Machine mentality and the nature of the ground relation.
Minds and Machines, Vol. 11, Issue 3, 2001. © Kluwer Academic Publishers.

Machine Mentality and the Nature of the Ground Relation

Darren Whobrey

Department of Computer Science

City University, London

Abstract

John Searle distinguished between *weak* and *strong* artificial intelligence (AI). This essay discusses a third alternative, *mild* AI, according to which a machine may be capable of possessing a species of mentality. Using James Fetzer's conception of minds as semiotic systems, the possibility of what might be called "mild AI" receives consideration. Fetzer argues against strong AI by contending that digital machines lack the ground relationship required of semiotic systems. In this essay, the implementational nature of semiotic processes posited by Charles S. Peirce's triadic sign relation is re-examined in terms of the underlying dispositional processes and the ontological levels they would span in an inanimate machine. This suggests that, if non-human mentality can be replicated rather than merely simulated in a digital machine, the direction to pursue appears to be that of mild AI.

Keywords: artificial intelligence, causal systems, dispositional processes, emergence, ground relation, machine mentality, mind, numeric systems, ontological levels, replication, semiotic processes, simulation, symbolic systems.

1 Introduction

John Searle (1984) has characterised the domain of artificial intelligence (AI) 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 essay discusses aspects of 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 approach adopted will be to explore whether mind can be replicated, as opposed to merely simulated, in digital machines. James Fetzer (1990, p. 17) 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." Fetzer suggests 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, suggesting digital machines lack the right kind of semiotic ground relations when viewed from a computational perspective. 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.

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 essay is to determine whether semiotic processes, which form the basis of a semiotic definition of mind, 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 it 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 previous results 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 first half of this essay explores Fetzer's analysis of the analogical argument for strong AI in some detail by 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 main classes of system used in theories of mind, 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 discussed there 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 Peirce's Semiotic as a Basis for a Theory of Mind

To start with, a brief review is presented of the terminology used by Peirce in his semiotic (theory of signs) – the later sections discuss his theory in more detail. Peirce set about developing what he called an architectonic for philosophy, a comprehensive view that encompassed pragmatism, semiotic, phenomenology and metaphysics based on synechism (his theory of continuity) – see Charles Hartshorne (1958). For Peirce, the semiotic was just one part of his philosophical architectonic in which the nature and role of mind was central. Charles Morris (1938) distinguished three branches in the semiotic field: 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. Herein, Peirce's ideas are treated in a broader sense as a basis for exploring the overall nomic structure of mind. While the parts of his architectonic are very much interrelated, for present purposes only aspects of the semiotic are drawn upon.

In relation to the mind, an important concept that arises from Peirce's theory is the process of signification. Central to this is the suggestion that some object has significance to the mind only as a result of the causal consequences invoked by the interpretation of a representation of the object. Following from this, and in simple terms, a sign is defined as the product of a specific mental process of signification involving a particular type of object, its representation and form of interpretation. A key part of the operation of mind would then involve the interaction of a multitude of these signs during the signification of the reality perceived by the creature embodying the mind. One purpose of the following sections is to refine the nature of this signification process.

Turning now to an overview of the main terms used by Peirce, starting with what he called the *dynamical object*. This is any actual object as it exists independently of any interpretation by a mind – however, it could be a mental, abstract object. A *ground relation* is said to transform an aspect of the dynamical object (i.e. that feature of the object being signified) into a representation called the *immediate object* (the nature of a ground relation and the representation is discussed later). The representation causes an effect, called the *interpretant*, in the subject acting as interpreter. The nature of an interpretant and how a representation causes an effect is discussed later (an interpretant is treated as a kind of dispositional process – discussed later). 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. The following sections discuss in more detail the nature of these terms, relations and the processes involved.

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. However, quoting Carl Hausman (1993, p. 72), “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.” This seems to suggest that there are four components to the triadic relation, not three, but, as Hausman suggests and the later sections will explain, under this more detailed analysis the ground relation should be understood as part of the interpretant. Paying attention to details such as this will play an important part in determining whether machines can have minds under the semiotic conception of mind.

Finally, it is useful to summarise, for the purposes of what follows, how 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 (1993, p. 86).

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.

3 How Mild AI Addresses the Problems Faced by Strong AI

The purpose of the first part of this section is to set out the analogical argument for strong AI as framed by Fetzer, and then to summarise the main points of Fetzer's counter argument. This ends with a review of William Rapaport's counter argument to Fetzer. The second part presents the main premise of mild AI: that the ontological levels within the system have to be considered – the meaning of *ontologic level* is discussed later. This suggests an alternate way of relating mind to digital machines.

3.1 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 (1990, p. 41-58). 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. Wherefore, the last two have the ability to reason logically and for criticism, respectively. A criterion for a system to be a mind arises from its capacity to make a mistake since it may take "something to stand for something other than that for which it stands" (p. 40). This also implies the system is conscious according to Fetzer's definition (p. 81): "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" (p. 16). 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 (p. 277), illustrated here in **Table 1**. This marks out four premises for each of which a parallel is suggested between human beings and digital machines. The strong AI argument runs: if stimuli and responses parallel inputs and outputs, and mental processes parallel (certain) programs, then digital machines can have minds. The following sections analyse this argument in more detail.

	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 (1990, p. 277).

Summarising Fetzer's counter argument, Premise 3 is singled out as the point at which the analogical argument for strong AI breaks down. This is suggested to happen 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 (illustrated in Figure 1).

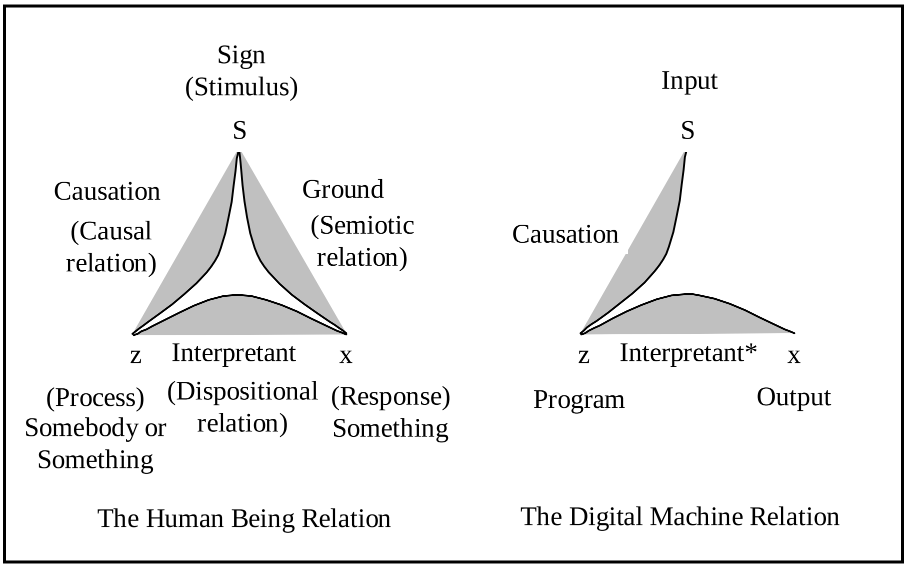


Figure 1. Comparing the operation of human beings & digital machines in terms of Peirce's Sign triad. After Fetzer (1990, p. 277-278).

The meanings of the terms and relations in Figure 1, and the parallel between the Human Being Relation and the Digital Machine Relation, are discussed at length in the remainder of this essay. At this point the figure is introduced to highlight what Fetzer suggests is the main problem with the analogical argument for strong AI, namely 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 (1990, p. 278), “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.” These points are taken up and analysed more closely in what follows.

Fetzer (1991) 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. To sum up, Fetzer (1990, p. 88) 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.

To counter Fetzer’s claims, Rapaport (1998, p. 405) suggests: “What is important for computationalism properly treated is whether cognitive processes are algorithmic (i.e. computable) in the Turing-machine sense.” Rapaport then goes on to consider whether there are any cognitive processes that cannot be computed, suggesting there are not and, therefore, that computationalism is true. However, this assumes that what is important about cognitive processes is the result of the process, the mind being characterised by a sequence of discrete states, and that an equivalence can be made between this and the results of a string of computations. In contrast, the semiotic conception of mind suggests that, operationally, what is important is the structure of the occurrent causal relations between processes and the nomic structure, on top of this causal structure, required to produce the web of semiotic processes that give rise to the mind.

Hence, the problem with Rapaport's counter argument is that an attempt is made to equate the algorithmic structure of the system (such as the causal relation between the nodes of the semantic network that make up the "mind" of Cassie, a computational cognitive agent that can interpret narratives) with the representation of the relational structure of the contents of the world perceived by the mind to the mind that these relations make up, which includes the mind itself! For example, Fetzer is suggesting that the mind of Cassie needs to represent and interpret the relations between the nodes in the semantic network. So, when Rapaport (p. 411) talks about the nodes of Cassie's semantic network being "grounded" by their location in the semantic-network, this is equating logical truth grounding (i.e. denotational semantics) with the occurrent representation and interpretation of ground relations. That is, a semiotic ground is a process asserting a (dispositional) relational effect with a certain form on recipient processes, whereas logical truth grounding is concerned with the logical consistency of the states of the system. Consequently, in the case of Cassie, her mind should be identified with the process the semantic network supports, and it is with respect to this process that ground relations must be signified. This suggests that understanding how ground relations are represented, from what perspective, to whom, and what role they play in the semiotic sense is a precursor to determining whether mild AI is possible.

3.2 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 (1993, p. 126). However, Searle was quick to refute this counter argument. To use the "skin-of-an-onion" analogy given by Turing (1950), 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 – this is discussed shortly. In this regard, for the purposes of analysis the essay 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 (1993). In addition, it does not rely upon or imply non-deducible emergent properties – see Claus Emmeche et al. (1997). 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 Semiotic Processes, Dispositions & System Laws

The previous section presented ontological levels as the main instrument used here to explore the promise of mild AI. Immediately this suggests an analysis may be fruitful of the nomic relations between the ontological levels within the system. A precursor step involves determining how to accommodate the semiotic processes implied by the semiotic conception of mind adopted here. This section explains how semiotic processes are most naturally understood in terms of dispositional processes and that within this conception certain laws can be set out stemming from the structure implied by the dispositional support of semiotic processes.

4.1 The Importance of Dispositions to Semiotic Processes

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 (1990, p. 78) and (1991). 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 (1977, p. 401).

For present purposes, the following four dispositional conceptions introduced by Fetzer (1981, p. 40) 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 (1981, p. 38). 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, an assumption of this essay is to consider semiotic processes as being 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.

4.2 Theories of Content & Nomic System Structure

This subsection introduces Fetzer's laws for the nomic structure of semiotic systems. These are extended in a later section to accommodate the ontologic levels within the system, which will serve to help clarify the relationship between computational processes and semiotic processes.

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 (1997) and Barbara von Eckardt (1993). Fetzer (1990, p. 86) 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 – dispositions are discussed later. 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.

Fetzer (1990, p. 288) 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 ontologic level structure of semiotic systems. To begin with a couple of Fetzer’s laws are introduced as they relate specifically to dispositions. A later section will show how the ontologic level structuring effect can be incorporated into them. 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 (1990, p. 288) 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 – the justification of this point is returned to later. 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. The following sections examine 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 Fetzer (1996, p. 109):

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

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

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.” One task is to refine the nature of the system structure that gives rise to these laws – this is explored elsewhere – see Whobrey (1999). In particular, examining graded dispositions 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.

5 Ontological System Levels

To explore how semiotic processes might be replicated in a machine the essay 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.

5.1 Causal Relevance Model of Explanation

This discussion 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 (1977, p. 407) 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 (1981, p. 110) 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 (1981, p. 49).

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 (1993, p. 22). 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 (1993, p. 77). 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.

5.2 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 (1982) 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 elsewhere – see Whobrey (1999).

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 $S_i = \langle E_i, F_i, R_i \rangle_{Ci}$, 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 ≥ 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,* in the sense 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. 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.

5.3 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. 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 (1994). 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.

5.4 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 (1986, p. 124) 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 view as detailed by Emmeche et al. (1997). 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 (1994) 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 (1996). In this regard, Whobrey (1999) presents a formal characterisation of the ontological levels, laws and properties of a causal system.

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 (described later), 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 – although their notation can be used as a basis for formally specifying emergent properties. Thinking about it in terms of emergent properties, in this particular case, a crucial point made in Whobrey (1999) 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 as well. 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 more involved, in that to get emergent mental properties a "causal accessibility requirement" 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 (1993, p. 63). 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 (1981, p. 144). Hence, 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. Consequently, 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.

5.5 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. The hypothetical system is divided into three ontological levels for the purpose of analysis: an implementation level, a dispositional level, and a phenomenal level. Adopting the notation used in Whobrey (1999):

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 essay develops it will be shown that they can be treated as distinct system levels in the ontological sense developed above. 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 a causal accessibility requirement gives rise to the phenomenon of concepts at the phenomenal level. The next section discusses the nature of these dispositional processes in more detail.

6 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. Fetzer (1977)) and the following subsections.

6.1 Functional-Flows & Temporal-Representations

As a foundation for the following discussion on “graded” dispositions, some terminology is helpful from dynamic systems theory – see Norton (1995). 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.

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. A definition of causal structure can be cast in terms of the relations between property instantiations – see Whobrey (1999).

6.2 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. Fetzer 1981, p. 51 – 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. 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. Fetzer 1977, p. 415), 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 Fetzer (1996, p. 109). 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 of graded 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” (Fetzer 1981, p. 113) 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).

6.3 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 (1982). These nets are capable of memorising patterns that can be later re-invoked when prompted by a similar pattern. Kosko (1987) 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. (1988) and Fujiwara et al. (1987). 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 its 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 (1982). 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.

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 δs' 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 (1977, p. 403), draws an important distinction between "*predicate constants*…, such as 'H', and the *sentential functions* that may be constructed from them, such as ⌈Hx⌉." 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, an instance of a property was effectively a function of a particular configuration of elements – defined formally in Whobrey (1999), see definition D-SLP: $p^{n}_{ijx} = \omega^{n}_{ij}\,\varsigma^{n}_{ij}(C^{n}_{ijx}, E^{n}_{i})$. 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.

6.4 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* instances of dispositions (DIDs), 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. Therefore, 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.

7 Grounding Semiotic Processes

In terms of the operational and structural properties of semiotic processes, the meaning of a sign (its semantics) could be 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.

7.1 The Ground Relation

As mentioned in the introduction, 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. 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 (1993, p. 89).

Often the ground is said to express the meaning of the sign and is equated with its content. However, Daniel Chandler (1995) 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.” Theories of content attempt to provide explanations for the production of meaning in systems that do not succumb to the interpreter-regress problem (see Eckardt (1993)) and satisfactorily elucidate any semantic primitives – see Fodor (1990) and Umberto Eco (1976). In contrast, this essay 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.

7.2 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. 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 (1990, p. 40).

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 neutral 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.

7.3 Peirce’s Triadic Sign Relation

The introduction reviewed Peirce’s terminology for the aspects of semiotic processes and included Fetzer’s version of the sign triad. Figure 2 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.

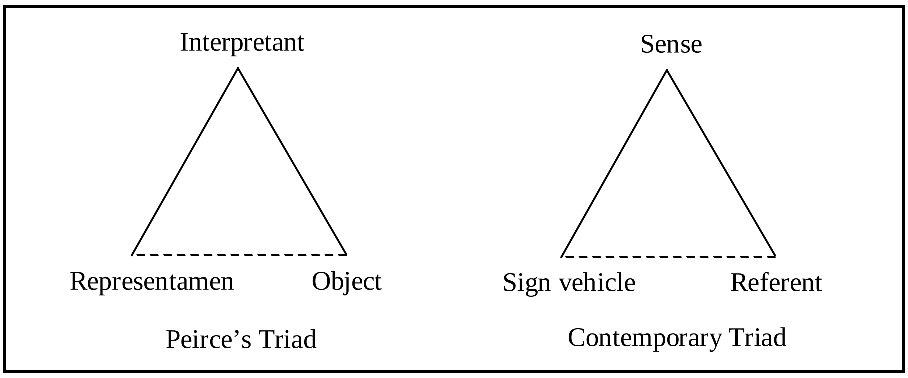


Figure 2. Simplified diagrams of semiotic relationships.
The first shows the key aspects of Peirce’s conception of the triadic sign relation – see Eco (1976, p. 59). The second shows a contemporary version using “more familiar” terms – see Chandler (1995).

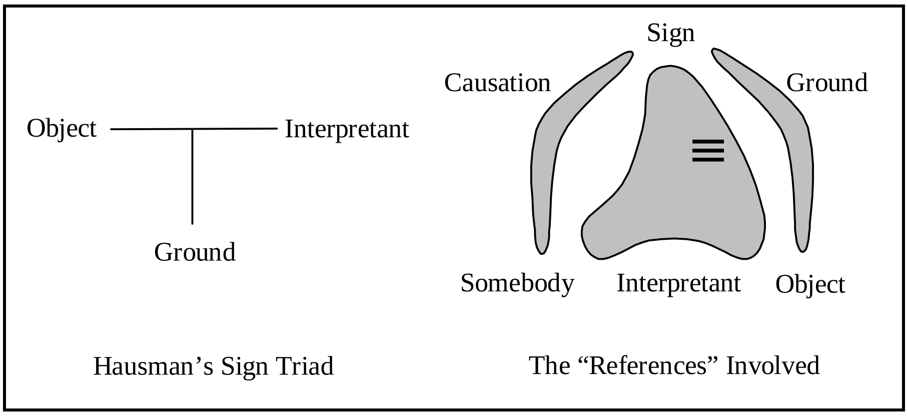
The diagrams are a simplification of the sign triad and also lean toward a subjective characterisation stemming from the way signification appears from a first person perspective. For the present purposes, Carl Hausman (1993, p. 71) presents a more useful characterisation, by reading Pierce more closely, 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 Figure 3 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.



Hausman’s Sign Triad

The “References” Involved

Figure 3. Triadic sign relation showing ground relation.
The first shows the ground version of the triadic sign relation according to Hausman – see Hausman (1993, p. 72). The second shows the references involved according to Peirce – see Peirce 5.283.

To help understand the form of semiotic processes the following subsections present augmented diagrams of the sign triad that include annotations for the types of the relations and the ontological level of the various entities involved.

7.4 Fetzer's Interpretation of the Triadic Sign Relation

This subsection analyses Fetzer's Human Being and Digital Machine Relations from his analysis of the analogical argument for strong AI introduced above. 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.

The introduction presented Fetzer's diagram for the Human Being and Digital Machine Relations. Noticeable in comparing the two relations is the absence of the ground relation in the case of digital machines. To recall, Fetzer (1991, p. 278) 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. 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. 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 subsection, the relationships within Fetzer's Human Being Relation are elaborated in terms of Peirce's "References".

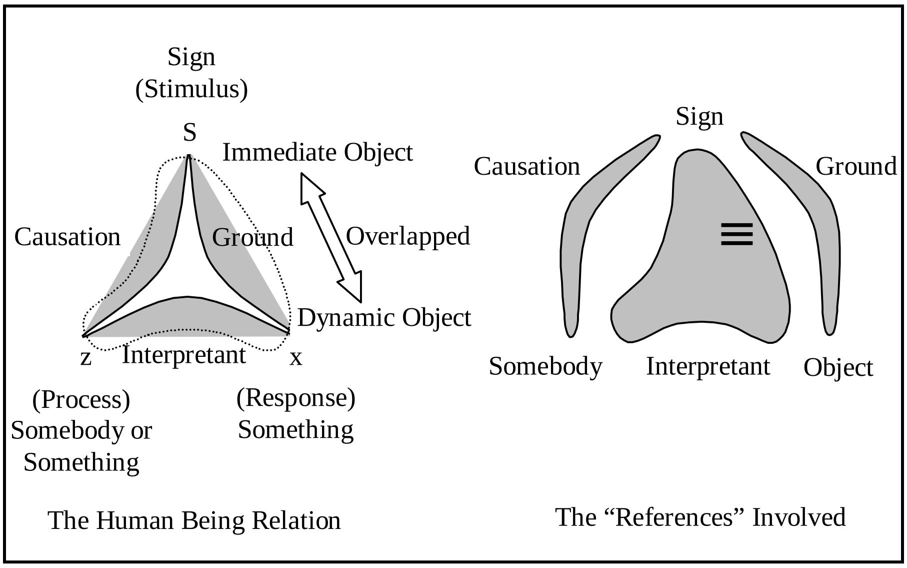


Figure 4. Comparing Fetzer's Human Being Relation with Peirce's "References".

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 (1991, p. 277). 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. This is in accord with Hausman's (1993, p. 72-73) interpretation of Peirce's triad.

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, for 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 subsection.

7.5 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.

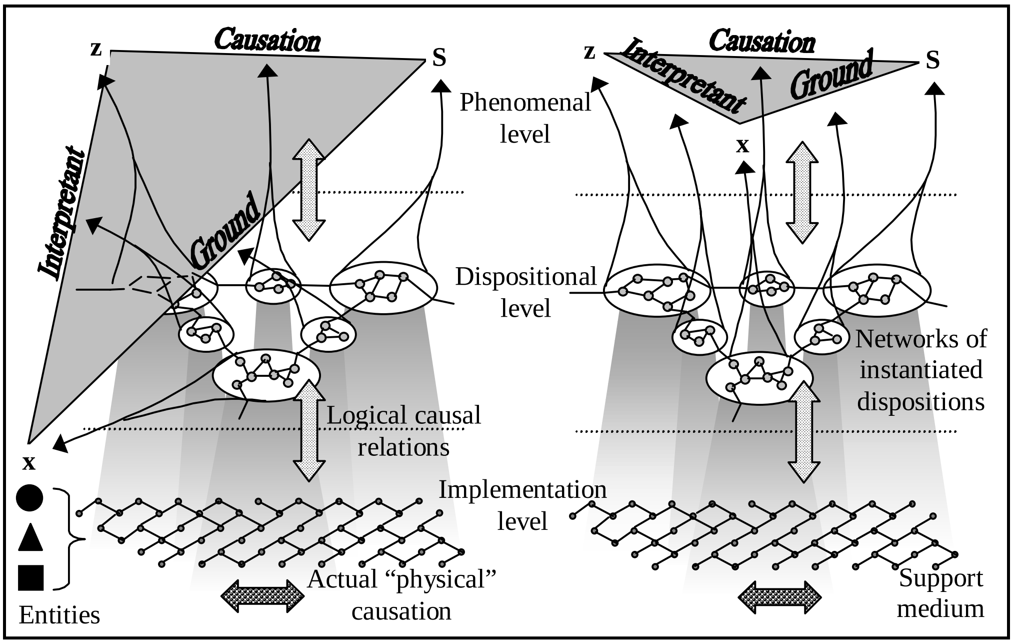


Figure 5. Ontological level interpretation of Fetzer's Human Being Relation.
The dotted lines represent level boundaries. Left diagram has a physical entity as dynamical object, right diagram has a sign (thought) as dynamical object.

Figure 5 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.

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 Whobrey (1999) 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 Whobrey (1999) for details. 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 Whobrey (1999). 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 (1993, p. 296) and Fetzer (1998).

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 (1993, p. 104, p. 125). Whobrey (1999) 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 6. 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.

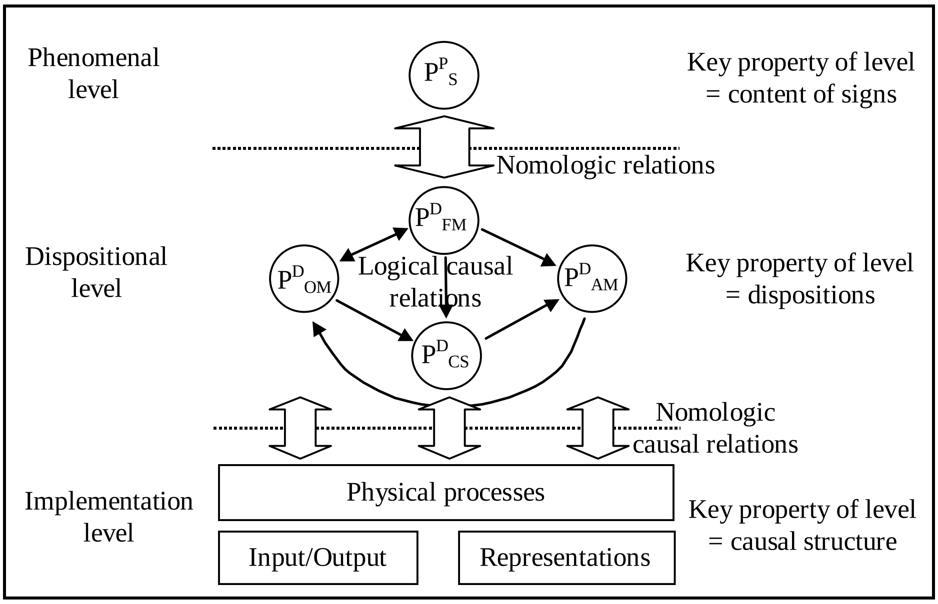


Figure 6. The system wide dispositional structure underlying the ground relation. P^P_S = property conveying phenomenal content of sign, P^D_{OM} = network of instantiated disposition properties for singularisation mechanism, P^D_{FM} likewise for fixing mechanism, P^D_{CS} category structure, and P^D_{AM} attentional mechanism – see Whobrey (1999).

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 essay.

8 Comparing Brains & Digital Machines

In studying the logical structure and the mathematical dynamical form of semiotic systems, biological minds can be treated as a type of numeric system and digital computers as a type of symbolic system. This section discusses criteria for comparing these two types of systems and under what conditions (relevant to the semiotic conception of mind) one can be seen as a replication of the other.

8.1 Characterising Conventional Kinds of Systems

This subsection reviews definitions for formal systems, digital machines, and the two main systems being analysed. Later this will be used to answer whether the simulation of mind is also replication.

8.1.1 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 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. (1986). 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 – see Machover (1996) and Hodges (1993) for further details.

8.1.2 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 (1998) and Wood (1987). Turing (1950) 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.”

8.1.3 Conventional Definition of Symbolic & Numeric Systems

Newell and Simon (1976) 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 (1995). 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 (1995).

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.

8.2 Comparing Numeric & Symbolic Systems

The objective now is to highlight the relevant differences and relationships between numeric and symbolic systems.

8.2.1 Working Definition for Numeric and Symbolic Systems

Consider general definitions for these two types of systems:

Define a numeric system S_N as consisting of two ontologic 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 ontologic 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.

8.2.2 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 given in the appendix of Whobrey (1999).

8.2.3 Comparing the Nature of the System 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 certain electrical circuits, are based on the dynamics of finite objects, which means that state changes are only statistically continuous.

8.2.4 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, e.g. as detailed in the definition of causal structure (D-CS) given in Whobrey (1999). 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 of Whobrey (1999)). 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 is tempting to describe T as a mapping, but this would be looking at it from the base level.

8.3 Replicating Systems

The goal of this subsection 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.

8.3.1 Simulation, Emulation, Replication

This subsection 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 Kuipers (1986). 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 (1978, p. 191).

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 (1990, p. 17) 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 (1984) 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.

8.3.2 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 of Whobrey (1999) 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 in Whobrey (1999) 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. 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 that 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, since from the dispositional approach causal structure is all that matters for reproducing mind, it follows that, to this extent at least, mild AI looks feasible. The next subsection considers whether it is practically possible to replicate mind on digital machines.

8.3.3 Producing a Mind in Future Computers

The potential ability of a digital machine to implement the hypothetical system as discussed above, and so replicate a human mind based on the dispositional approach, 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 Gordon Moore (1965). 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.

Table 2 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 DID	0.1 to 10
F	% neurons using flows	50 to 100
G	# processor instructions to replicate DID	10×B to 100×B
H	# DID updates per second	100 to 1000

Table 2. Factor estimates for replicating a human mind via dissociated instances of dispositions (DID).

The estimate for the number of neurons in a human brain comes from the DARPA neural network study – see DARPA (1988). 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 (1995)), 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 $50] \times$ # neurons MIPs = $[5\times10^4$ to $5\times10^{13}]$ MIPs.

A conservative estimate (taking the lower bounds only for A, B and H) gives:

Conservative estimate of processor performance: 5×10^8 MIPs.

This is of the same order of magnitude as the estimate suggested by Hans Moravec (1999, p.91) based on the performance of today's experimental robots.

Of equal importance is the processor storage capacity requirement (i.e. memory size). The storage capacity required to hold state information on each dissociated instance of a disposition is estimated as follows:

Processor storage capacity per DID: # bytes ≈ J × B, where J ≈ 2 bytes.

Lower and upper bound on storage requirement for replicating the human mind (scaling upper bound by ten to allow for sub-neuron structure of micro-tubules):

[2k to 20k] × # neurons bytes = $[2 \times 10^4$ to $2 \times 10^8]$ 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 is 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.

9 Semiotic Systems & Ontological Levels Revisited

From an operational perspective, the requirements developed for a semiotic system in Whobrey (1999), led to the conclusion that to generate mind and consciousness requires a system containing a hierarchy with at least three ontological levels. 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.

9.1 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. 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 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. 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.

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. 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.

9.2 Adding Ontological Levels to Fetzer's Laws

The analysis suggests an ontological level mediates between brain and mind. Consequently, Fetzer's laws (see Fetzer (1990, p. 284)), 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 (see Whobrey (1999)) makes relating a static brain-state with a mind-state problematical. The other laws can be supplemented in a similar manner.

9.3 Symbolic Replication of Semiotic Systems

Whobrey (1999) presented a theorem (T-WSCDE) for equating systems that effectively concluded with three conditions 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* discontinuous with respect to time. Finally, observable properties must be measurably equivalent to an observer from a given perspective and for a given resolution.

It was 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 mind 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. Secondly, dissociated disposition dynamics may permissibly be based on statistical properties of the underlying medium, such as averages. This holds so long as any granularity introduced into the dynamics is below the measurable resolution of the observer. In this case, this is relative to the first person perspective view of consciousness. 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, 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.

10 Summary

The field of AI was characterised in terms of 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. 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 ontological 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.

Consequently, the success of mild AI depends on whether mind 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. Above it was suggested that replicating the right causal structure in a system is sufficient for producing a semiotic system as a foundation for producing a mind.

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.

11 References

Baas, N.A. (1994). Emergence, hierarchies, and hyperstructures. In: *Artificial Life*, ed. C.G.Langton, Vol.III, 515-537. Addison-Wesley.

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.

Chandler, D. (1995). *The act of writing: a media theory approach*. Aberystwyth, University of Wales.

Cohen, B. Hardwood, W.T. Jackson, M.I. (1986). *The specification of complex system.* Addison-Wesley.

Copeland, B.J. (1993). *Artificial intelligence, a philosophical introduction*. Blackwell.

DARPA, (1988). *DARPA neural network study*. AFCEA International Press.

Dennett, D.C. (1978). *Brainstorms: philosophical essays on mind and psychology*. Bradford Books, Harvester Press.

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.

von Eckardt, B. (1993). *What is cognitive science?* Bradford Books, MIT Press.

Eco, U. (1976). *A theory of semiotics*. Indiana University Press.

Emmeche, C. Koppe, S. Stjernfelt, F. (1997). Explaining emergence: towards an ontology of levels. *Journal for General Philosophy of Science*. Vol.28, 83-119.

Fetzer, J.H. (1977). A world of dispositions. *Synthese*. Vol.34, 397-421.

Fetzer, J.H. (1981). *Scientific knowledge*. Dordrecht, Holland: D.Reidel.

Fetzer, J.H. (1986). Methodological individualism. *Synthese*. Vol.68, 99-128.

Fetzer, J.H. (1990). *Artificial intelligence: its scope and limits*. Kluwer Academic Publishers.

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.

Fetzer, J.H. (1993). *Philosophy of science*. Paragon House.

Fetzer, J.H. (1996). *Philosophy and cognitive science*. Paragon House.

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.

Fodor, J.A. (1990). *A theory of content and other essays*. MIT Press.

Fujiwara, S. Okajima, K. Tanaka, S. (1987). A Heteroassociative memory network with feedback connection. *IEEE International Conference on Neural Networks,* Vol.II, 711-718.

Hartshorne, C. Weiss, P. Burks, A. (1958). Collected papers of Charles Sanders Peirce. Eds. Harvard University Press.

Hausman, C.R. (1993). *Charles S. Peirce's evolutionary philosophy*. Cambridge University Press.

Hodges, W. (1993). *Model theory*. Cambridge University Press.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings National Academy Science USA, Biophysics*. Vol.79, 2554-2558.

Kosko, B. (1987). Adaptive bi-directional associative memories. *Applied Optics*. Vol.26, No.23, 4947-4960.

Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence*. Vol.29, 289-338.

Machover, M. (1996). *Set theory, logic and their limitations*. Cambridge University Press.

Mellor, D.H. (1993). Supervenience? No chance! Reply to Menuge. *Analysis*. Vol.53, No.4, 236-239.

Moore. G. (1965). *Moore's Law*. See www.intel.com/intel/museum.

Moravec, H. (1999). *Robot: mere machine to transcendent mind*. Oxford University Press.

Morris, C.W. (1938). *Foundations of the theory of signs*. Chicago University Press.

Nanopoulos, D.V. (1995). Theory of brain function, quantum mechanics and superstrings. *CERN-TH/95-128*. nanopoud@cernvm.cern.ch.

Newell, A. Simon, H.A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, Vol.3, 205-211.

Newell, A. (1982). The knowledge level. *Artificial Intelligence*. Vol.18, 87-127.

Norton, A. (1995). Dynamics: an introduction. In: *Mind as Motion*, ed. R.F.Port and T.van Gelder, MIT Press, 45-68.

Rapaport, W.J. (1998). How minds can be computational systems. *Journal of Experimental and Theoretical AI*. Vol. 10, 403-419.

Rey, G. (1997). *Contemporary philosophy of mind*. Blackwell.

Savage, J.E. (1998). *Models of computation*. Addison-Wesley.

Searle, J.R. (1984). *Minds, brains and science: the 1984 Reith lectures*. BBC publication.

Smullyan, R.M. (1994). *Diagonalization and self-reference*. Oxford Logic Guides, Clarendon Press. No.27.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, Vol. 59, 422-460.

Whobrey, D.J.R. (1999). Aspects of qualitative consciousness: a computer science perspective. PhD Thesis. Department of Computer Science. City University, London. Online version at: www.mildai.org.

Wood, D. (1987). *Theory of computation*. Harper and Row.