems* S_i *and* S_j*, at levels* l_i^m *and* l_j^n*, and for a designated set of property instances* P_{hx}^n*, a causal structure* G_j^n *is structurally equivalent to* G_i^m*, if it can be collapsed to an isomorphic structure,* $G_i^m \equiv \chi_j^n(G_j^n)$*, that preserves the properties:* $P_{jx}^n \equiv P_{ix}^m$*, where* $(k>m,h=i) \wedge (k>n,h=j)$*.*

The next theorem brings together a number of the definitions that together imply the causal structure of a continuous process can be approximated by a discrete process:

Theorem for Weak Structural Continuous-Discrete Equivalence. (T-WSCDE)

A discrete process $\ddot{\pi}_j^n(t)$ *is weakly relative measurable dynamically equivalent to a continuous process* $\bar{\pi}_i^m(t), \bar{\pi}_i^m(t) \equiv \ddot{\pi}_j^n(t)$*, if there exists a maximal base time interval* $\tau_j^n > 0$*, and the processes are weakly structurally equivalent,* $G_i^m \equiv \chi_j^n(G_j^n)$*.*

Proof: this follows from definitions D-WRMDE, D-LMBTI and D-WSE. In other words, the causal structure of certain continuous processes can be reproduced on a suitable digital machine provided timing demands are met.

In summary, the definitions suggest that a number of conditions apply to comparing systems. These centre on system level equivalence, dynamical equivalence and structural equivalence. Common to these is the perspective of the observer, the fidelity of their measurements and the properties of interest. A distinction was made between strong and weak equivalence, where strong equivalence demands all sublevels are equivalent, while weak equivalence only applies to the levels concerned.

Weak equivalence essentially requires an appropriate isomorphism exist between the levels.

6 Defining Function & Representation

This section is motivated by the need to distinguish two types of function: functional-maps versus functional-flows, and two types of representation: stationary versus temporal.

6.1 Defining Functional-Map

The term *functional-map* refers to the conventional mathematical sense of function as a relation between two sets or spaces. Given a domain X and a codomain Y, a function f associates an element of X with an element of Y: y = f(x). The function f is then defined as the set of ordered pairs $\langle x, f(x)\rangle$. A map is a function and a given codomain. Thus, a function maps an element from a domain to a corresponding element in a codomain – see E.J.Borowski [Boro89], Cohn [Cohn93] and Swokowski [Swok79]. This type of function will be denoted by the normal function symbol, or a capital 'M' subscript when emphasis is required, $f_M : X \rightarrow Y$, or $f_M : x \mapsto y$.

6.2 Defining Static-Functional-Flows

The evolution of a system can be described by a temporal sequence of state values, i.e. a trajectory through its state space. This is commonly called a *flow* and represented by the flow function $\phi(t, x)$ which gives the state value at time t given the present state x – see Arrowsmith and Place [Arro92], and Norton [Nor95]. The flow function covers the space of permissible trajectories, which are determined by the systems structure. Representing this in terms of functional-maps involves a forward map f, and a backward map g, such that the next state $y(t+\delta t) = f(x(t), g(y(t)))$. Effectively, the system has a memory of its prior state. An example of a discrete version of such a system is the finite-state-machine (FSM) mentioned earlier, but now the emphasis is on the trajectories of the system rather than its state.

When embedded in a larger system the flow of a subsystem will have a particular causal profile (i.e. perform a particular role) as it mediates between other subsystems. Thus, in contrast to a functional-map, the term *functional-flow* is used to denote the temporal profile of the flow of a subsystem and its impact in a system, which is represented by the equation given above for $y(t+\delta t)$. When the maps are static structures over a period T, this is called a static-functional-flow over T. This type of function will be denoted by the function symbol embellished by a double arrow, or a capital 'F' subscript for emphasis, $\overleftrightarrow{f}_F : X, T \rightarrow Y$, or $\overleftrightarrow{f}_F : x, t \mapsto y$. Note the extra time dimension compared to f_M.

6.3 Defining Dynamic-Functional-Flows

A dynamic-functional-flow is a functional-flow that is itself changing. In terms of the forward and backward maps mentioned above, their structure changes over a period T. This can be modelled as a second order continuous state machine. A conventional finite state machine has a static structure, which determines the state transitions. In some systems the state structure may be determined through a process of adaptation – see Aleksander [Alek96]. In contrast, a second order state machine is not adapting, but *continually* evolving, the change in its state structure is a continuous function with respect to time. In some cases it may be able to adapt, as well, in the conventional sense. Representing a second order state machine involves partitioning its functional structure such that the transfer functions, i.e. the maps f and g, are themselves functions of the current state, $f = F(x(t))$, $g = G(x(t))$. Notice that this introduces constraints on the structure of the state space, and hence the flows achievable by the system. This type of function will be denoted by the function symbol embellished by a single arrow, or a capital 'FD' subscript for emphasis, $\vec{f}_{FD} : X, T \rightarrow Y$, or $\vec{f}_{FD} : x, t \mapsto y$.

6.4 Defining Stationary & Temporal Representations

Following from the definition of functional-maps and functional-flows, the term *stationary-representation* refers to any relational structure that acts as a representation for a system at a particular instance in time. Conversely, the term *temporal-representation* refers to a changing relational structure whereby the change acts as a representation for a system over a period of time. A painting is an example of the first kind, and a symphony an example of the latter.

Definition of System Representation. (D-SR)

A representation R^n_{ijx}, *at a level* l^n_i *in a system* S_i *is a tuple of property instances,* $R^n_{ijx} = \langle \psi^n_{ijx}(t) \mid x \in X,\ C^n_{ijx}(t) \models C^n_i \rangle$, *where the set X is determined by the form of the representation. An interpretation* I^m_k *of the representation will only be valid for an observer from a given first or third person perspective.*

Suppose $R^n_{ijx}(t)$ *represents a changing relational structure. This is called a stationary-representation, denoted* $R^n_{Sijx}(t)$, *to an observer under an interpretation* I^m_k *if, for some* δ, $R^n_{ijx}(t \pm \delta t) \neq_{I^m_k} R^n_{ijx}(t)$, *and* $dR^n_{ijx}(t)/dt \approx 0$ *over the open interval* $(t-\delta t, t+\delta t)$. $R^n_{ijx}(t)$ *acts as a temporal-representation to an observer over the period T, denoted* $R^n_{Tijx}(T)$, *iff,* $\Box\ R^n_{ijx}(t_1) \neq_{I^m_k} R^n_{ijx}(t_2)$, *for some* $t_1 \in T$ *and* $t_2 \in T$.

More succinctly, a stationary-representation can be defined at an unconnected point in time, whereas a temporal-representation is defined as a flow over a period of time. A stationary-representation can be cast in terms of Ramsey sentences, whereas a temporal-representation cannot – see Davies [Dave98, p262]. The qualifiers static and dynamic will be prefixed to these terms when dealing with particular types of functional-flows.

7 Dynacept Mood Detector Example Programme Listings

For communications theory used to simulate neuron connectivity on a multiprocessor see D.Whobrey [Who89].

7.1 Dface.h

```
/* NeSeL v2.0 April 1989 */
/* DFace neural net header file */

#define MAX_WIDTH 127 /* max pulse width */
#define PKT_EMPTY ((PktHnd)(0))
/* signifies No-Comms-Packet on channel */
#define RT2PIE 2.50662828 /* sqrt(2*pi) */

#define NUM_FEAS 11 /* num components to face */
#define NUM_EXPS 5
/* num expressions made from state of features */
#define NUM_ESIM 4 /* num simple emotions */
/* The feature detectors */
#define F_NOSE_BRIDGE 0
#define F_NOSE_BOTTOM 1
#define F_MOUTH_LSIDE 2
#define F_MOUTH_LIPS 3
#define F_MOUTH_RSIDE 4
#define F_EYE_LEFT 5
#define F_EYE_RIGHT 6
#define F_BROW_LEFT_L 7
#define F_BROW_LEFT_R 8
#define F_BROW_RIGHT_L 9
#define F_BROW_RIGHT_R 10
/* Note the following double as conx indices */
/* The static expressions */
#define E_MELLOW 0
#define E_HAPPY 1
#define E_ANGRY 2
#define E_SAD 3
#define E_SQUINT 4
/* The dynamic expressions */
#define D_LAUGH 0
#define D_FROWN 1
#define D_CRY 2
#define D_WINK 3

/* Internal code to distinguish feature, expression & emotion
nodes */
#define ND_FEA 0
#define ND_EXP 1
#define ND_EMO 2
#define ND_ALT 3
#define ND_PRO 4

/* Position of features on face */
#define F_POSITION_X 45,45,35,35, 0,25,75,30,30,70,70
#define F_POSITION_Y 60,60,85,85, 0,30,30,20,20,20,20
#define F_LENGTH   30,10,10,30,10,10,10,10,10,10,10

/* Feature angles that define the static expressions */
#define F_MELLOW 1,32,  96,32,32, 32,96, 96,32,96,32
#define F_HAPPY  10,32,112,32,16, 32,96, 80,48,80,48
#define F_ANGRY  5,24, 80,32,48, 40,88, 90,55,72,38
#define F_SAD    16,32, 85,32,43, 16,112,90,24,102,40
#define F_SQUINT 8,32, 115,32,10, 20,96, 110,48,80,48
```

7.2 Dface.c

```
/* NeSeL v2.0 April 1989. DFace neural net */
#include <stdio.h>
#include <string.h>
#include <ctype.h>
#include <malloc.h>
#include <stdarg.h>
#include <math.h>
#include "nsltype.h"
#include "nslstd.h"
#include "nslios.h"
#include "nslwsio.h"
#include "dface.h"
#include "dface.nsh"

/* ------------------ COMMS routines ----------------- */
/* Normal Distribution Look-Up table parameters */
#define DIST_X 32
#define DIST_Y 100
#define DIST_H (DIST_Y/2)
#define DIST_M (DIST_H-1)
#define DIST_D (DIST_Y-1)
/* Look-up table for normal distribution for comms delay */
int dist_table[DIST_X][DIST_Y+1]={0};

/* Work out quick look up table for path length normal
distribution */
void calc_dist(int len,int max)
{ float cc,mm,kk,ss,x1,x2,xx,dd,sum=(float)0.0,aa;
  int ff,bb,nn,no,iy,ix=1;
  dist_table[len][0]=(int)max;
  x1=(float)len; x2=(float)max; nn=(int)x1;
  x1-=(float)0.5; x2+=(float)0.5;
  mm=(x1+x2)/(float)2.0;
  dd=(float)0.02; ff=50;
  if(len<1) goto skip;
  ss=(float)len/com_sigma_coeff;
 /* approximation for sigma */
  kk=(float)2.0*ss*ss;
  cc=ss*(float)RT2PIE;
  for(xx=x1;xx<x2;xx+=dd) sum+=Gauss(xx,kk,mm,cc);
  xx=x1;
  while(ix<DIST_H)
   {iy=ff; aa=(float)0.0;
    while(iy--) {aa+=Gauss(xx,kk,mm,cc); xx+=dd;}
    iy=(int)(aa*(float)DIST_D/sum);
    if(iy<1) iy=1;
    if(xx+dd>mm) iy=DIST_H;
    while(iy--&&ix<DIST_H) dist_table[len][ix++]=(int)nn;
    ++nn;
   } skip:
  nn=(int)dist_table[len][DIST_M];
dist_table[len][DIST_H]=(int)nn;
  ix=DIST_H; iy=DIST_H+1; no=nn; bb=nn;
  if(bb<max) {if(len%2) ++bb; else if(max%2) ++bb;}
  while(iy<DIST_Y)
   {nn=(int)dist_table[len][ix];
    if(nn!=no) {if(bb<max) bb++; no=nn;}
    dist_table[len][iy]=(int)bb;
    ++iy; --ix;
   }
  for(iy=1;iy<DIST_Y;++iy) {if(dist_table[len][iy]) --
dist_table[len][iy];}
}

/* Allocate mem for comm channel: pick random vals for path
len.
  sigma=SL/coeff, S=packet size=1, L=min path len, M=max
len.
  range=6sigma: L..M, [L>=com_min,M<=com_max].
Formula: M=L+6*L/coeff.
*/
void com_mem(packet_in_synapse *dp)
{ pulse *apt; int L,M;
  float vary=(float)1.0+(float)6.0/com_sigma_coeff;
  /* work out max L for current com_sigma_coeff */
  L=(int)((float)com_max/vary); if(L<1) L=1;
  /* work out new path min */
  L=1+L-com_min; L=com_min+(int)rnd(L)-1;
  /* now get the new path max for this min len */
  M=(int)(L*vary); /* L + 6 sigma */
  dp->len=(int)L; dp->max=(int)M;
  if(!dist_table[L][0]) calc_dist(L,M);
  ++M; /* plus one for terminal PKT_EMPTY slot */
  if(!(apt=(pulse *)malloc(sizeof(pulse)*M)))
    {printf("no space\n"); w_exit(0);}
  dp->delay=apt;
  while(M--) (*apt++)=PKT_EMPTY;
}

/* free memory used for comms delay */
void com_free(packet_in_synapse *dp)
{ if(dp->len) {free(dp->delay); dp->delay=NULL;}
}

typedef struct
{    int m_size;
     int m_users;
     int m_locked;
     int m_untouched;
} PktData;

PktData **PacketTable=NULL;
int PacketTableSize=20000; /* number of comm cells i.e. L *
num_net_conxs */
int PacketMaxHds=0;
#define PacketIsFree(p) (!((p)->m_locked) && !((p)-
>m_users) && !((p)->m_untouched))
#define PacketInit(p) {(p)->m_users=0; (p)->m_locked=1;
(p)->m_untouched=1;}

PktHnd PacketAlloc(int nsize)
{    PktHnd hd; PktData *p;
     if(!PacketTable)
     {    PacketTable=
(PktData **)malloc(sizeof(PktData *)*PacketTableSize);
for(hd=0;hd<PacketTableSize;++hd)
PacketTable[hd]=NULL;
     }

     for(hd=1;hd<=PacketMaxHds;++hd)
     {    p=PacketTable[hd];
          if(p && (p->m_size==nsize) && PacketIsFree(p))
          {    PacketInit(p);
               return hd;
          }
     }

     for(hd=1;hd<PacketTableSize;++hd)
     {    if(!PacketTable[hd])
          {    if(hd>PacketMaxHds) ++PacketMaxHds;
               p=(PktData *)malloc(sizeof(PktData)+nsize);
               PacketTable[hd]=p;
               p->m_size=nsize;
               PacketInit(p);
               return hd;
          }
     }
     return 0;
}

void PacketFree()
{    if(PacketTable)
     {    PktHnd hd;
          for(hd=1;hd<=PacketMaxHds;hd++)
```

```
			if(PacketTable[hd]) free(PacketTable[hd]);
		free(PacketTable);
		PacketTable=NULL; PacketMaxHds=0;
	}
}

PktHnd PacketValid(PktHnd hd)
{	if(PacketTable && (hd>0) && (hd<=PacketMaxHds)
&& PacketTable[hd]
&& !PacketIsFree(PacketTable[hd])) return hd;
	return 0;
}

void *PacketDataPt(PktHnd hd)
{	PktData *p; PktHnd old=hd;
	if(!(hd=PacketValid(hd))) return NULL;
	p=PacketTable[hd]; ++p;
	return (void *)p;
}

void PacketAccess(PktHnd hd)
{	if(!PacketValid(hd)) return;
	++(PacketTable[hd]->m_users);
	PacketTable[hd]->m_untouched=0;
}

void PacketRelease(PktHnd *phd)
{	PktHnd hd=PacketValid(*phd);
	if(hd) --(PacketTable[hd]->m_users);
	*phd=0;
}

void PacketUnlock(PktHnd *phd)
{	PktHnd hd=PacketValid(*phd);
	if(hd) PacketTable[hd]->m_locked=0;
	*phd=0;
}
/* Get input packet. This models comms delay.
*  Propagate signal down connection channel
*  All inputs are strobed every cycle.*/
PktHnd PacketGet(NSLNSLnet *net,packet_in_synapse *dp)
{	PktHnd *fp,hd=inx(PktHnd,dp->input);
	PacketAccess(hd);
	if(com_off||com_max<2)
{PacketRelease(&dp->last); dp->last=hd;}
	else
	{	int nn;
		if(!dp->len) com_mem(dp);
		nn=dp->max; fp=dp->delay;
		while(nn--) {*fp=(*(fp+1)); ++fp;}
/* last slot always empty */
		nn=(int)dist_table
[dp->len][1+(int)(rndf()*(float)DIST_D)];
		fp=dp->delay+nn;
		while(*fp!=PKT_EMPTY) ++fp;
		*fp=hd;
		if(*(fp=dp->delay)!=0)
		{	PacketRelease(&dp->last); dp->last=(*fp);}
	}
	return dp->last;
}

pulse PulsePacketGet(PktHnd hd)
{	pulse *p=(pulse *)PacketDataPt(hd);
	if(p) return *p;
	return 0;
}

pulse PeekPulse(packet_in_synapse *p)
{	return(PulsePacketGet(p->last));
}

/* --------------- FACE GUI routines ------------------ */
/* Convert pulse val to angle */
int val2deg(pulse val)
{ long ang=360*(long)val; ang/=(1+MAX_WIDTH);
 return((int)ang);
}

/* draw a feature */
void face_draw_feature(fio_ds *fp,feat_ds *ap,feat_ds *bp)
{ int w=fp->scr_f,x,y,a; long xy; pulse val=ap->output_value;
 w_penchar(w,0);
 if(ap->oldpos.x) /* erase old line */
  {w_position(w,ap->oldpos.x,ap->oldpos.y);
   w_pencolor(w,wnd_white); w_pendown(w,1);
   w_turnto(w,ap->oldang); w_move(-w,ap->fealen);
  }
 if(ap->feapos.x)
  {x=ap->feapos.x; y=ap->feapos.y;}
 else {x=bp->nxtpos.x; y=bp->nxtpos.y;}
 if(x)
  {a=val2deg(val);
   w_position(w,x,y);
   w_pencolor(w,wnd_black); w_pendown(w,1);
   w_turnto(w,a); w_move(-w,ap->fealen);
   ap->oldang=a; ap->oldpos.x=x; ap->oldpos.y=y;
   xy=w_position(w,(int)-1,(int)-1);
   ap->nxtpos.x=w_x(xy); ap->nxtpos.y=w_y(xy);
  }
}

/* Convert pulse val to ascii, for display purposes */
char potasc(pulse val)
{ if(val<1) return('A');
 val/=(pulse)6; return((char)('B'+val));
}

/* Convert val in [0..1] to ascii */
char intasc(float val)
{ if(val<=0.0) return('A');
 if(val>=1.0) return('Z');
 val*=(float)26.0; return((char)('A'+(int)val));
}

#define NOO_OFS 1 /* wnd y offset for state vals */
#define NOO_WID 16 /* width of state dsp column */
/* State display column offsets, for formatting purposes */
int col_ofs[]={1,10,30,50};
/* Annotations for the various moods */
char *face_var[]={
"Mellow","Happy","Angry","Sad","Squint","Laugh",
"Frown","Cry","Wink","Crying","Winking","Neither"
};

/* Display state of Expression & Emotion nodes */
void face_state(NSLNSLnet *net,void *np,int ntyp)
{ long num,adjnum; float aa; int
ww,ax,by=NOO_OFS,yd,xx,nn;
 fio_ds *fp; feat_ds *lp; expr_ds *op; Production_ds *rp;
 fp= &(dsnp(nsl_handle(NULL,"\\face_io_node"),
face_io_ds)->fio);
 ww=fp->scr_s;
 switch(ntyp)
  {case ND_EXP: op=(expr_ds *)np;
    nn=op->size.input; num=op->node.name;
adjnum=num-fp->b_feat; break;
   case ND_PRO: rp=(Production_ds *)np;
    nn=(int)rp->state.num_elements; num=rp->node.name;
      adjnum=NUM_FEAS+NUM_EXPS;
      break;
   case ND_FEA: lp=(feat_ds *)np; num=lp->node.name;
      adjnum=num-fp->b_feat; break;
     case ND_ALT: /* TO DO */
     default: return;
  }
 if(dsp_flag)
  {xx=(int)adjnum; if(num==fp->b_emto) ++xx;
   yd=1+w_y(w_size(ww,TRUE))-NOO_OFS;
   ax=col_ofs[xx/yd]; by+=xx%yd;
```

```
   w_moveto(-ww,ax,by); w_printf(-ww,"%ld",num);
   switch(ntyp)
    {case ND_EXP:
            for(xx=0;xx<nn;++xx)
               w_printf(-ww,"%c",
potasc(PeekPulse(&(op->input[xx].input))));
        w_printf(ww,"%c ",potasc(op->output_value));
            break;
     case ND_PRO:
            for(xx=0;xx<nn;++xx)
             w_printf(-ww,"%c",
intasc(rp->state.p_elements[xx]));
            break;
     case ND_FEA:
            w_printf(ww,"%c ",potasc(lp->output_value));
            break;
       case ND_ALT: /* TO DO */
       default: ;
    }
  }
 ww=fp->scr_v; by=1+(int)adjnum-NUM_FEAS;

 switch(ntyp)
  {case ND_EXP:
           aa=((float)op->output_value)/(float)MAX_WIDTH;
           w_position(ww,1,by); w_clear_line(-ww);
      w_printf(ww,"%s=%g",face_var[by-1],aa);
      break;
    case ND_PRO:
       if(num==fp->b_emto) by+=NUM_ESIM;
       for(xx=0;xx<nn;++xx)
       {  aa=rp->state.p_elements[xx];
         w_position(ww,1,by+xx); w_clear_line(-ww);
               w_printf(ww,"%s=%g",face_var[by-1+xx],aa);
       }
     break;
      case ND_FEA: case ND_ALT: /* TO DO */
    default:;
   }
}

/* Display general information i.e. permissible driver patterns
*/
void face_info(fio_ds *fp)
{ int w=fp->scr_i;
 w_clear(w);
 w_printf(w,"driver(\"pattern\",x):\n 0=mellow 1=happy
2=angry 3=sad 4=squint");
}

/* Clean the face */
void face_clean(fio_ds *fp,va_list args)
{ BOOL refresh;
 refresh=va_arg(args,BOOL);
 w_clear(fp->scr_f); w_clear(fp->scr_v);
 face_info(fp);
 if(refresh) {w_refresh(fp->scr_f,TRUE);
w_refresh(fp->scr_v,TRUE);}
}

BOOL face_setup(NSLNSLnet *net,fio_ds *fp)
{ BOOL dumy=FALSE; int w; char *ap,buff[100];
 sprintf(buff,"%s_",net->ins_name); ap=buff+strlen(buff);
 fp->b_feat=1;
 strcpy(ap,"state");
 fp->scr_s=
w_open(buff,TRUE,TRUE,FALSE,TRUE,FALSE,
40,8,wnd_black,wnd_white);
 strcpy(ap,"info");
 fp->scr_i=
w_open(buff,TRUE,TRUE,FALSE,TRUE,FALSE,
40,4,wnd_black,wnd_white);
 strcpy(ap,"values");
 fp->scr_v=
w_open(buff,TRUE,TRUE,FALSE,TRUE,FALSE,
16,12,wnd_black,wnd_white);
 strcpy(ap,"dface");
 fp->scr_f=
w_open(buff,TRUE,TRUE,FALSE,FALSE,FALSE,
100,110,wnd_black,wnd_white);
 fp->scr_t=fp->scr_s;
 w_title(fp->scr_f,NIO_TL,NIO_DRAW,FALSE,
"Mr Moody");
 w_title(fp->scr_s,NIO_BL,NIO_DRAW,FALSE,
"DFace Demo");
 w_title(fp->scr_i,NIO_BL,NIO_DRAW,FALSE,
"DFace Info");

 /* reposition windows */
 w=fp->scr_s; if(w>=0) {w_origin(w,770,440);
w_resize(w,336,140);}
 w=fp->scr_i; if(w>=0) {w_origin(w,750,290);
w_resize(w,336,84);}
 w=fp->scr_v; if(w>=0) {w_origin(w,870,30);
w_resize(w,144,182);}
 w=fp->scr_f; if(w>=0) {w_origin(w,640,30);
w_resize(w,192,180);}

 face_clean(fp,(va_list)&dumy);
 return(TRUE);
}

/* table holding feature angles for the expressions */
pulse moods[NUM_EXPS][NUM_FEAS]={
{F_MELLOW},
{F_HAPPY},
{F_ANGRY},
{F_SAD},
{F_SQUINT}
};
/* Put expression onto face */
int face_pattern(NSLNSLnet *net,fio_ds *fp,va_list args)
{ int xx,nn;
 nn=(int)va_arg(args,int);
 for(xx=0;xx<NUM_FEAS;++xx)
   dsx(pulse,fp->b_feat+xx,feat_ds,next)=moods[nn][xx];
 return(0);
}

/* Update face after a cycle or when initialising */
int face_update(NSLNSLnet *net,fio_ds *fp,va_list args)
{ int xx; feat_ds *ap=NULL,*bp;
 sprintf(fp->buff,"%ld",net->time);
 w_title(fp->scr_t,NIO_TR,NIO_DRAW,FALSE,fp->buff);
 for(xx=0;xx<NUM_FEAS;++xx)
 {  bp=ap; ap=dsnp(fp->b_feat+xx,feat_ds);
    face_draw_feature(fp,ap,bp);
 }
 w_update(fp->scr_t); w_update(fp->scr_f);
 w_update(fp->scr_s);
 return(0);
}

/* Perform any user driver commands */
int face_user(fio_ds *fp,va_list args)
{ return(0);
}

#include "dface.nsc"
#include "dface.nsd"

int NSLmain(int argc,char *argv[])
{    nsl_mon_wnds();
     nsl_install(&network,"aux_dface");
     /*nsl_driver(&network,NSLIO_SETUP,(char *)0L);*/
     nsl_monitor(
"printf(\"For demonstration try: call dface;\\n\");"
);
     return(0);
}
```

7.3 Dface.n

```
/* NeSeL v2.0 April 1989. DFace neural net */
/* Compile with nsl1 options -aqvmh; and no options to nsltox
*/
pragma header 0
#include <stdio.h>
#include <string.h>
#include <ctype.h>
#include <malloc.h>
#include <stdarg.h>
#include <math.h>

#include "nsltype.h"
#include "nslstd.h"
#include "nslios.h"
#include "nslwsio.h"

#include "dface.h"
pragma header -1

#define COM_OFF 1 /* set to 1 to turn comms off */
#define COM_MIN 1 /* minimum comm path length */
#define COM_MAX 10 /* maximum comm length */

#define MAX_DIFF (float)MAX_WIDTH
#define DEF_MEM 0
/* default value for dendrite memories */
#if(COM_OFF==1) /* comms off */
#define DEF_RATE 1.0
/* default rate of change of state wrt dir vector */
#define DEF_DECAY 0.9 /* default rate of decay of state */
#define DEF_RATE_E 0.8
/* default rate for dynamic emotions */
#define DEF_DECAY_E 0.9
/* default decay for dynamic emotions */
#endif
#if(COM_OFF!=1) /* comms on */
#define DEF_RATE 0.95
#define DEF_DECAY 0.95
#define DEF_RATE_E 0.8
#define DEF_DECAY_E 0.95
#endif

/* Alteration defaults */
float altsg=(float)12.0,altsc=(float)-0.7;
 /* transfer sigmoid parameters */
#define GP_SIG ((float)2.3548)
/* gauss pulse sigma coeff=2.0*(sqrt(2.0*log(2.0))) */
#define DEF_ALT_P ((float)4.0) /* number of groups */
#define DEF_ALT_Q ((float)3.0)
/* number of elements per group */
#define DEF_ALT_A GROUP_INIT(DEF_RATE,0)
/* rate of change for state elements */
#define DEF_ALT_B GROUP_INIT(DEF_DECAY,0)
/* rate of decay for state elements */
#define DEF_ALT_W GROUP_INIT(0.5,0) /* initial state */
#define DEF_ALT_DW GROUP_INIT(0.1,0)
 /* initial direction */
#define L_FEAS 0..(NUM_FEAS-1) /* feature list range */

float exsg=(float)100.0,exsc=(float)-0.08;
/* expression sigmoid parameters */
float mood_rate=(float)0.125;
/* rate at which to vary feature angles */
int mood_period=2;
/* multiplcation factor for cycles between moods */
/* comms parameter for normal distribution */
int
com_off=COM_OFF,com_min=COM_MIN,com_max=COM
_MAX,dsp_flag=1;
float com_sigma_coeff=(float)12.0;

typedef int pulse;
typedef int PktHnd;
/* group vectors are relayed via handles */
/* Point co-ordinates for features on face */
typedef struct
{ int x,y;
} coors;

typedef struct
{ float default_value;
  float num_elements;
  int max_elements;
  float *p_elements;
} group_ds;

#define GROUP_INIT(a,b) {(float)a,(float)b,0,(float *)0L}

node
{ int len,max;
  PktHnd last,*delay,*input;
} packet_in_synapse={0,0,0,0L};

node
{ pulse mem; /* memorised value c.f. weight */
  packet_in_synapse input;
} dendrite={DEF_MEM};

node
{ pulse output_value; PktHnd output;
  pulse next; float state;
  int fealen,oldang;
  coors feapos,oldpos,nxtpos;
  NSLNSLio args;
} feat_ds={0,0,MAX_WIDTH+1,(float)-
1.0,0,0,{0,0},{0,0},{0,0}};

node
{ pulse output_value; PktHnd output;
  dendrite input[];
  NSLNSLio args;
} expr_ds={0,0};

node
{ int scr_t, scr_f, scr_v, scr_i, scr_s;
  long b_feat,b_expr,b_emto;
  char buff[20];
} fio_ds={buff={0}};

node
{ fio_ds fio;
  NSLNSLio args;
} face_io_ds;

node
{ PktHnd output;
  packet_in_synapse input[];
  NSLNSLio args;
} ConxToGroup_ds={0};

node
{ PktHnd output;
  group_ds **glist;
  packet_in_synapse input[];
  NSLNSLio args;
} AddGroups_ds={0,0L};

node
{ int member;
  PktHnd output;
  packet_in_synapse input;
  NSLNSLio args;
} FilterElement_ds={0,0};
```

```
node
{ group_ds state;
 PktHnd output;
 packet_in_synapse input[];
 NSLNSLio args;
} Reduction_ds={GROUP_INIT(0.0,0.0),0};

node
{ float p; /* number of groups (adaptable) */
 float q; /* number of elements per group (adaptable) */
 group_ds A; /* rate of change for state elements */
 group_ds B; /* rate of decay for state elements */
 group_ds W; /* current state position */
 group_ds dW;
/* direction vector in which to move state position */
 PktHnd output;
 packet_in_synapse input;
 NSLNSLio args;
}
Alteration_ds={DEF_ALT_P,DEF_ALT_Q,DEF_ALT_A,D
EF_ALT_B,DEF_ALT_W,DEF_ALT_DW,0};

node
{ group_ds state;
 PktHnd output;
 packet_in_synapse input;
 NSLNSLio args;
} Production_ds={GROUP_INIT(0.0,0.0),0};

extern void PacketFree();
extern void *PacketDataPt(PktHnd hd);
extern void PacketAccess(PktHnd hd);
extern void PacketRelease(PktHnd
*phd),PacketUnlock(PktHnd *phd);
extern PktHnd PacketAlloc(int nsize),
PacketGet(NSLNSLnet *net,packet_in_synapse *dp);
extern pulse PulsePacketGet(PktHnd hd);

/* returns true if arg is positive or equals zero */
BOOL ispos(float a)
{ return((BOOL)(a>=(float)0.0));
}

/* Convert pulse val to be in interval [0,1] */
float Pulse2Norm(pulse val)
{ return (float)val/(float)(MAX_WIDTH);
}

float Gauss(float xx,float kk,float mm,float cc)
{ xx-=mm; xx*=xx; xx/=kk;
 return((float)exp((double)(-xx))/cc);
}

float GaussPulse(float i, float j, float n, float p, float ss)
{    i/=n; j/=p;
     i-=j; i/=ss; i*=i; i*=(float)(-0.5);
     return((float)exp((double)i));
}

/* Save a pulse value in a packet */
void PulsePacketAlloc(PktHnd *phd, pulse n)
{    PktHnd hd;
     PacketUnlock(phd);
     hd=PacketAlloc(sizeof(pulse));
     *phd=hd;
     *((pulse *)PacketDataPt(hd))=n;
}

group_ds *GroupPacketGet(PktHnd hd)
{    return (group_ds *)PacketDataPt(hd);
}

/* Alloc memory for each group element */
void GroupPacketAlloc(PktHnd *phd,int nn)
{    PktHnd hd; group_ds *p;
     PacketUnlock(phd);
     hd=PacketAlloc(sizeof(group_ds)+nn*sizeof(float));
     *phd=hd;
     p=(group_ds *)PacketDataPt(hd);
     p->p_elements=(float *)(p+1);
     p->max_elements=nn;
     p->num_elements=(float)nn;
     p->default_value=(float)0.0;
}

/* Alloc space for group via packet table & init values */
group_ds *GroupAlloc(group_ds *p, int nn)
{    PktHnd hd=0; group_ds *q; int i;
     GroupPacketAlloc(&hd,nn);
     q=(group_ds *)PacketDataPt(hd);
     if(p)
     {    p->p_elements=q->p_elements;
          p->max_elements=q->max_elements;
          p->num_elements=q->num_elements;
          q=p;
     }
     for(i=0;i<nn;i++) q->p_elements[i]=q->default_value;
     return q;
}

face_io_ds face_io_node()
{ switch($args.cmd)
   {case NSLIO_SETUP:
     face_setup(net,&$fio); face_update(net,&$fio,$args.args);
break;
    case NSLIO_INIT: break;
    case NSLIO_CLOSE:
     w_close($fio.scr_f); w_close($fio.scr_i);
     w_close($fio.scr_v); w_close($fio.scr_s);
       PacketFree( ); break;
    case NSLIO_REDRAW: face_clean(&$fio,$args.args);
break;
    case NSLIO_PATTERN:
face_pattern(net,&$fio,$args.args); break;
    case NSLIO_PRE: break;
    case NSLIO_POST: face_update(net,&$fio,$args.args);
break;
    case NSLIO_USER: face_user(&$fio,$args.args); break;
    default:;
   }
 $args.cmd=NSLIO_NULL;
}

long tbug_time; long tbug_node; char *tbug_name;
#define TBUGSET(a) tbug_time=net->time; \
tbug_node=node->node.name; tbug_name= #a;\
w_printf(wnd_std,"%ld,%ld%s\n",\
tbug_time,tbug_node,tbug_name);

/* feature node: vary feature angle at rate */
feat_ds feature()
{ pulse nn=$next,xx=$output_value;
 TBUGSET(feature)
 if($args.cmd!=NSLIO_NULL)
 {  $args.cmd=NSLIO_NULL;
    return;
 }
 if(nn<=MAX_WIDTH && xx!=nn)
  {if($state<(float)0.0) $state=(float)xx;
  $state+=(((float)nn)-$state)*mood_rate;
    $output_value=(pulse)($state+(float)0.5);
  }
 PulsePacketAlloc(&$output,$output_value);
 face_state(net,$,ND_FEA);
}

/* Compare inputs against memory pattern.
* A product error is used rather than, say,
* the angle between vectors,
* so that a single input mismatch isn't ignored.
```

```
* Responds with probability of expression.
*/
expr_ds expression()
{ float dd,err; int xx,nn=$size.input;
 TBUGSET(expression)
 if($args.cmd!=NSLIO_NULL)
 {   if($args.cmd==NSLIO_CLOSE)
         for(xx=0;xx<nn;++xx) com_free(&$input[xx].input);
     $args.cmd=NSLIO_NULL;
     return;
 }
 err=(float)1.0;
 for(xx=0;xx<nn;++xx)
 {
dd=(float)PulsePacketGet(PacketGet(net,&$input[xx].input));
     dd-=((float)$input[xx].mem);
   dd/=MAX_DIFF; if(dd<(float)0.0) dd=(-dd);
   err*=((float)1.0-
generic_sigmoid(dd,(float)1.0,(float)0.0,exsg,exsc));
 }
 $output_value=(pulse)1
+(pulse)((float)(MAX_WIDTH-1)*err);
 PulsePacketAlloc(&$output,$output_value);
 face_state(net,$,ND_EXP);
}

/* maps n inputs to an n dim group output vector */
ConxToGroup_ds ConxToGroup()
{    int xx,nn=$size.input; float *pe;
 TBUGSET(ConxToGroup)
     if($args.cmd!=NSLIO_NULL)
     {    if($args.cmd==NSLIO_CLOSE)
              for(xx=0;xx<nn;++xx) com_free(&$input[xx]);
          $args.cmd=NSLIO_NULL;
          return;
     }
     GroupPacketAlloc(&$output,nn);
     pe=((group_ds *)PacketDataPt($output))->p_elements;
     for(xx=0;xx<nn;++xx)
          pe[xx]=Pulse2Norm(PulsePacketGet(
PacketGet(net,&$input[xx])));
}

/* Append groups into one i.e. form cross-product */
AddGroups_ds AddGroups()
{    int k=0,empty=0,xx,nn=$size.input; group_ds *py;
 TBUGSET(AddGroups)
     if($args.cmd!=NSLIO_NULL)
     {    if($args.cmd==NSLIO_CLOSE)
          {    for(xx=0;xx<nn;++xx) com_free(&$input[xx]);
               if($glist) {free($glist); $glist=NULL;}
          }
          $args.cmd=NSLIO_NULL;
          return;
     }
     if(!$glist)
$glist=(group_ds **)malloc(sizeof(group_ds *)*nn);
     for(xx=0;xx<nn;++xx)
     {    $glist[xx]=
GroupPacketGet(PacketGet(net,&$input[xx]));
          if($glist[xx]) k+=(int)$glist[xx]->num_elements;
else ++empty;
     }
     if(empty)
          GroupPacketAlloc(&$output,0);
/* xmit a null group for now */
     else
     {    int i,j,m=0;
          GroupPacketAlloc(&$output,k);
          py=(group_ds *)PacketDataPt($output);
          for(xx=0;xx<nn;++xx)
          {    j=(int)$glist[xx]->num_elements;
               for(i=0;i<j;i++)
py->p_elements[m++]=
$glist[xx]->p_elements[i];
               }
          }
}

/* Extract one element from group */
FilterElement_ds FilterElement(member)
{    group_ds *pg,*py;
 TBUGSET(FilterElement)
     if($args.cmd!=NSLIO_NULL)
     {    if($args.cmd==NSLIO_CLOSE) com_free(&$input);
          $args.cmd=NSLIO_NULL;
          return;
     }
     pg=GroupPacketGet(PacketGet(net,&$input));
     GroupPacketAlloc(&$output,(int)1);
     py=(group_ds *)PacketDataPt($output);
     py->p_elements[0]=
( (pg&&(((int)pg->num_elements)>$member))
          ?(pg->p_elements[$member]):((float)0.0) );
}

/* Generic Gaussian Group Alteration
* 1) Approximate continuous distrbutions by discrete ones
* 2) Degree of partitioning may be an evolving parameter
*/
Alteration_ds GenericGGAlteration()
{    int i,j,n=0,p=0; float sum,sq,N,*g,*dW,*W,*A,*B;
group_ds *pg,*py;
 TBUGSET(GenericGGAlteration)
     if($args.cmd!=NSLIO_NULL)
     {    if($args.cmd==NSLIO_CLOSE) com_free(&$input);
          $args.cmd=NSLIO_NULL;
          return;
     }
     if(pg=GroupPacketGet(PacketGet(net,&$input)))
     {    g=pg->p_elements; n=(int)pg->num_elements;
          if(n>0)
          {    p=(int)$p;
               if(!$A.p_elements)
               {    GroupAlloc(&$A,p); GroupAlloc(&$B,p);
GroupAlloc(&$W,p); GroupAlloc(&$dW,p);
               }
               A=$A.p_elements; B=$B.p_elements;
               W=$W.p_elements; dW=$dW.p_elements;

               /* compute direction vector */
               sq=(float)1.0/(GP_SIG*$q*$p);
               for(j=0;j<p;j++)
               {    sum=(float)0.0;
                    for(i=0;i<n;i++)
                         sum+=g[i]*GaussPulse((float)i,(float)j,
(float)(n-1),(float)(p-1),sq);
                    dW[j]=sum;
               }
               /* normalise direction vector */
               N=(float)0.0;
               for(j=0;j<p;j++) N+=dW[j];
               if(N!=(float)0.0) for(j=0;j<p;j++) dW[j]/=N;
               /* move state position */
               for(j=0;j<p;j++)
               {    W[j]=B[j]*W[j]+A[j]*dW[j];
     W[j]=generic_sigmoid(W[j],(float)1.0,(float)0.0,
altsg,altsc);
               }
     }    }
     /* output == new state position */
     GroupPacketAlloc(&$output,p);
     py=(group_ds *)PacketDataPt($output);
     for(j=0;j<p;j++) py->p_elements[j]=W[j];
     face_state(net,$,ND_ALT);
}
/* Generic Identity Production
*  Maps input to output
*/
Production_ds GenericIdentityProduction()
```

```
{   int i,n=0; group_ds *pg,*py;
 TBUGSET(GenericIdentityProduction)
    if($args.cmd!=NSLIO_NULL)
    {   if($args.cmd==NSLIO_CLOSE) com_free(&$input);
        $args.cmd=NSLIO_NULL;
        return;
    }
    if(pg=GroupPacketGet(PacketGet(net,&$input)))
    {   if((n=(int)pg->num_elements)>0)
            if(!$state.p_elements) GroupAlloc(&$state,n);
    }
    GroupPacketAlloc(&$output,n);
    py=(group_ds *)PacketDataPt($output);
    for(i=0;i<n;i++)
    {   py->p_elements[i]=pg->p_elements[i];
        $state.p_elements[i]=pg->p_elements[i];
    }
    face_state(net,$,ND_PRO);
}

/* Generic Linear Reduction
*  Take  linear combination of input groups
*/
Reduction_ds GenericLinearReduction()
{   int i,j,k=0,m=0,xx,nn=$size.input;
group_ds *pg,*py;
 TBUGSET(GenericLinearReduction)
    if($args.cmd!=NSLIO_NULL)
    {   if($args.cmd==NSLIO_CLOSE)
            for(xx=0;xx<nn;++xx) com_free(&$input[xx]);
        $args.cmd=NSLIO_NULL;
        return;
    }
    for(xx=0;xx<nn;++xx)
    {
    if(!(pg=GroupPacketGet(PacketGet(net,&$input[xx]))))
continue;
        j=(int)pg->num_elements;
        if(k==0)
        {   k=j;
if(k>0 && !$state.p_elements) GroupAlloc(&$state,k);
for(i=0;i<k;i++) $state.p_elements[i]=$state.default_value;
        }
        else if(k!=j) break;
        ++m;
        for(i=0;i<k;i++)
 $state.p_elements[i]+=pg->p_elements[i];
    }
    if(m>1) for(i=0;i<k;i++) $state.p_elements[i]/=(float)m;
    GroupPacketAlloc(&$output,k);
    py=(group_ds *)PacketDataPt($output);
    for(i=0;i<k;i++) py->p_elements[i]=$state.p_elements[i];
}
net ExpressionReduction()
{ feature()[NUM_FEAS];
 expression()[NUM_EXPS]
    [E_MELLOW]/input[L_FEAS]=feature[L_FEAS]
    {..\mem=(pulse [])[F_MELLOW]},
  [E_HAPPY]/input[L_FEAS]=feature[L_FEAS]
    {..\mem=(pulse [])[F_HAPPY]},
  [E_ANGRY]/input[L_FEAS]=feature[L_FEAS]
    {..\mem=(pulse [])[F_ANGRY]},
  [E_SAD]/input[L_FEAS]=feature[L_FEAS]
    {..\mem=(pulse [])[F_SAD]},
  [E_SQUINT]/input[L_FEAS]=feature[L_FEAS]
    {..\mem=(pulse [])[F_SQUINT]};
 ConxToGroup()
    output:,
  input[*]=expression[E_HAPPY,E_MELLOW,
E_MELLOW,E_ANGRY,E_MELLOW,
E_SAD,E_MELLOW,E_SQUINT]/output;
 ConxToGroup()
    mellow:,
    input[*]=expression[E_MELLOW]/output;
} =
{ feature[*]={(pulse [])[F_MELLOW]},
 feature[*]\fealen=(int [])[F_LENGTH],
 feature[*]\feapos.x=(int [])[F_POSITION_X],
 feature[*]\feapos.y=(int [])[F_POSITION_Y]
}

net DynSupportExpression( )
{   ExpressionReduction( )
        feature: feature,
        mellow: mellow;
    GenericGGAlteration( )
        input=ExpressionReduction/output;
    GenericIdentityProduction( )
        output: output,
        input=GenericGGAlteration/output;
} =
{ GenericGGAlteration\p=(float)4.0;
 /* laugh, frown, cry, wink groups */
 GenericGGAlteration\q=(float)1.0;
/* number of elements per group */
}
net DynSupportEmotion()
{   GenericLinearReduction()
        input: input;
    GenericGGAlteration()
        input=GenericLinearReduction/output;
    GenericIdentityProduction()
        output: output,
        input=GenericGGAlteration/output;
} =
{ GenericGGAlteration\p=(float)3.0;
 /* crying, winking, neither groups */
 GenericGGAlteration\q=(float)1.0;
/* number of elements per group */
 GenericGGAlteration\A=GROUP_INIT(DEF_RATE_E,0);
/* rate of change */
 GenericGGAlteration\B=GROUP_INIT(DEF_DECAY_E,0);
/* rate of decay */
}
net network( )
{   DynSupportExpression( )
        mellow: mellow;
    FilterElement(D_CRY)
        cry: output,
        input=DynSupportExpression/output;
    FilterElement(D_WINK)
        wink: output,
        input=DynSupportExpression/output;
    AddGroups( )
        input=cry+wink+mellow;
    DynSupportEmotion( )
        input=AddGroups;
    face_io_node( );
} =
{ face_io_node={{0,0,0,0,0,
    a_app(s_node(),..\DynSupportExpression\feature),
    a_app(s_node(),..\DynSupportExpression
\GenericIdentityProduction),
    a_app(s_node(),..\DynSupportEmotion
\GenericIdentityProduction)
    }}
}
```

Appendix.

References

[Aitk90] Aitkenhead, A.M. Slack, J.M. (1990). *Issues in cognitive modelling*. Lawerance Erlbaum.

[Alk87b] Aleksander, I. Kan, W. (1987). A probabilistic logic neuron network for associative learning. *IEEE International Conference on Neural Networks*. Vol.II, 541-548.

[Alek96] Aleksander, I. (1996). *Impossible minds*. Imperial College Press, London.

[Ali98] Ali, S.M. Zimmer, R.M. (1998). Beyond substance and process: a new framework for emergence. In: *Toward a science of consciousness II*, eds. S.R.Hameroff, A.W.Kaszniak, A.C.Scott. MIT Press.

[Arro92] Arrowsmith, D.K. Place, C.M. (1992). *Dynamical systems*. Chapman & Hall.

[Atk83] Atkins, P.W. (1983). *Molecular quantum mechanics*. Oxford University Press.

[Audi95] Audi, R. (1995). *The Cambridge dictionary of philosophy*. Cambridge University Press.

[Baar88] Baars, B.J. (1998). *A cognitive theory of consciousness*. Cambridge University Press.

[Baas94] Baas, N.A. (1994). Emergence, hierarchies, and hyperstructures. In: *Artificial Life*, ed. C.G.Langton, Vol.III, 515-537. Addison-Wesley.

[Baas96] Baas, N.A. (1996). A framework for higher order cognition and consciousness. In: *Toward a science of consciousness*, eds. S.R.Hameroff, A.W.Kaszniak, A.C.Scott. MIT Press. 633-648.

[Bae94] Baeten, J.C.M. Bergstra, J.A. Bol, R.N. (1994). A real time process logic. *Lecture Notes in AI, Temporal Logic*. No.827, 30-47.

[Bail98] Bailey, A. (1998). The five kinds of levels of description. In: *Toward a science of consciousness*, Vol.II, ed. S.R.Maeroff, A.W.Kaszniak, A.C.Scott. MIT Press.

[Brn95] Barenco, A. Bennett, C.H. Cleve, R. DiVincenzo, D.P. Margolus, N. Shor,P. Sleator, T. Smolin, J. Weinfurter, H. (1995). Elementary gates for quantum computation. *Physical Review A*. Pre-print.

[Bar87] Barlow, H. (1987). The biological role of consciousness. In: *Mindwaves*. Ed. by C.Blakemore and S.Greenfield. Blackwell.

[Bech91] Bechtel, W. Abrahamsen, A. (1991). *Connectionism and the mind: an introduction to parallel processing in networks*. Blackwell.

[Biro93] Biro, J.I. (1993). Consciousness and objectivity. In: *Consciousness*. Ed. by M.Davies, G.W.Humphreys. Blackwell.

[Blak87] Blakemore, C. Greenfield, S. (1987). *Mindwaves*. Blackwell.

[Blc78] Block, N. (1978). Troubles with Functionalism. In: *Perception and Cognition: Minnesota Studies in the Philosophy of Science, Vol.IX*, ed. by Savage.

[Blc95] Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and brain Science*. Vol.18, No.2, 227-287.

[Bode89] Boden, M.A. (1989). *Artificial intelligence in psychology*. Bradford Books, MIT Press.

[Bolc95] Bolc, L. Szalas, A. (1995). *Time and logic: a computational approach.* UCL Press.

[Bor96] Boreale, M. De Nicola, R. (1996). A symbolic semantics for the π-calculus. *Information and Computation.* No.126, 34-52.

[Boro89] Borowski, E.J. Borwein, J.M. (1989). *Dictionary of mathematics*. Collins.

[Bors70] Borst, C.V. (1970). *The mind / brain identity theory*. Macmillan.

[Bren74] Brentano, F. (1874/1973). *Psychology from and empirical standpoint.* Translated by A.Rancurello et al. Routledge & Paul.

[Bri93] Brisset, P. Ridoux, O. (1993). Continuations in λ Prolog, Logic Programming. In: *Proceedings Tenth International Conference On Logic Programming,* MIT Press, ed. by D.S.Warren. 29-43.

[Bro97] Brooks, R.A. (1997). Intelligence without representation. In: *Mind Design II*. Ed. by J.Haugeland. MIT Press.

[Cmpb94] Campbell, J. (1994). *Past, space and Self.* MIT Press.

[Cmp81] Campbell, K. (1981). The metaphysics of abstract particulars. In: *Midwest studies in philosophy VI: The foundations of analytic philosophy*, ed. by P.French, T.Uehling, H.Wettstein, University of Minnesota Press.

[Chf91] Chaffee, S.H. (1991). *Explication, Communications concepts, vol. 1.* Sage.

[Chan95] Chandler, D. (1995). *The act of writing: a media theory approach.* Aberystwyth, University of Wales.

[Chal96] Chalmers, D.J. (1996). *The conscious mind.* Oxford University Press.

[Chn87] Chen, S-S. (1987). Automated reasoning on neural networks: a probabilistic approach. *IEEE International Conference on Neural Networks.* Vol.II, 373-378.

[Chis57] Chisholm, R.M. (1957). Intentional inexistence. In: *Perceiving: a philosophical study.* Chapter 11, p168-85. Cornell University Press.

[Chom87] Chomsky, N. (1987). On the nature, use and acquisition of language. In: *Language in a psychological setting, Sophia Linguistica XII.* Sophia University Press.

[Chur83] Churchland, P.M. Churchland, P.S. (1983). Stalking the wild epistemic engine. *Nous*. Vol.17, 5-18.

[Clar89] Clark, A. (1989). *Mircocognition: philosophy, cognitive science and PDP*. Bradford Books, MIT Press.

[Clar93] Clark, A. (1993). *Sensory qualities*. Clarendon Press Oxford.

[Cohe86] Cohen, B. Hardwood, W.T. Jackson, M.I. (1986). *The specification of complex system*. Addison-Wesley.

[Cohn93] Cohn, P.M. (1993). *Algebra*. John Wiley and Sons, 2nd Ed. Vols.1, 2 and 3.

[Cos87] Cohen, M.S. Julian, W.H. (1987). Neural networks with a Hopf bifurcation: slowly modulated waves. *IEEE International Conference on Neural Networks*. Vol.II, 161-168.

[Coi96] Cointe, P. Dony, C. Malenfant, J. (1996). A semantics of introspection in a reflective prototype-based language. *LISP and Symbolic Computation*. No.9, 153-179.

[Cope93] Copeland, B.J. (1993). *Artificial intelligence, a philosophical introduction*. Blackwell.

[Cra98] Crane, T. Wiggins, D. (1998). Metaphysics. In: *Philosophy*. Vol.1. Ed. by A.C.Grayling. Oxford University Press.

[Crc90] Crick, F. Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neurosciences*. Vol.2, 263-275.

[Crc94] Crick, F. (1994). *The astonishing hypothesis*. New York: Scribner's.

[Crc95] Crick, F. Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*. Vol. 375, 121-123.

[Crc98] Crick, F. Koch, C. (1998). Consciousness and neuroscience. *Cerebral Cortex*. Vol.8, 97-107.

[Crc99] Crick, F. Koch, C. (1999). The unconscious homunculus. In: *The neuronal correlates of consciousness*, ed. by T.Metzinger. MIT Press.

[Darp88] DARPA, (1988). *DARPA neural network study*. AFCEA International Press.

[Dave93] Davies, M. Humphreys, G.W. (1993). *Consciousness*. Basil Blackwell.

[Dave98] Davies, M. (1998). The philosophy of mind. In: *Philosophy*, ed. by A.C.Grayling, Oxford University Press. Vol.1.

[Denn71] Dennett, D.C. (1971). Intentional systems. *Journal of Philosophy*. Vol.68, 87-106.

[Denn78] Dennett, D.C. (1978). *Brainstorms: philosophical essays on mind and psychology*. Bradford Books, Harvester Press.

[Denn88] Dennett, D.C. (1990). Quining Qualia. In: *Mind and cognition*, ed. by W.G.Lycan, Blackwell. 519-547.

[Denn91] Dennett, D.C. (1991). *Consciousness explained*. Penguin Press.

[Denn96]	Dennett, D.C. (1996). *Kinds of mind.* Phoenix.
[Desc85]	Descartes, R. (1985). Principles of Philosophy. In: *The philosophical writings of Descartes.* Vols. I & II. Translated by J.Cottingham et al., Cambridge University Press.
[Dev90]	Devitt, M. (1990). A narrow representational theory of mind. In: *Mind and cognition.* Ed. by W.G.Lycan. Blackwell.
[Dret80]	Dretske, F.I. (1980). The intentionality of cognitive states. *Midwest Studies in Philosophy.* University of Minnesota Press. Vol. V, 281-294.
[Dret86]	Dretske, F.I. (1986). Misrepresentation. In: *Belief: Form, Content and Function*, ed. by R.J.Bogdan, Oxford University Press.
[Dry88]	Dreyfus, G. Guyon, I. Nadal, J.P. Personnaz, L. (1988). Storage and retrieval of complex sequences in neural networks. *Physical Review A*, Vol.38, No.12, 6365-6372.
[Drey79]	Dreyfus, H.L. (1979). *What computers can't do.* Harber and Row.
[Drey97]	Dreyfus, H.L. (1997). *What computers still can't do.* MIT Press.
[Dun93]	Dunlop, C.E.M. Fetzer, J.H. (1993). *Glossary of cognitive science.* Paragon House.
[Ecka93]	von Eckardt, B. (1993). *What is cognitive science?* Bradford Books, MIT Press.
[Eco76]	Eco, U. (1976). *A theory of semiotics.* Indiana University Press.
[Edl85]	Edelman, G.M. Finkel, L.H. (1985). Interaction of synaptic modification rules within populations of neurons. *Proceedings National Academy of Science USA.* Vol.82, 1291-1295.
[Edel89]	Edelman, G.M. (1989). *The remembered present: a biological theory of consciousness.* Basic Books.
[Edel92]	Edelman, G.M. (1992). *Bright air, brilliant fire.* Penguin Press.
[Eke93]	Ekert, A. (1993). Quantum computation. *American Institute of Physics.* pre-print.
[Emm97]	Emmeche, C. Koppe, S. Stjernfelt, F. (1997). Explaining emergence: towards an ontology of levels. *Journal for General Philosophy of Science.* Vol.28, 83-119.
[Eng98]	Engel, A.K. Fries, P. Roelfsema, P.R. Konig, P. Singer W. (1998) Temporal binding, binocular rivalry, and consciousness. *Association for the scientific study of consciousness: electronic seminars.* http://server.phil.vt.edu/Assc.
[Fau89]	Faustini, A.A. Mitchell, W.H. (1989). The intensional logic language. In: *Tense, Proceedings of the 1989 International Symposium on Lucid and Intensional Programming.* Arizona State University.
[Fetz77]	Fetzer, J.H. (1977). A world of dispositions. *Synthese.* Vol.34, 397-421.
[Fetz81]	Fetzer, J.H. (1981). *Scientific knowledge.* Dordrecht, Holland: D.Reidel.
[Fetz86]	Fetzer, J.H. (1986). Methodological individualism. *Synthese.* Vol.68, 99-128.

[Fetz90] Fetzer, J.H. (1990). *Artificial intelligence: its scope and limits.* Kluwer Academic Publishers.

[Fetz91] Fetzer, J.H. (1991). Primitive concepts: habits, conventions, and laws. In: Definitions and Definability: Philosophical Perspectives, ed. by J.H.Fetzer, D.Shatz, G.Schlesinger. 51-68.

[Fetz93] Fetzer, J.H. (1993). *Philosophy of science.* Paragon House.

[Fetz96] Fetzer, J.H. (1996). *Philosophy and cognitive science.* Paragon House.

[Fetz98a] Fetzer, J.H. (1998). People are not computers: (most) thought processes are not computational procedures. *Journal of Experimental & Theoretical Artificial Intelligence.* Vol.10, 371-391.

[Fetz98b] Fetzer, J.H. (1998). Consciousness and cognition: semiotic conceptions. Personal communication.

[Fla92] Flanagan, O. (1992). *Consciousness reconsidered.* MIT Press.

[Flan89] Flanders, H. (1989). *Differential forms with applications to the physical sciences.* Dover.

[Flor86] Flores, F. Winograd T. (1986). *Understanding computers and cognition.* Addison Wesley.

[Fod75] Fodor, J.A. (1975). *The language of thought.* Crowell.

[Fod87] Fodor, J.A. (1987). *Psychosemantics.* MIT Press.

[Fod88] Fodor, J.A. Pylyshyn, Z.W. (1988). Connectionist and cognitive architecture: a critical analysis. In: *Connections and Symbols*, *Cognition Special Issue*, ed. by S.Pinker, J.Mehler. Vol.28.

[Fod90] Fodor, J.A. (1990). *A theory of content and other essays.* MIT Press.

[Foo95] Foo, N. (1995). Ontological revision, conceptual structures: applications, implementation and theory. In: *Lecture Notes in AI*, ed. by J.F.Sowa et al., Springer-Verlag. No.954, 16-31.

[Frnk95] Franklin, S. (1995). *Artificial minds.* MIT Press.

[Fri98] Frieden, B.R. (1998). *Physics from Fisher information.* Cambridge University Press.

[Fji87] Fujiwara, S. Okajima, K. Tanaka, S. (1987). A Heteroassociative memory network with feedback connection. *IEEE International Conference on Neural Networks*, Vol.II, 711-718.

[Gaa96] van der Gaag, L.C. (1996). Bayesian belief networks: odds and ends. *The Computer Journal.* Vol.39, No.2, 97-113.

[Gai78] Gaines, B.R. (1978). Fuzzy and probability uncertainty logic's. *Information and control.* Vol.38, 154-169.

[Gai87b] Gaines, B.R. (1987). Uncertainty as a foundation of computational power in neural networks. *IEEE International Conference on Neural Networks.* Vol.III, 51-57.

[Gasp91] Gasper, P. (1991). Causation and explanation. In: *The philosophy of science.* Ed. R.Boyd, P.Gasper, J.D.Trout. MIT Press.

[Geld95] van Gelder, T. Port, R.F. (1995). It's about time: an overview of the dynamical approach to cognition. In: *Mind as Motion*, ed. R.F.Port and T.van Gelder, MIT Press, 1-43.

[Gibs79] Gibson, J.J. (1979). *The ecological approach to visual perception*. LEA.

[Gil91] Gilman, R.H. (1991). Formal languages and infinite groups. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*.

[Goe88b] Goel, A. Chandrasekaran, B. Allemang, D. (1988). Connectionism and information. *AI Magazine*. Winter, 24-34.

[Gold98] Gold, I. (1998). Remarks. Comments on Engel et al. [Eng98].

[Gold96] Golden, R.M. (1996). Mathematical methods for neural network analysis and design. MIT Press.

[Gol93] Goldman, A.I. (1993). The psychology of folk psychology. Behavioural and Brain Sciences. Vol. 16, 15-28.

[Gdm92] Goodman, L.E. (1992). The trouble with phenomenalism. *American Philosophy Quarterly*. Vol.29, No.3, 237-252.

[Grah95] Graham, I. (1995). *Object oriented methods*. Addison-Wesley.

[Gre87] Green, P.E. Michalson, W.R. (1987). Real-time evidential reasoning and network based processing. *IEEE International Conference Neural Networks*. Vol.2, 359-365.

[Greg89] Gregory, R.L. (1989). *The Oxford companion to the mind*. Oxford University Press.

[Gss68] Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Science USA*. Vol.60, 758-765.

[Gss75b] Grossberg, S. Ellias, S.A. (1975). Pattern formation, contrast control, and oscillations in short term memory of shunting on-center off-surround networks. *Biological Cybernetics*. Vol.20, 69-98.

[Gss88b] Grossberg, S. (1988). *Neural networks and natural intelligence*. Bradford Books, MIT Press.

[Gss95] Grossberg, S. (1995). Neural dynamics of motion perception, recognition learning, and spatial attention. In: *Mind as Motion*, ed. R.F.Port and T.van Gelder, MIT Press, 449-489.

[Grs93] Grossman, R.L. Larson, R.G. (1993). An algebraic approach to hybrid systems. *http://www.math.uic.edu*.

[Guz97] Guzeldere, G. (1997). The many faces of consciousness: a field guide. In: *The Nature of Consciousness*, ed. by Block et al., MIT Press. 1-67.

[Gul80] van Gulick, R. (1980). Functionalism, information and content. *Nature and System*. Vol. 2, 139-162.

[Gul93] van Gulick, R. (1993). Understanding the phenomenal mind: are we all just armadillos? In: *Consciousness*. Ed. by M.Davies, G.W.Humphreys. Blakwell.

[Hack87] Hacker, P.M.S. (1987). *Appearance and reality*. Blackwell.

[Hart58] Hartshorne, C. Weiss, P. Burks, A. (1958). Collected papers of Charles Sanders Peirce. Eds. Harvard University Press.

[Haug85] Haugeland, J. (1985). *AI: the very idea.* MIT Press.

[Haug97] Haugeland, J. (1997). *Mind design II.* MIT Press.

[Haus93] Hausman, C.R. (1993). *Charles S. Peirce's evolutionary philosophy.* Cambridge University Press.

[Haye96] Hayes, M.H. (1996). *Statistical digital signal processing and modeling.* Wiley.

[Heil98] Heil, J. (1998). *Philosophy of mind.* Routledge.

[Hmp52] Hempel, C. (1952). Fundamentals of concept formation in empirical science. University of Chicago Press.

[Hmp66] Hempel, C. (1966). *Philosophy of Natural Science.* Prentice-Hall.

[Hill91] Hill, C.S. (1991). Sensations: a defence of type materialism. *Cambridge University Press.*

[Hil94] Hill, P.M. (1994). A module system for meta-programming. In: *Logic Program Synthesis and Transformation*, LNCS, ed. by Fribourg and Turini, Springer-Verlag. No.883, 395-409.

[Hodg93] Hodges, W. (1993). *Model theory.* Cambridge University Press.

[Hofs83] Hofstadter, D.R. (1983). Escher, Godel and Bach.

[Hond90] Honderich, T. (1990). *Mind and brain: a theory of determinism.* Clarendon Press.

[Hop82] Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings National Academy Science USA, Biophysics.* Vol.79, 2554-2558.

[Hop84] Hopfield, J.J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings National Academy Science USA, Biophysics.* Vol.81, 3088-3092.

[Hum87] Humphrey, N. (1987). The inner eye of consciousness. In: *Mindwaves.* C. Blakemore and S.Greenfield. Blackwell.

[Iwa94] Iwasaki, Y. Simon, H.A. (1994). Causality and model abstraction. *AI.* Vol.67, 143-194.

[Jack87] Jackendoff, R.S. (1987). *Consciousness and the computational mind.* Bradford Books, MIT Press.

[Jack82] Jackson, F. (1982). Epiphenomenal qualia. *American Philosophical Quarterly.* 127-36.

[John83] Johnson-Laird, P.N. (1983). *Mental models.* Cambridge University Press.

[Jibu95] Jibu, M. Yasue, K. (1995). *Quantum brain dynamics and consciousness.* J.Benjamins.

[Kim93] Kim, J. (1993). *Supervenience and mind.* Cambridge University Press.

[Klee84] de Kleer, J. Brown, J.S. (1984). A qualitative physics based on confluences. *Artificial Intelligence.* Vol.24, 7-83.

[Klir88] Klir, G.J. Folger, T.A. (1988). *Fuzzy sets, uncertainty, and information.* Prentice Hall.

[Ksk87] Kosko, B. (1987). Adaptive bi-directional associative memories. *Applied Optics.* Vol.26, No.23, 4947-4960.

[Kosk92] Kosko, B. (1992). *Neural networks and fuzzy systems.* Prentice-Hall.

[Kyb91] Kyburg, H. (1991). Normative and descriptive Ideals. In: *Philosophy and AI*, ed. Cummins and Pollock, MIT Press, 129-140.

[Kup86] Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence.* Vol.29, 289-338.

[Lepo91] Lepore, E. Van Gulick, R. (1991). *John Searle and his critics.* Basil Blackwell.

[Lev93] Levine, J. (1993). On leaving out what it's like. In: *Consciousness.* Ed. by M.Davies and G.W.Humphreys. Basil Blackwell.

[Lew94] Lewis, J. Kieburtz, R.B. (1994). *Algebraic Design Language.* Pacific Software Research Center, Oregon Graduate Institute of Science and Technology, Portland.

[Lin95] Lind, D. Marcus, B. (1995). *An Introduction to Symbolic Dynamics and Coding.* Cambridge University Press.

[Lock75] Locke, J. (1975). *An essay concerning human understanding.* Ed. by P.H.Nidditch. Oxford University Press.

[Lop94] Lopes, J.G. Wermelinger, M. (1994). Basic conceptual structures theory. *Lecture Notes in AI.* Springer-Verlag. No.835, 144-159.

[Lyca90] Lycan, W.G. (1990). *Mind and cognition.* Blackwell.

[Lyca96] Lycan, W.G. (1996). *Consciousness and experience.* Bradford Books, MIT Press.

[Lyn95] Lyons, W. (1995). *Approaches to intentionality.* Oxford University Press.

[Mach96] Machover, M. (1996). *Set theory, logic and their limitations.* Cambridge University Press.

[Mali98] Malim, T. Birch, A. (1998). *Introduction to psychology.* Macmillan Press.

[Marr82] Marr, D.C. (1982). *Vision: a computation investigation into the human representation and processing of visual information.* San Francisco: Freeman.

[Mass98] Massey, B. Ward-Smith, J. (1998). *Mechanics of fluids.* Stanley Thornes.

[Mau89] Maudlin, T. (1989). Computation and consciousness. *Journal of Philosophy.* Vol.86, 407-432.

[McGi91] McGinn, C. (1991). *The problem of consciousness.* Blackwell.

[Mead88] Mead, C. (1988). *Analog VLSI and neural systems.* Addison-Wesley.

[Mll93] Mellor, D.H. (1993). Supervenience? No chance! Reply to Menuge. *Analysis.* Vol.53, No.4, 236-239.

[Metz95] Metzinger, T. (1995). Faster than thought. In: *Conscious experience*, ed. by T.Metzinger. Imprint Academic.

[Mil92] Milner, R. Parrow, J. Walker, D. (1992). A calculus of mobile processes I. *Information and computation*. Vol. 100, 1-40.

[Min94] Mineau, G.W. (1994). Views, mappings and functions: essential definitions to the conceptual graph theory. *Lecture Notes in AI*. Springer-Verlag. No.835, 160-174.

[Mins68] Minsky, M.L. (1968). *Semantic information processing*. (Editor). MIT Press.

[Mins69] Minsky, M.L. Papert, S. (1969). *Perceptrons*. MIT Press, new ed. 1988.

[Mins87] Minsky, M.L. (1987). *The society of mind*. Heinemann.

[Mood74] Mood, A.M. Graybill, F.A. Boes, D.C. (1974). *Introduction to the theory of statistics*. McGraw-Hill, 18th Print 1987.

[Moo65] Moore. G. (1965). *Moore's Law*. See www.intel.com/intel/museum.

[Mor38] Morris, C.W. (1938). *Foundations of the theory of signs*. Chicago University Press.

[Nag71] Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese*. Vol. XXII, 396-413.

[Nag74] Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*. Vol.83, 435-450.

[Naka90] Nakahara, M. (1990). *Geometry, topology and physics*. IOP Publishing.

[Nan95] Nanopoulos, D.V. (1995). Theory of brain function, quantum mechanics and superstrings. *CERN-TH/95-128*. nanopoud@cernvm.cern.ch.

[Nel93] Nelkin, N. (1993). The connection between intentionality and consciousness. In: *Consciousness*, eds. M.Davies and G.W.Humphreys, Blackwell.

[Neu56] von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. *Collected works*. Ed. A.H.Taub, Pergamon Press, Vol.5, 372.

[Nwl61] Newell, A. Simon, H.A. (1961). GPS, a program that simulates human thought. In: *Computers and thought*. Ed. E.A.Feigenbaum, J.Feldman, McGraw-Hill, 279-293.

[Nwl76] Newell, A. Simon, H.A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, Vol.3, 205-211.

[Nwl82] Newell, A. (1982). The knowledge level. *Artificial Intelligence*. Vol.18, 87-127.

[Nwl90] Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

[Ngy96] Nguyen, H.T. Walker, E.A. (1996). *A first course in fuzzy logic*. CRC Press.

[Nort95]	Norton, A. (1995). Dynamics: an introduction. In: *Mind as Motion*, ed. R.F.Port and T.van Gelder, MIT Press, 45-68.
[Org94]	Orgun, M.A. Ma, W. (1994). An overview of temporal and modal logic programming. *Lecture Notes in AI, Temporal Logic*. Springer-Verlag. No.827, 445-479.
[Pain93]	Pain, H.J. (1993). *The physics of vibrations and waves*. Wiley. 4th ed.
[Park88]	Parkinson, G.H.R. (1988). *Leibniz*. Everyman Classics.
[Pat96]	Pattee, H.H. (1996). Evolving self-reference: matter, symbols, and semantic closure. *CC-AI*. Vol.12, Nos.1-2, 9-27.
[Paw94]	Pawlak, Z. (1994). Hard and soft sets. *Rough sets, fuzzy sets and knowledge discovery*, ed. by W.P.Ziarko, Springer-Verlag. 130-135.
[Pear88]	Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann.
[Penr89]	Penrose, R. (1989). *The Emperor's new mind*. Oxford University Press.
[Penr94]	Penrose, R. (1994). *Shadows of the mind*. Vintage Science.
[Pen94]	Penther, B. (1994). A dynamic logic of action. *Journal of Logic, Language, and Information*. Vol.3, 169-210.
[Peyt87]	Peyton Jones, S.L. (1987). *The implementation of functional programming languages*. Prentice-Hall.
[Pola68]	Polanyi, M. (1968). Life's irreducible structure. *Science*. Vol. 160, 1308-1313.
[Pre95]	Prehofer, C. (1995). A call-by-need strategy for higher-order functional-logic programming. In: *Logic Programming, Proceedings International Symposium,* ed. by J.Lloyd, MIT Press. 147-161.
[Rebe86]	Reber, A.S. (1986). *Dictionary of psychology*. Penguin books.
[Resc69]	Rescher, N. (1969). *Many-valued logic*. McGraw-Hill.
[Res92]	Restall, G. (1992). Arithmetic and truth in Lukasiewicz's infinitely valued logic. *Logique and Analyse*. No.139-140, 303-312.
[Rey97]	Rey, G. (1997). *Contemporary philosophy of mind*. Blackwell.
[Rob95]	Robinson, D.J.S. (1995). *A course in the theory of groups*. Springer.
[Rse62]	Rosenblatt, F. (1962). *Principles of neurodynamics*. Spartan.
[Rsnf92]	Rosenfield, I. (1992). The strange, familiar and forgotten. Picador.
[Rsnt91]	Rosenthal, D.M. (1991). *The nature of mind*. Oxford University Press.
[Russ95]	Russell, S. Norvig, P. (1995). *Artificial intelligence: a modern approach*. Prentice-Hall.
[Ryc95]	Rychlak, J.F. (1995). A teleological critique of modern cognitivism. *Theory and Psychology, SAGE*. Vol.5, No.4, 511-531.
[Ryl49]	Ryle, G. (1949). *The concept of mind*. Penguin.
[San98]	Sainsbury, M. (1998). Philosophical logic. In: *Philosophy*, ed. by A.C.Grayling. Vol.1, 61-122.

[San91] Sanford, A.J. (1991). *Cognition and cognitive psychology.* Lawerance Erlbaum Associates.

[Svg98] Savage, J.E. (1998). *Models of computation.* Addison-Wesley.

[Schi90] Schiffman, H.R. (1990). *Sensation and perception.* Wiley.

[Schm93] Schmidt, G. Ströhlein, T. (1993). *Relations and graphs.* Springer-Verlag.

[Scot77] Scott, A.C. (1977). *Neurophysics.* Wiley Interscience.

[Seag91] Seager, W. (1991). *Metaphysics of consciousness.* Routledge.

[Sear83] Searle, J.R. (1983). *Intentionality.* Cambridge University Press.

[Sear84] Searle, J.R. (1984). *Minds, brains and science: the 1984 Reith lectures.* BBC publication.

[Sear92] Searle, J.R. (1992). *The rediscovery of the mind.* MIT Press.

[Sear97] Searle, J.R. (1997). *The mystery of consciousness.* Granta.

[Sear98] Searle, J.R. (1998). How to study consciousness scientifically. In: *Toward a science of consciousness.* Ed. S.R.Hameroff, A.W.Kaszniak, A.C.Scott. MIT Press. Vol.2.

[Sej86c] Sejnowski, T.J. Rosenberg, C.R. (1988). *NETtalk: a parallel network that learns to read aloud.* EECS Technical Report, John Hopkins University.

[Shal98] Shallice, T. (1998). Modularity and consciousness. In: *The nature of consciousness.* Ed. by N.Block, O.Flanagan, G.Guzeldere. MIT Press.

[Shp96] Shapiro, Y. (1996). The perceiving brain: Turing's machine or Darwin's machine? *Theory and Psychology, SAGE.* Vol.6, No.2, 195-228.

[Shea97] Shear, J. (1997). *Explaining consciousness – the "hard problem".* Ed. J.Shear. MIT Press.

[Shep83] Shepherd, G.M. (1983). *The synaptic organization of the brain.* Oxford University Press.

[Shep83b] Shepherd, G.M. (1983). *Neurobiology.* Oxford University Press.

[Sim96] Simon, H.A. (1996). *The sciences of the artificial.* MIT Press.

[Skog96] Skogestad, S. Postlethwaite, I. (1996). *Multivariable feedback control: analysis and design.* Wiley.

[Sla90] Slaney, J. (1990). A general logic. *Australasian Journal of Philosophy.* Vol.68, No.1, 74-88.

[Slo97] Sloman, A. (1997). *The evolution of what?* http://www.cs.bham.ac.uk/~axs.

[Smrt62] Smart, J.J.C. (1962). Sensation and brain processes. *The philosophy of mind*, ed. by V.C.Chappell, New York: Dover, 1981, 16-172.

[Smit86] Smith, P. Jones, O.R. (1986). *The philosophy of mind.* Cambridge University Press.

[Smit96] Smith, B.C. (1996). *On the origin of objects.* MIT Press.

[Smul94] Smullyan, R.M. (1994). *Diagonalization and self-reference.* Oxford Logic Guides, Clarendon Press. No.27.

[Sow95] Sowa, J.F. (1995). Syntax, semantics, and pragmatics of contexts. In: *Conceptual structures: applications, implementation and theory, Lecture Notes in AI*, ed. by J.F.Sowa et al., Springer-Verlag. No. 954, 1-15.

[Spe90] Specht, D.F. (1990). Probabilistic neural networks. *Neural Networks.* Vol.3, 109-118.

[Stap93] Stapp, H.P. (1993). *Mind, matter, and quantum mechanics.* Springer Verlag.

[Ster90] Sterelny, K. (1990). *The representational theory of mind.* Blackwell.

[Sti83] Stich, S. (1983). The syntactic theory of mind. In*: From folk psychology to cognitive science*, MIT Press, chapter 8.

[Sti92] Stich, S. (1992). What is a theory of mental representation. *Mind.* Vol.101, No.402, 242-261.

[Stil95] Stillings, N.A. Weisler, S.E. Chase, C.H. Feinstein, M.H. Garfield, J.L. Rissland, E.L. (1995). *Cognitive science, an introduction.* MIT Press.

[Swok79] Swokowski, E.W. (1979). *Calculus with analytic geometry.* Prindle, Weber & Schmidt.

[Tar69] Tarski, A. (1969). Truth and proof. *Scientific American.* Vol.194, 63-77.

[Thi91] Thijsse, E.G.C. (1991). Logic's of consciousness explained and compared. *Logique and Analyse.* Vol.133-134, 221-250.

[Tur50] Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, Vol. 59, 422-460.

[Vlm98] Velmans, M. (1998). Goodbye to reductionism. In: *Towards a science of consciousness*, ed. by S.Hameroff, A.Kaszniac, A.Scott, MIT Press. 45-52.

[Vil94] Vila, L. (1994). A survey of temporal reasoning in artificial intelligence. *AICOM.* Vol.7, No.1, 4-27.

[West87] Westphal, J. (1987). *Colour: a philosophical introduction.* Blackwell.

[Who89] Whobrey, D.J.R. (1989). A communications chip for multiprocessors. *CONPAR88.* Cambridge University Press.

[Who89b] Whobrey, D.J.R. (1989). *NeSeL: users guide & reference manual.* Department of Computer Science, City University, London.

[Win92] Winston, P.H. (1992). *Artificial Intelligence.* Addison-Wesley.

[Wood87] Wood, D. (1987). *Theory of computation.* Harper and Row.

[Zal93] Zalta, E.N. (1993). Twenty-five basic theorems in situation and world theory. *Journal of Philosophical Logic.* Vol.22, 385-428.

[Zal97] Zalta, E.N. (1997). *Principia Metaphysica.* Draft, Stanford University, http://mally.standord.edu.

[Zeki93] Zeki, S. (1993). *A vision of the brain.* Blackwell.