Chapter 4

The cool universe of digitality has absorbed the world of metaphor and metonymy.

Jean Baudrillard, *Simulations*.

Artificiality

4.1. Overview

In this chapter, the notion of artificiality is investigated. An interpretation of the concept due to Simon (1969) is briefly described and links between this concept and notions of functionalism and computationalism are examined. Attempts at unifying various "sciences of the artificial" (artificial analogues of natural phenomena) under Simon's concept are shown to be problematic, thereby necessitating a reformulation of the notion of artificiality. An alternative definition of the concept is developed via comparison and contrast with the idea of naturality (nature or `the natural'). Three interpretations of the ontological status of artificiality (simulation, replication, emulation/duplication/realization) are described and "weak" and "strong" notions of artificiality are defined. The Turing Test, a standard by which the ontological status of artificialities (artificial phenomena or artificials) may be evaluated with respect to corresponding naturalities (natural phenomena or naturals), is introduced. The possibility of unification under the new interpretation of artificiality is considered. Three natural phenomena (matter, life, mind) are then investigated and artificialities corresponding to these phenomena (AI, A-Life, A-Physics) are described. Finally, the idea of artificial reality (A/VR) is briefly examined.

4.2. What is *Artificiality* ?

The *Oxford English Dictionary* (Second Edition, 1989) defines artificiality as

> 1. The quality or state of being artificial; artificial character or condition. 2. with *pl*. an artificial thing or characteristic.

and artificial as

> 1. Made by or resulting from art or artifice; contrived, compassed, or brought about by constructed skill, and not spontaneously; not natural. *a*. Artificial in result, as well as in process. *b*. Of natural products or results, artificially produced, e.g. Artificial light. 2. Made by art in imitation of, or as substitute for, what is natural or real. 3. Merely made up; factitious; hence, feigned, fictitious. 6. Displaying much skill; *a*. of things: skilfuly made or contrived. *b*. of persons: skilled in constructive art, skilful.

Although the above definitions capture the manifold meanings of artificiality, they suffer from a lack of precision and do not, therefore, provide suitable foundations upon which to establish a science or philosophy of the artificial. Herbert Simon's *The Sciences of the Artificial* (1969, 1981) was probably the first major work in which an attempt at a precise definition of the concept of artificiality was made. Although it was originally conceived in the context of specific domains for a specific purpose (section 4.2.1), an appreciation of Simon's concept of artificiality is necessary for (at least) three reasons: First, it represents one of the earliest attempts at unifying phenomenal domains based on a postulated analogy between artificials and naturals; second, by focusing on issues of *production* (synthesis) over *interpretation* (analysis), Simon's concept of artificiality provides an appropriate point from which to initiate an investigation of the ontological issues associated with the distinction in *poiēsis* (coming-forth, bringing-forth) between naturals and artificials (chapter 6); third, and relatedly, this shift in focus serves to redefine the debate over "strong" *vs.* "weak" artificiality in terms of this ontological distinction as opposed to the conventional epistemological distinction between appearance and reality[1].

4.2.1. "The Sciences of the Artificial"

Simon (1969) originally developed his concept of artificiality in the context of three domains, viz. (1) economics, (2) psychology of cognition, (3) planning and engineering design; however, in (Simon,80), the concept of artificiality was extended to include cognitive science. It should be appreciated at the outset that a connection between Simon's concept of artificiality and the notion of *teleology* or goal-directedness is implied in the choice of domains. The significance of this link, more specifically, the problems to which it gives rise, are addressed in section 4.2.5.

The idea of artificiality can be introduced by way of an examination of the contrast between artificial and natural sciences. Simon (1981) offers the following definition of

---

[1] In chapter 6, it will be shown that the ontological artificial-natural distinction is *grounding* relative to the epistemological artificial-real distinction.

a *natural* science, viz.

> a natural science is a body of knowledge about some class of things - objects or phenomena - in the world: about the characteristics that they have; about how they behave and interact with each other. (p.3)

However, this definition applies equally well to the *artificial* sciences (that is, the sciences of artifacts). Simon clarifies the distinguishing characteristics of artificial sciences via an informal statement of the artificiality thesis, viz.

> certain phenomena are `artificial' in a very specific sense: they are as they are only because of a system's being molded, *by goals or purposes*, to the environment in which it lives [emphasis added]. (p.ix)

Thus, according to Simon, the defining characteristic of artifacts is that they are teleological. However, whether the teleology is ontical or epistemological, that is, whether artifacts are *actually* goal-directed or whether or it is merely the case that they cannot be *described* in non-teleological terms is an open issue[2]. For present purposes it should suffice to note that Simon's view of artificiality implies teleology, and teleology of a specific kind, viz. the adaptation of a system to its environment. Clearly, a connection between the concept of artificiality and Darwinian evolution by natural selection is intended. The establishment of such a link and, additionally, Simon's assertion that the role played by natural selection in the biological world is analogous to that played by rationality in the sciences of human behaviour facilitates the identification of naturality (or nature) with artificiality and visa-versa. However, Simon (1981) cautions against naive identifications of artifacts and naturals, thereby contesting a possible subsumption of artificiality into naturality. As he states,

> a forest may be a phenomenon of nature; a farm certainly is not. (p.5)

From this statement, it can be inferred that in order to prevent the subsumption of artifacts into nature, it is necessary to recognize man's products as artificial in the sense of *artifactual* (implying synthetic or man-made). However, it could be argued, assuming a *naturalistic* position in which everything is considered a part of nature, that man himself is a product of nature; consequently, man's products must also be natural products and hence, artificiality *can* be subsumed into naturality[3]. In fact, this position

[2] The opposing positions in the debate over the ontological status of intentionality - genuine (or originary) and *as-if* (or derivative) - are presented in (Dennett,87) (Searle,92) (Dennett,95) (Searle,95).

[3] Dennett (1995) appeals to an argument of this kind in his attempt at undermining the distinction between naturals and artificials and, in particular, Searle's (1992) argument for genuine (or originary) intentionality.

appears[4] to be supported by Simon himself, viz.

> those things we call artifacts are not apart from nature. They have no dispensation to ignore or violate natural law. (p.6)

Simon (1981) defines the `boundaries for the sciences of the artificial' as follows:

- Artificial things are synthesized (though not always or usually with full forethought) by man.
- Artificial things may imitate appearances in natural things while lacking, in one or many respects, the reality of the latter.
- Artificial things can be characterized in terms of functions, goals, adaptation.
- Artificial things are often discussed, particularly when they are being designed, in terms of imperatives as well as descriptives.

According to Simon, artificiality is (i) synthetic (or artifactual), (ii) apparent, (iii) teleological, and (iv) normatively-interpretable. However, it is an open question whether or not these terms can be applied to nature: On creationism (chapter 6), phenomena regarded as natural by human beings are held to have been synthesized by a supernatural agency; as a corollary of this, natural phenomena may be imitations (that is, appearances) of another order of reality (`supernature'). This view leads to a replacement of the simple artificial-natural and appearance-reality dualities by a possibly infinite series of such dualities. Hence, it is at least *conceivable* that naturality may be subsumed into artificiality[5]. This position is supported by the following observation:

> [artificial systems] are adapted to man's goals and purposes. They are what they are in order to serve man's desire to fly or to eat well. As man's aims change, so too do his artifacts - and vice-versa. (p.6)

However, from this statement it should be appreciated that nature could only be subsumed into artificiality if the former could be *objectively* identified as teleological. Since this may be impossible (logically or/and empirically), it provides a basis upon which to differentiate artificiality from naturality. For example, the behaviours of artificial systems, biological systems and human beings are distinguishable with respect to the nature of the associated teleology: Rationally-directed human behaviour is *a priori* goal-directed or intentional; adaptive evolutionary behaviour (assuming the standard

---

4 In fact, this argument is a *non sequitur* since the *consistency* of artifacts with natural law does not entail the *subsumption* (and identification) of artificials into naturals: It is quite conceivable, under the assumption of pluralism (chapters 5 and 6), that artifacts are ontologically-*distinct* from yet also ontologically-*dependent* on naturals.

5 This position is consistent with the adoption of an idealistic solution to the measurement problem in quantum theory (section 4.7.4.4). On this view, the creation of matter and, thereby, the `natural' (physical) world, is held to be a consequence of the collapse of the quantum wave function by consciousness (Goswami,93).

Darwinian account) is *a posteriori* goal-directed or *teleonomic*[6] (Mayr,82). Thus, the epistemic status of the teleology in natural and artificial systems with respect to human beings is different. This, it is argued, is a *necessary* consequence of the difference in ontical status of artifacts and naturals: The former are man-made whereas the latter are not. Thus, in contradistinction to assertions by Levin (1979) and McGinn (1987) to the effect that the intrinsic nature or *Being* of an object is logically independent of the manner of its genesis or *becoming*, it is asserted that *how* things come into existence, what Heidegger refers to as their *historicity* (chapters 1 and 6), is critically significant; to assert otherwise is to assume an *a priori* Platonic position in which the phenomenal substance-attribute duality is considered ontologically primitive, and the coming-forth or *poiēsis* of substances and attributes (the latter of which are contextually-defined on the Heideggerian view) is regarded as irrelevant (chapter 6). However, adopting a substance-attribute position engenders serious problems: Specifically, variants of Zeno's paradoxes, viz. how *dynamic* processes can arise in *static* substances, a problem which led Plato to deny the reality of change (Rescher,96); more generally, the broader problem of emergence (chapter 3).

Close examination of the above statements reveals Simon's concept of artificiality as somewhat ambiguous with respect to its ontological status: On the one hand, artificiality is viewed as, in some (possibly weak) sense, identical to naturality, viz. human rationality is *analogous* to Darwinian evolution; on the other, it is maintained that artificiality and naturality are distinct, viz. a farm is *not* a natural object. However, if Simon's concept of artificiality is to provide a basis for unifying matter, life and mind when the latter are defined in computational terms, it is necessary that an isomorphism (that is, one-one mapping) of some kind exist between naturality and artificiality. This leads directly to a consideration of the link between artificiality and functionalism.

4.2.2. Artificiality and Functionalism

Functionalism is the view that the state of a system is defined by its functional role, that is, in terms of the causal relationships between the state, other states, and inputs and outputs to and from the system (chapters 1 and 2). A functionalist account of systems is consistent with both top-down and bottom-up approaches to system construction, that is, with rationally-directed design or/and emergent evolution; hence, it is readily linked to the notion of artificiality as conceived by Simon. Functionalism necessitates adopting a `black box' conception of systems in which the latter are held to be hierarchically decomposable or *fractionable* (Rosen,93) into collections of sub-systems (chapter 3) with

---

[6] According to Campbell (1985), "*teleonomy* asserts that the function of a biological structure .. is not what the structure will usefully do for the organism but the effect that the homologous structure in the ancestors had on survival in past generations. Teleonomic function thereby refers to past effects instead of present purpose." (p.153)

sub-systems at the lowest (most primitive) level in the hierarchy being realized directly (physically). Furthermore, systems are held to be *partitionable*, by which it is meant that the *internal* `environment' of a system is separable from its *external* environment. Since the notion of partitioning is central to Simon's concept of artificiality and the possibility of functionalism, it merits further examination.

The concept linking artificiality, functionalism and partitioning is *teleology*. According to Simon (1981), fufilment of purpose or adaptation by an artificial system to a goal involves a relation among three terms:

1. the *goal* (teleology or purpose) of the system

2. the *character* (or ontology, that is, internal `environment') of the system

3. the *environment* of the system

Nature impinges on the artifact through two of the three relations that characterise it, viz. the being of the artifact itself (2), and the external environment in which it acts (3). Hence, an artifact may be conceived as an *interface* between an internal environment (the substance and organization of the artifact), and an external environment (the surroundings in which it operates). The external environment can therefore be viewed as determining the conditions for goal-attainment. As Simon states, "if the inner environment is appropriate to the outer environment, or vice-versa, the artifact will serve its intended purpose." (p.9) There are essentially two advantages in partitioning the internal from the external environment when studying an adaptive artificial or natural system: (i) *prediction* of behaviour from a system's goals and its external environment requiring only minimal assumptions about the internal environment; (ii) *homeostasis* or maintenance of system state by isolating the internal environment of the artifact from the detail of its external environment (Simon,81). Moreover, according to Simon,

> in the best of all possible worlds - at least for a designer - we might even hope to combine the two sets of advantages we have described that derive from factoring an adaptive system into goals, outer environment, and inner environment. We might hope to be able to characterize the main properties of the system and its behaviour without elaborating the detail of *either* the outer or inner environments. We might look toward a science of the artificial that would depend on the relative simplicity of the interface as its primary source of abstraction and generality. (p.12)

As a corollary to (i) and (ii), it becomes possible for different implementations of internal environments to accomplish identical or similar goals in identical or similar external environments. For example,

> airplanes and birds, dolphins and tunafish, weight-driven clocks and spring-driven clocks, electrical relays and transistors. (p.11)

This position is equivalent to a claim for *multiple-realizability*, that is, the existence of a one-many mapping between phenomena and media in which such phenomena can be realized. A *sufficient* condition for multiple-realizability is, therefore, production of *behaviour* (or functionality) identical to that of the target phenomenon. On an *essentialist* view (in which all members of a class share a set of common properties), this may be formally stated as follows:

> A system *s* belongs to a phenomenal class *P* if *s* generates a behaviour *B* which is identical to that generated by *all* known members of *P*, that is, $\{\forall x \in P \mid B(s) \equiv B(x), s \neq x\} \Leftrightarrow s \in P$.

Assuming a *non-essentialist* position (in which some members of a class share common properties):

> A system *s* belongs to a phenomenal class *P* if *s* generates a behaviour *B* which is identical to that generated by *at least one* known member of *P*, that is, $\{\exists x \in P \mid B(s) \equiv B(x), s \neq x\} \Leftrightarrow s \in P$.

According to Simon (1981), "resemblance in behaviour of systems without identity of the inner systems is particularly feasible if the aspects in which we are interested arise out of the *organization* of the parts, independently of all but a few properties of the individual components." (p.21) Thus, the teleology (or *final causality*) associated with artificiality and functionalism is reducible to organization (or *formal causality*) on Simon's definition of the concept. This is significant since it implies the tacit adoption of Platonism (chapter 2) in postulating (1) the *inessentiality* of the connection between form (behaviour) and matter[7] (medium) and (2) the relative significance of the former over the latter; in short, multiple-realizability is grounded in the assumption that the form-matter or behaviour-medium relation is *contingent* rather than *necessary*.

Although Simon identifies artificial with artifactual (or man-made), he does not hold this term as *necessarily* equivalent to *synthetic*. The distinction between the two is clarified by way of an example, viz.

> a gem made of glass colored to resemble sapphire would be called artificial, while a man-made gem chemically indistinguishable from sapphire would be called synthetic. (p.7)

On this basis, *artificing* would seem to allow for multiple-realizability with respect to both substance and process whereas *synthesis* allows for multiple-realizability only with respect to process[8]; artificiality is, therefore, a less constrained (or constraining) concept than synthesis. As Simon (1981) states,

---

7 On the "strong" or *ontological* version of the Church-Turing Thesis described in chapter 2, matter is reducible to computation and computation, in turn, to form.

8 In chapters 6 and 7, a similar distinction is made between "hard" (or pure) and "soft" (or impure) artifacts: In the former, both matter and form are made while in the latter only form is made, matter being *given*.

> artificiality connotes *perceptual* [epistemological] similarity but *essential* [ontological] difference, resemblance from without rather than within [emphasis added]. (p.17)

He offers the following clarification of this view, viz. "the artificial object *imitates* the real by turning the same face to the outer system, by adapting, relative to the same goals, to comparable ranges of external tasks [emphasis added]" (p.17). This leads directly to consideration of an important issue associated with artificiality and functionalism which concerns the ontological status of artificialities (that is, artificial analogues of natural phenomena): Whether a functional representation of a phenomenon - the causal `role' - when suitably instantiated by a physical substrate or medium - the causal `occupant' (Sterelny,90) - is a simulation or a realization of the phenomenon (section 4.3.3). Even if a functional-realist position is adopted in which function is viewed as *intrinsic* to systems (chapters 2 and 3), resolution of the problem depends on the availability of a complete functional description of all *necessary* causal relations at every level in the systemic hierarchy. If it is impossible to attain such a description then, ultimately, in some future unforseen context, the behaviour of a functional realization (that is, an instantiation) of the original systemic phenomenon will begin to deviate from the observed behaviour of the phenomenon itself. In this event, the inadequacies of the former, with respect to its capacity for attaining functional isomorphism with the latter, will become manifest. This leads to what may be described as `phenomenal leakage' or the appearance of `side effects'. Simon (1981) describes it as follows:

> Often we shall have to be satisfied with meeting the design objectives only approximately. Then the properties of the inner system will 'show though'. That is, the behaviour of the system will only partly respond to the task environment; partly it will respond to the limiting properties of the inner system. (p.16)

Thus, side effects occur if an essentialist-functionalist perspective is adopted in which causal relations within systems are distinguished as functional and non-functional. However, this distinction leads to further problems (other types of side effect or phenomenal leakage) since it requires specification of either a *context*-in-which (objectivist) or *perspective*-from-which (subjectivist) the distinction can be made. These issues will be discussed in greater detail in chapter 7 when the distinction between "hard" (or *pure*) and "soft" (or *impure*) artifacts is examined.

4.2.3. Artificiality and Computationalism

In the previous sections, the connection between artificiality and functionalism has been examined. A possible connection between artificiality and computationalism is implied by the fact that the latter is a type of functionalism (chapters 1 and 2). According to Simon (1981), however, this link is more than just possible; its establishment follows almost inevitably once the issue of *substrates* supporting functionality is addressed, viz.

> no artefact devised by man is so convenient for this kind of functional description as a digital computer. It is truly protean, for almost the only ones of its properties that are detectable in its

> behaviour (when it is operating properly!) are the *organizational* properties. (p.22)

This view is further reinforced by Simon's identification of the essence of computation with the mutually-related concepts of *function* and *organization*:

> A computer is an organization of elementary functional components in which, to a high approximation, only the function performed by those components is relevant to the behaviour of the whole system. (p.22)

The link between computers and physical symbol systems (chapter 2) is described as follows: "the computer is a member of an important family of artifacts called symbol systems, or more explicitly, *physical symbol systems* [PSSs]. Another important member of the family (some of us think, anthropomorphically, it is the *most* important) is the human mind and brain." (pp.26-27) Simon's conception of the relation between artificiality, PSSs, and digital computers may be formally stated as follows: $\{A \cap P\} \supset C$, where A denotes the set of all artificialities, P the set of all PSSs, and C the set of all digital computers. (The possibility of natural PSSs and artificial non-PSSs is supported on this definition.) Crucially, Simon maintains that

> computers have transported symbol systems from the platonic heaven of ideas to the empirical world of actual processes carried out by machines or brains, or by the two of them working together. (p.28)

Thus, there is an *almost* natural link between artificiality and computationalism[9].

4.2.4. Unification of Domains

Collingwood (1945) has argued for two senses by which nature (or `naturality') may be understood, viz. nature as (i) the *principle* behind natural things, something which makes its possessor behave the way it does, the source of which, by definition, is within the thing itself, and (ii) the sum total or *aggregate* of natural things. By analogy, artificiality may be understood as (i) the principle underlying the production of artifacts (via design or evolution as discussed in section 4.2.1) and (ii) the sum total or aggregate of artifacts, viz. artificiality as "the *fields of knowledge* of which the subject-matter is partly man-made [emphasis added]" (Vickers,81). Thus, in addition to its definition as a *principle*, viz. "a Theory of Design, where `artifacts' (the fabricated implements of humankind) are endowed with `intelligence'" (Van Gigch,90), whereby `intelligence' is understood adaptability, artificiality can also be used to denote the unification of artificial domains, understood as implying the identification and subsequent consolidation of artificialities (that is, artifactual analogues of natural phenomena) under the unifying *principle* of artificiality as artifactuality. This dual interpretation enables artificiality to meet one of the requirements of a unifying conceptual framework: On the one hand, it is defined in

---

9 Only *almost* since it is an objective of this thesis to argue that artificiality and computation are anthropocentrically artifactual (chapter 7), and hence, not natural (except in a trivial, derivative sense).

terms of the *universal* as opposed to the *particular*, and is concerned with the *essence* (or *what-ness*) of artificial domains (ontology) and how this can be known (epistemology); on the other, it simultaneously refers to the totality of such domains as instances of phenomenal artificiality. (On the basis of the definition of artificiality given in sections 4.2.1 and 4.2.2, the artificial `science' associated with an artificial `domain' may be identified with the organized body of theoretical and empirical knowledge about adaptation with respect to systems belonging to the given domain.)

As stated previously in section 4.2.1, the concept of artificiality can be applied to anything that can be regarded as adapted to some situation, in particular to the living things that are assumed to have evolved as a consequence of natural selection. On this basis, attempts have been made to extend the concept both `upwards' and `downwards' in the phenomenal hierarchy via the adoption of an evolutionary epistemology: For example, Darwinian explanations of consciousness, language and culture (Dennett,95) on the one hand, and evolutionary accounts of cosmology (Smolin,97) on the other. For present purposes it suffices to accept that Simon's concept of artificiality appears to provide a basis for unifying matter, life and mind contingent on the possibility of teleological interpretations of such phenomena.

4.2.5. Problems with Simon's Concept

Prem (1995) argues that Simon's concept of artificiality leads to the following relation: Artificial is to normative and synthetic as natural is to descriptive and analytic. He maintains this on the basis of the following statement by Simon (1969), in which it is asserted that the scientist of the artificial

> is concerned with how things *ought* to be - how they ought to be in order to *attain goals* and to function (p.7).

Teleology (functionality, goal-directedness, adaptation) is central to Simon's concept of artificiality and it is precisely this fact that renders this concept problematic, preventing its adoption as a means by which to unify artificialities (that is, artificial analogues of natural phenomena). However, this appears to contradict the closing statement of the previous section to the effect that unification of phenomena is possible under Simon's concept. In fact, this is not the case since it was explicitly stated that unification is only possible *if* phenomena can be interpreted *teleologically*. The problem with this position is that on the current scientific view, matter is defined in non-teleological terms. Hence, unless one is prepared to endorse some form of neo-Aristotelianism in which *entelechies* (that is, self-directing agencies) are permitted, artificial matter cannot be integrated with other intrinsically teleological artificialities (such as life and mind[10]) under Simon's

---

[10] On the standard scientific view, both life and mind are teleological; however, teleology in the former is *teleonomic* (that is, *a posteriori* or epistemically teleological) whereas - on both folk psychological and sophisticated philosophical accounts (Searle,82) (Searle,92) - teleology in the latter is *intentionalistic* (that is,

concept of artificiality. This *teleology problem* arises from Simon's adoption of a mechanistic interpretation of biology and an organismic interpretation of machines (Newell,76) and his postulation of a *hybrid* concept grounded in these notions as the metaphysical basis upon which to formulate a concept of artificiality. However, as a consequence of the teleology problem (with respect to matter), Simon's ontical primitives, viz. functional artifacts, cannot be ontically primitive with respect to the *unified* framework of computationally emergent artificiality presented in chapter 5. For this reason, it is necessary to consider alternative conceptions of artificiality.

Negrotti (1991e) presents the following list of characteristics as definitive of the concept of artificiality:

1. An artificial device is a machine which reproduces some essential functions of a sub-system of a natural system.

2. In the artificial sub-system the number of homologous components is reduced.

3. In the artificial sub-system the homologous components are structurally different.

4. In the artificial sub-system new components may appear.

5. In the artificial sub-system new types of internal and external relations may appear.

6. In the artificial sub-system, some kinds of internal or external relations are lost, and others may be added.

7. Every artificial device is a machine, but not all machines are to be conceived as *artificial* devices.

8. The performances of an artificial device usually show a different spectrum (sometimes wider and sometimes narrower) compared to the one shown by the correspondent natural sub-system.

9. The research and the development of enhancements of the artificial device consist in the deepening of its own artificial characteristics as such and, usually, this moves the new generations of the device further and further from the natural sub-system.

10. The artificial device will be accepted as a good reproduction of the natural system if, and only if, its functioning allows a good reproduction of the main and *essential* features and performances of the natural sub-system.

Detailed examination of the entries in the above list reveals a number of issues which are central to the debate over artificiality. For example, in (1) it is assumed that nature is *ontically* systemic; moreover, natural systems are assumed to be nearly-decomposable (chapter 3); in (2) it is implied that artificiality involves *abstraction* from naturality; (3-6) support the concept of *multiple-realizability*; (7) is consistent with a *mechanistic* interpretation of nature; (8) indicates the likelihood of *`side-effects'* (section 4.2.2); (9) describes *artifact evolution*; and in (10) a *substantialist* (or substance-attribute) ontology

a priori or ontically teleological).

is assumed. Negrotti's concept of artificiality appears to be an improvement on that due to Simon since teleology is not explicitly associated with artificiality. However, (1) makes reference to functionality which appears to be a teleological concept[11]. Ultimately, what is required is a conception of artificiality in which artifacts are defined in non-functional terms.

4.3. Artificial and Natural

One possible approach by means of which a non-teleological conception of artificiality may be formulated is via an examination of related and opposing terms. A natural place to begin is with the idea of *nature* itself. However, scientific and philosophical literature associated with this concept is extremely vast and a survey of it is beyond the aim and scope of this study[12]. Consequently, recourse has been made to definitions provided by various philosophical dictionaries. This move may be justified on the grounds that what is being sought is a new understanding of naturality and artificiality rather than a historical understanding of these concepts[13]. As Heidegger might say, what is required is to think these ideas afresh, to *question* concerning artificiality.

4.3.1. What is *Nature* ?

The *Oxford Companion to Philosophy* (1995) defines *natural* as follows:

> Belonging to or concerned with the world of nature, and so accessible to investigation by the natural sciences. `Natural' may be contrasted with various terms such as `artificial', `unnatural', `supernatural', `non-natural'. The first three of these occur in ordinary language, though `unnatural' in particular leads to problems about its real meaning. But `non-natural' is a philosopher's term, and (non-naturalistic) is the usual contrast terms to `natural' or `naturalistic' in philosophy. Roughly it refers to what cannot be studied by the methods of the natural sciences, or defined in terms appropriate to them, and is applied to subject-matters that are essentially abstract, or outside space and time.

Lacey (1995) clarifies the philosophical meaning of `non-natural' by stating the

---

11 This view is contested by Searle (1992, 1995) who maintains, on Darwinian grounds, that the teleology of natural systems incapable of supporting conscious intentional states and processes is *extrinsic*, that is, epistemically imposed by an external observer. This position is supported by Nagel (1961), Simon (1971) and the contributors to (Rescher,86). However, it is important to appreciate that even if functional accounts of phenomena in non-teleological (or Darwinian) terms are possible, there remains the problem of explaining the *origin* of the replicating species (functional entities) in such accounts. In short, naturalistic-functionalism, while discharging the *proxima*l (or local) teleology problem (of particulars), does nor discharge the *distal* (or global) form of this problem; hence, the persistence of the Intelligent Design argument under `enlightened' creationism (chapter 6).

12 Collingwood (1945) provides a good introduction to this subject.

13 Ironically, in chapter 6, it will be argued that this *new* understanding of (the distinction between) naturality and artificiality is grounded in the notion of *historicity*.

following as a necessary condition for naturalism:

> The important thing for the naturalist in the metaphysical sphere is that the world should be a unity in the sense of being amenable to a *unified* study which can be called the study of nature, though it may not always be easy to say what counts as a sufficient degree of unification. (pp.604-605)

Ruse (1995) presents the following three definitions of *nature*:

- Everything that there is in the physical world of experience, broadly construed. The universe and its constituents, in short.
- The living or animate as contrasted with the non-living or inanimate.
- That which sees everything, especially the organic world, set off against humans and the consequences of their labours.

Angeles (1981) presents the following comprehensive definition of *nature*:

> (L., *natura*, from *natus*, `born', `produced'; the past participle of *nasci*, `to be born'). 1. The universe. The existing system of all that there is in time and space. Everything that happens (good and bad). 2. The powers (forces) that cause (produce, create) existing phenomena. 3. The origin (or foundation) of everything. 4. The ground for the explanation of things. 5. The essence of a thing; its essential characteristics. 6. The natural endowments of a thing. 7. The physical constitution of a thing. 8. An original, primitive state of things unadulterated and uncultivated by humans. That which happens without human interference.

Flew (1979) defines *nature* as follows "the content, structure, and development of the spatio-temporal world *as it is in itself*. Sometimes man is allowed to be part of Nature and sometimes not, whereas for the theist Nature is always the work but never a part of God." On this view, nature is identified with the Kantian noumenal world or reality as contrasted with the phenomenal world of appearance (chapter 1). In Aristotelian philosophy, nature (or naturality) is also identified with the totality of natural beings (Runes,60); this is analogous to the way artificiality may be defined as the totality of artificial things (section 4.2.4).

An alternative approach is to define naturality negatively by comparison and contrast with a number of opposing terms (Runes,60), viz.

- artificial
- unnatural (or abnormal)
- conventional (or customary)
- intellectual (or deliberate)
- subjective

Additionally, both Flew (1979) and Runes (1960) identify a link between naturality and the *absence* of will or volition[14]. This interpretation of naturality can be restated in Heideggerian terms as follows: `natural' implies transparent (non-reflective, non-thematic) coping in-the-world (chapters 1 and 6). The link between naturality and volition can also be formulated in essentialist terms, viz. naturality as associated with *necessity*[15] and artificiality with *contingency*. Such a scheme is consistent with Simon's interpretation of artificiality and the concept of multiple realizability (section 4.2.2).

Another way of defining natural is in theistically-nomic terms, viz. nature as the law-like contrasted with supernature as the miraculous. However, this contrast is readily shown to be problematic on epistemological grounds: What might be deemed a miraculous event at one point in time could be given a naturalistic[16] explanation at a later more advanced stage of knowledge[17].

According to the dictionary definitions given in section 4.2, `artificial' can be associated with the following: (i) design in process, (ii) design in product, (iii) imitation of nature, or/and (iv) substitution of nature. In support of this view, Hepburn (1967b) identifies artificiality with the imposition of some kind of `alien' causality on a natural causal substrate, viz.

> in one group of cases the natural is contrasted with the artificial or conventional. This contrast requires some conception of how the object or organism would behave by reason of its immanent causality alone, the causal factors that are particular to that type of thing and make it whatever it is - a stone, a fish, or a man. *The artificial and conventional are seen as interferences, modifying by an alien causality the characteristic patterns of behaviour* [emphasis added]. (p.454)

On the basis of this definition, artificiality can be identified with a network of causal relations that *supervene* (chapter 3) on the causal network of the natural substrate. Although this formulation of the concept of artificiality does not make explicit reference to functionality, its identification with the notion of supervenience supports the establishment of such a link. If such a definition is adopted, it becomes identical with Simon's original concept. Hepburn further maintains that separating nature from artifact is not easy since "organism and environment, individual and cultural climate, are in ceaseless interplay." (p.454) This point is extremely important since it appears to (i) support the functionalist thesis and (ii) undermine the distinction between artificiality and naturality. Detailed arguments against this inference are presented in chapter 6 based on

---

14 On this view, sleeping would be considered natural whereas deciding to write a novel would not.

15 For example, sleep is *necessary* for the maintenance of human biological existence.

16 Clearly, this argument appeals to a God-of-the-gaps conceptualization of supernaturality.

17 While this is certainly possible for events occurring *within* the world, the possibility remains of a miraculous interpretation of the world *as a whole*, if the latter is a meaningful concept. The implications of this Kantian distinction are examined in chapter 6.

(1) the essential, historical and unitary relation between Being and becoming and (2) the distinction in Being between naturals and artificials the follows from the distinction in their respective modes of *poiēsis* (coming-forth, bringing-forth). For present purposes, it is merely assumed that the distinction is valid.

The various distinctions between naturality and its opposing terms[18] are graphically summarized in Fig 4.1:

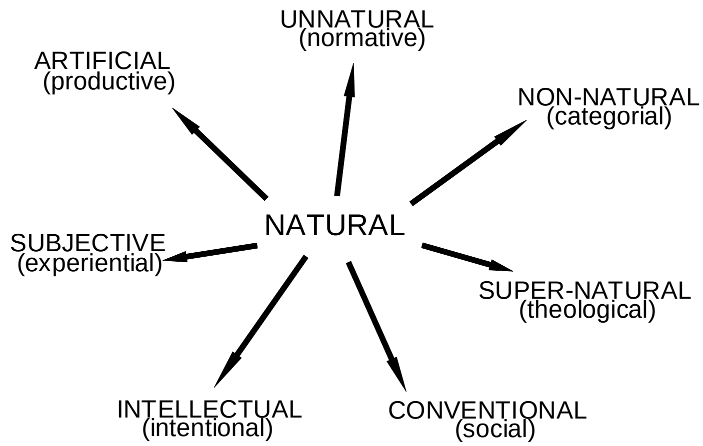

Fig 4.1 Naturality and its oppositions.

Specific distinctions between naturality and artificiality are listed in Table 4.1:

NATURALITY as	*ARTIFICIALITY* as
The World Living things Non-artifactuals The set of naturals Causality Being Modes (or ways) of Being Ontology Noumena	Processes of design Products of design The set of artifacts (or artifactuals) Imitation Substitution Phenomena

[18] In this connection, it is interesting to note with Heidegger (1939) that "whatever range has been attributed to the word `nature' in the various ages of Western history, in each case the word contains an interpretation of beings as a whole, even when `nature' seems to be meant as only one term in a dichotomy. In all such dichotomies, `nature' is not just one of two equal terms but `essentially' holds the position of priority, inasmuch as the other terms are always and primarily differentiated by contrast with - and therefore are determined by - *nature*." (p.184)

Table 4.1 Naturality-artificiality distinctions.

4.3.2. Towards A New Concept of Artificiality

Sokolowski (1988) presents the following example in order to distinguish two ways in which the term `artificial' may be used:

> The word artificial is used in one sense when it is applied, say, to flowers, and in another sense when it is applied to light. In both cases something is called artificial because it is fabricated. But in the first usage artificial means that the thing seems to be, but really is not, what it looks like. The artificial is merely the apparent; it just shows how something else looks. Artificial flowers are only paper, not flowers at all; anyone who takes them to be flowers is mistaken. But artificial light is light and it does illuminate. It is fabricated as a substitute for natural light, but once fabricated it is what it seems to be. In this sense the artificial is not the merely apparent, not simply an imitation of something else. The appearance of the thing reveals what it is, not how something else looks. (p.45)

Artificiality is, therefore, interpretable in two senses: (1) artificiality as *appearance* and (2) artificiality as *reality*. (On both schemes, the *artifactual* aspect of artificiality, that is, its *made-ness* (chapters 6 and 7), is accepted as an irreducible fact which cannot be the subject of further study. This follows, according to Heidegger (1959), from the reduction of Being to a mere binary predicate: A thing either exists or does not exist; its mode of *poiēsis* (becoming, coming-forth) is considered irrelevant to its Being which is interpreted in essentialist terms.) On the first interpretation, *material* identity is a *necessary* condition for phenomenal identity: A paper flower and a real flower are not phenomenally identical since natural flowers are not made of paper. On the second interpretation, *functional* identity is a *sufficient* condition for phenomenal identity. The physicality of the realizing substrate is viewed as a merely *contingent* factor; thus, artificial light and natural light are both instances of the class of illuminating things. The two interpretations follow from the adoption of different ontological schemes: (1) assumes a "weak"[19] form of *identism* or *physicalism*; (2) is grounded in *functionalism* and the concept of multiple-realizability. It is significant to note that, again, a connection has emerged between functionality and artificiality. Furthermore, the ontological question concerning artificiality has been replaced by an epistemological question: Questioning concerning the possible difference in *poiēsis* (coming-forth) of phenomena (naturals and artificials) has given way to inquiry into the reality status of such phenomena, thereby undermining the *productive* foundation underlying Simon's concept of artificiality.

The reconceptualization of the question of artificiality in terms of the epistemological problem of the relation between phenomena and noumena or appearance and reality (chapter 1) leads directly to an examination of the notion of simulation and other related concepts.

---

19 Here, `weak' is taken to mean "not necessitating the epistemological reduction of phenomena".

4.3.3. Simulation, Replication and Emulation

Perhaps the earliest reference to the concept of simulation is Plato's notion of *mimesis*, implying the imitation and representation of something by something else. The modern technical literature on simulation is vast; however, a precise formulation of the concept remains somewhat elusive although attempts have been made in this regard (Rasmussen,95). Various definitions of the concept have been proposed in the scientific and philosophical literature. For example, Searle (1980) defines simulation in terms of `black box' functionality, viz. "the right input and output and a program in the middle that transforms the former to the latter." This black box conception is supported by Prem (1995) who maintains that "a characteristic feature of simulations is that not all entities decode into something in the real world." (p.3) Simulation therefore implies abstraction which in turn implies functionalism. It is significant to note, following Webb (1991), that simulation can be distinguished from *replication*: In the former, causal capacities and structure are *represented* whereas in the latter they are *reproduced*. The distinction follows from (i) the assertion that reproduction is non-representational and (ii) the observation that simulation is grounded in representation or *modelling*. The latter necessitates that reference be made to the *interpretative* (epistemological) and *productive* (ontological) roles of a *modeller* (chapter 7), viz.

> simulating a system involves *someone* constructing a correspondence between the causal capacities and structure of a system and the capacities and structure of the simulation, so that the simulation produces behaviour that corresponds to that of the system. (p.248)

Webb states the conditions for replication in functionalist terms as follows: "For the replication of the behaviour to occur, there must *at some level* be an identity of capacities and structure between the replication and the system." (p.249) However, this position is problematic since it is unclear whether the identity relation is ontic or epistemic, that is, intrinsic to the system or projected onto the system by the modeller in its role as *interpreter*. Fetzer (1990) presents a somewhat different argument based on a semiotic-representationalist position: The simulation-replication distinction does not arise as a consequence of the representation-reproduction distinction, but as a result of the distinction between extrinsic and intrinsic representations respectively. As he states,

> simulations might .. be said to involve figurative representations (from an `external' point of view), whereas replications involve literal representations (from an `internal' point of view) - assuming that [the latter] is something that humans can do. (p.63)

A third possibility is *emulation* (or duplication). Fetzer (1990) holds that emulation involves "affecting the right functions by means of the same - or similar - processes implemented within the same medium. A relation of emulation between systems of different kinds thus entails that they be constituted of similar material (or components)." (p.18). Emulation is consistent with weak identism/physicalism (section 4.3.2) since both positions assert identity of material substrate (medium) as a necessary condition for phenomenal identity. The difference between simulations, replications and emulations can be stated as follows:

- *simulation* of a natural system involves capturing the functional connections between inputs and outputs of the system in a causal model.

- *replication* of a natural system involves (1) capturing the functional connections between inputs and outputs of the system in a causal model (2) via processes that correspond to, viz. are the same as or are similar to, those of the natural system.

- *emulation* of a natural system involves (1) capturing the functional connections between inputs and outputs of the system in a causal model (2) via processes that correspond to, viz. are the same as or are similar to, those of the natural system (3) in the same medium that the natural system is realized.

Hofstadter (1981b) differentiates between simulation and emulation in the context of a computational medium (substrate) as follows:

> The verb `emulate' is reserved for simulations, by a computer, of another computer, while `simulate' refers to the modelling of other phenomena, such as hurricanes, population curves, national elections, or even computer users. (p.380)

Furthermore, he states that "simulation is almost always approximate, depending on the nature of the model of the phenomenon in question, whereas emulation is in a deep sense exact." (p.380) Emulation makes possible the functional *embedding* of tokens (instances of a type) within other tokens conditional on the assumption of type identity between the tokens: For example, it is possible to embed computers (Turing machines) within other computers (Turing machines) leading ultimately to the construction of a `virtual machine' hierarchy (chapters 2 and 5). The difference between simulation and emulation can be stated as follows:

> *simulation involves a mapping between heterogeneous types while emulation involves a mapping between tokens (or instances) of the same type.*

A corollary of the above statement is that computers are held to be capable of simulating non-computational things and emulating computational things. This point is extremely important since "strong" artificiality (section 4.3.5) rests on the assumption that what *appears* (phenomenally, epistemically) to be non-computational is in *reality* (noumenally, ontically) computational, thereby allowing for the possibility of emulating - as opposed to merely simulating - non-computational *phenomena*.

Tipler (1994) presents a different interpretation of the difference between simulations and emulations:

> In a simulation, a mathematical model of the physical object under study is coded in a program. The model includes as many attributes of the real physical object as possible (limited of course by the knowledge of these attributes and also by the capacity of the computer). The running of the program evolves the model in time. If the initial model is accurate, if enough key features of the real object are captured by the model, the time evolution of the model will mimic with fair accuracy the time

> development of the real object, and so one can predict the most important key aspects which the real object will have in the future .. An absolutely precise simulation of something is called an *emulation*. (pp.206-207).

On this view, material identity is not explicitly stated as a necessary condition for emulation; hence, both simulation and emulation can be defined in terms of a mapping between heterogeneous types. For Tipler, simulation becomes emulation if a behavioural isomorphism can be established between a phenomenon and the model of that phenomenon[20]. The implication is that if the mapping is exact, the model must be functionally isomorphic with the noumenal ground of the phenomenon. Crucially, on this basis, Tipler is led to uphold the "strong" artificiality thesis (section 3.4.5).

4.3.4. Simulation and Realization

It is important to be absolutely clear about the distinction between the simulation of a phenomenon and the duplication or *realization* of that phenomenon. For example, Hofstadter (1981a) maintains that artificial (simulated) hurricanes and natural hurricanes are instances of a common phenomenal class, viz. the class of hurricanes. However, this position necessitates the situatedness (or *endophysicality*) of artificial observers within computer simulations. On Hofstadter's view,

> if the program were incredibly detailed, it could include simulated people on the ground who would experience the wind and the rain just as we do when a hurricane hits. In their minds - or, if you prefer, in their *simulated* minds - the hurricane would not be a simulation but a genuine phenomenon complete with drenching and devastation. (p.74)

This view has been contested by Searle (1980,1992) and Sober (1991) who maintain that artificial analogues of natural phenomena and the natural phenomena themselves are ontically distinct: The former are non-causal *as artificial analogues of natural phenomena* whereas the latter are causal *as the natural phenomena themselves*. As Sober states,

> it is sometimes suggested that, when a computer simulation is detailed enough, it then becomes possible to say that a computer is an instance of the objects and processes that it simulates. A computer simulation of a bridge can be treated as a bridge, when there are simulated people on it and a simulated river flowing underneath .. The problem with computer simulations is not that they are simplified representations, but that they *are* representations. Even a complete description of a bridge - one faithful in every detail - would still be a very different object from a real bridge [emphasis added]. (p.764)

However, this argument can be contested on the grounds that it assumes *a priori* that a

---

[20] Tipler implicitly adopts the Newtonian modelling relation described in chapter 2.

particular observational perspective, viz. extrinsic or *exosystemic*[21] observation, is the *only* possible perspective that can be adopted. However, this assumption is untenable under naturalism since on such a view, human beings (and possibly[22] other entities) are observers that are intrinsic - or *endosystemic*[23] - to the world: A computer simulation is *only* a simulation to an observer who is able to transcend the simulation and attain a `God's Eye' view of the system, thereby allowing representation to be distinguished from reality; for observers *within* the simulation, representation *is* reality. However, it is unclear whether intrinsic observation *can* occur in an artificial, more specifically, computational, world. Rasmussen (1991b) has stated this as a *necessary* condition (postulate 4) within his metaphysical framework for artificial reality:

> One of the criteria for a process to be alive involves adaptive organism-environment responses. This implies that even the simplest living object, for example, a hypothetical process implemented on a computer, must have a primitive notion of itself and its surrounding environment. Such responses imply the existence of an internal model of the world. The living object *perceives* a reality .. Reality can, thereby, acquire its meaning through a *conscious* conception of the world, via an organization of the information we get from our senses [emphasis added]. (p.769)

Tipler (1994) argues for a similar position, basing his view on the anthropic principle[24], viz. "the simulations which are sufficiently complex to contain observers - thinking, feeling beings - as subsimulations exist physically." (p.210) However, Rasmussen (and Tipler) appears to endorse two conflicting positions: On the one hand, a commitment to Gibson's (1966) idea that perception is not grounded in conscious sensation (that is, secondary qualities or *qualia*) but on the detection of information (implicitly viewed either *in terms of* primary qualities or *as a* primary quality itself); on the other hand, a commitment to the view that meaning (semantics) is acquired via a "*conscious* conception of the world", implying a link between perception and consciousness and thus, between perception and sensation. However, this is, in fact, not the case since Rasmussen holds that consciousness can be defined in terms of the *organization* of sensory information, implying, thereby, a functionalist view of consciousness[25]. In this

---

21 That is, an observational perspective *external* to the system.

22 Only *possibly* since it is unclear whether an observation counts *as* an observation in the absence of the possibility for appreciating this fact (chapter 6).

23 That is, an observational perspective *internal* to the system.

24 Gale (1986) defines the *Anthropic Principle* briefly as "a causal link between the existence of intelligent observers and the properties of the universe which they observe." (p.104) A more detailed interpretation of the concept is presented in chapter 6.

25 In chapter 7, it will be shown that functionalist accounts of consciousness - like all materialist accounts (Searle,92) - fail to address the `hard' problem of consciousness, viz. how ontological subjectivity (that is, what-it-is-likeness) can emerge from an ontologically-objective substrate. This point is of critical significance since

connection, it is significant to note that Rasmussen appeals to a notion of modelling similar to that of Crutchfield (1994) which has been described by the latter as "somewhat different from the standard conception" (chapter 3). However, this notion of modelling is problematic since it leads back to the problem of observation: In short, is it meaningful to speak about the construction of models in the absence of an agency capable of consciously *interpreting* such constructions *as models* ? Additionally, there is a need to consider the issue of whether it is *necessary* that an organism construct an internal representation or *model* of its environment in order to respond adaptively in a homeostatic or evolutionary sense. For example, Maturana and Varela (1980) present a framework for organism-environment relations in which explicit internal models are absent; the `model' is implicit in the dynamical coupling relation between the systemic structures of organism and environment (chapters 3 and 6). Such *autopoietic* (or organizationally-homeostatic) systems are held to be ontically non-representational; however, artificial analogues of such systems realized in computational media (substrates) *are* representational at the most primitive level of their being. Hence, artificial analogues of autopoietic systems *are* models in the conventional sense, and can be interpreted as such by their creators[26]. Notwithstanding these problems, however, Hofstadter (1981a) maintains that the confusion surrounding the simulation-realization debate is epistemological, a consequence of confusing levels of description. On his view,

> the laws of physics don't get torn apart by real hurricanes [just as no computer ever gets torn apart in the process of simulating winds]. In the case of the simulated hurricane, if you go peering at the computer's memory expecting to find broken wires and so forth, you'll be disappointed. But look at the proper level. Look into the *structures* that are coded for in the memory. You'll see that some abstract links have been broken, some values of variables radically changed, and so forth. There's your flood, your devastation - real, only a little concealed, a little hard to detect .. You recognize a hurricane by its *effects*. You have no way of going in and finding some ethereal `essence of hurricane', some `hurricane soul', located right in the middle of the eye! It's the existence of a certain kind of *pattern* - a spiral storm with an eye and so forth that makes you say it's a hurricane. (pp.74-75)

The problem with this argument is that it assumes that syntax (structure) is intrinsic to physics. However, this computational-realist position has been contested by Searle (1992), Tallis (1994) and Lanier (1995b) who maintain that the stucture of computer hardware and software reflects intentional human concerns and is, thereby, *ontologically* parasitic on human intentionality. Furthermore, and relatedly, how is an effect recognized *as an effect*? For an effect to be detectable, it is necessary that some difference exist between states of the world prior to and posterior to causation (chapter 3). However, what is to count as a difference has two aspects, viz. ontological and

---

it undermines the possibility of *conscious* intrinsic or endosystemic observation and hence, of "strong" computationally emergent artificiality (chapter 5). This is because the latter *as* a "strong" assertion must be capable of artifactually-instantiating analogues of *all* natural phenomena and since it cannot instantiate consciousness, it is incomplete.

26 As shown in chapter 6, such systems are, at the most primitive level of their being, *allopoietic* or artifactual.

epistemological: Ontically, on a realist-objectivist view it is necessarily the case that for an effect to be counted as an effect, it must make a difference irrespective of whether or not the effect is observed; from an epistemological perspective, by contrast, it is necessarily the case that an observation is made and that observation is motivated by intentional (although not necessarily conscious) concerns, viz. selection of what is to count as of interest in the act of observation. On the latter view, effects are *observationally-relativistic*. A simple demonstration of the correctness of this assertion is given by considering what happens when an electric light is switched on in a room: If one is interested merely in the *function* of illumination, then the effect of pressing the light switch is that the room becomes illuminated[27]. Thus, the effect of the cause (pressing the switch) is that certain previously invisible objects in the room become visible; heating and other `side effects' of the light are usually ignored. While it might be argued that this is simply a case of epistemological incompleteness, viz. all effects are objectively present yet undisclosed in the specific observational act, it is precisely this point which provides support for observational-relativism; that is, there is always the possibility that there may exist undisclosed properties. This is what leads to the *List Fallacy* for naturals (section 4.4), that is, the problem of defining naturals in terms of a list of properties, since it is impossible to prove that lists are ever complete[28]. As will be shown in chapter 7, the List Fallacy is further compounded by the existence of $objective_1$ phenomena, whereby $objective_1$ is meant real, which are not $objective_2$, whereby $objective_2$ is meant accessible from a third-person (or externalistic) perspective (Nagel,86).

4.3.5. "Strong" and "Weak" Artificiality

The notion of `strength' as applied to artificiality was originally introduced by Searle (1980,1984) in the context of a critique of cognitive science and artificial intelligence (AI): "strong" AI is the *epistemologically-reductive* view that minds are computers; "weak" AI, on the other hand, is the *methodologically-reductive* view in which it is held that minds can be investigated via the use of computers[29]. The strength concept can also be applied to other `sciences of the artificial' such as artificial life (Sober,91) which leads to the following generic formulation: "strong" artificiality implies emulation, duplication or realization; "weak" artificiality implies simulation. "Strong" artificiality is consistent with computationalism (chapter 2); however, on the "weak" view of artificiality, computers are regarded as tools, that is, artifacts in the conventional anthropocentric

---

27 *Ceteris paribus*, that is, all other things being equal (or under normal conditions).

28 As will be argued in chapters 6 and 7, the List Fallacy does not arise for artifactuals since their being is specified (made). Thus, epistemological constraints follow ontological constriants (specifically, mode of *poiēsis* or coming-forth).

29 Brief descriptions of epistemological and methodological reductionism are given in chapter 3.

sense of the term. Pattee (1989) claims that many of the controversies associated with computational artificiality, for example, whether an artifactual analogue of a natural phenomenon is an instance of "strong" or "weak" artificiality, arise as a consequence of the multiple uses of the term `computation', viz.

- computer-dependent *realization* of phenomena
- computer *simulation* of phenomenal behaviour
- construction of theories of phenomena *deriving* from simulations
- *testing* of theories of phenomena via simulations

This position is supported by Haken et al. (1993) who, in the context of a discussion of the link between computers and metaphor, maintain that

> when we consider the machine as a metaphor and more specifically the computer as a metaphor for human thinking we find ourselves in an ambivalent situation. At the same time the computer is a tool for and the subject of thinking. (p.4)

The possibility of viewing the computer both as the `root metaphor' (chapter 1) of computationalism (chapter 2) and as a tool or artifact gives rise to two sets of oppositions involving the concept of artificiality, viz. (i) artificiality as appearance *vs.* reality and (ii) artificiality as artifactuality *vs.* naturality respectively. If (ii) can be displaced by (i) then "strong" artificiality - more specifically, "strong" computationally emergent artificiality (chapter 5) - is possible since on this view natural reals and artifactual reals are possible; that is, artifactuality is only *contingently*-related to appearance and naturality to reality[30]. However, if, as is argued in chapter 6, (ii) is grounding relative to (i), "strong" artificiality is impossible. In the following sections, "strong" artificiality is assumed to be possible and the means by which it can be realized are investigated, commencing with a discussion as to how the appearance-reality - or *as-if* (simulation) *vs. as-is* (realization) - question can be decided.

4.3.6. The Turing Test

The Turing Test was developed by Alan Turing (1950) in response to the question of whether or not a machine could be capable of thought. Turing argued that the original formulation of the question was problematic since terms such as `machine' and `thought' are ambiguous. Consequently, it was proposed that the question be replaced by a variant of what he referred to as the "imitation game":

> It is played with three people, a man (*A*), a woman (*B*), and an interrogator (*C*) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the two is the man and which is the woman. He knows them by labels *X* and *Y*, and at the end of the game he says either `*X* is *A* and *Y* is *B*' or `*X* is *B* and *Y* is *A*'.(p.40)

---

30 In short, the appearance-reality distinction can be overcome on this view because distinction in *poiēsis* (coming-forth, becoming) is inessential to the Being of phenomena.

The object of the game for *A* or/and *B* is to render *C* incapable of making a correct identification. Furthermore, *C* is only allowed to interrogate *A* and *B* in a way which does not immediately decide the issue. For example, in the original version of the game, *C* would not be allowed to see *A* or *B* since this would almost certainly make it possible to determine who was the man and who the woman. In Turing's variant, *A* is replaced by a computer and *C* is only allowed to communicate with *A* and *B* via typewritten answers; this follows from his associating intelligence with linguistic capability. If *A* cannot be identified as the computer, it is said to have `passed' the Turing Test. According to Turing, passing the test is a sufficient (but not necessary) condition for intelligence. Various objections to the test based on theological, ethical, logical, experiential, epistemological and even paranormal grounds were considered and refutations attempted by Turing[31].

Although the Turing Test is held to be scientific since it meets Popper's criteria of falsifiability and generates results which are repeatable and objective, it has been the subject of endless criticism within the philosophy of mind. This is a consequence of its being essentially *operationalistic* or/and *behaviouristic* in nature. Behaviourism and operationalism are philosophical positions in which mental concepts such as "thinking" are defined in terms of overt (or objective) behaviour or dispositions to behave, thereby providing a deductive basis upon which claims for the existence of mental phenomena can be made. The problem with these positions is that thinking is, at least in part, clearly an internal (subjective) activity (Searle,80) (Searle,92); consequently, behavioural or operational analyses of a mental concept will, of necessity, be incomplete, thereby giving rise to the possibility of what Sober (1991) has called type-1 errors, viz. the passing of the test by computers which do not think. However, other interpretations of the test are possible: For example, the *inductive* interpretation due to Moor (1976). On this position, it is not necessary for additional evidence to be gathered before a justified inductive inference (as to the presence of thought in a machine) can be made. However, there is then the problem of induction as identified by Hume, viz. on what basis is an inference to count as `justified' (Chalmers,82) ? A third possibility is to view the Turing Test as involving *abductive* (or heuristic) reasoning on the basis of overt behaviour, viz.

If an object *O* has property *P* it will generate behaviour *S*.
X generates behaviour *S*.
Therefore, it is *likely* that *X* has property *P*.

Other criticisms of the test have been based on assertions that it is too easy, too narrow or too shallow (Moor,92). Additionally, it could be argued that the Turing Test merely measures the ability of a candidate system to deceive its interrogator (Shieber,94) and that the test is, therefore, flawed since intelligence intrinsically involves truthfulness

31 It is interesting to note that Turing held the argument from extrasensory perception to be the only serious grounds for objecting to the validity of his test. However, the extent to which any of his refutations of the (other) arguments can be considered final and binding is highly contestable.

(Epstein,96). However, this argument assumes a *realist* position with respect to truth claims; on a *pragmatist* or evolutionary view (Dennett,95), truth is merely identical to what confers survival value, and is, therefore, defined with respect to the *contingency* of the environment. On such a view, there are no *necessary* truths[32] and it is quite possible - even probable - that, at least in some environments (places and times), truth *is* deception. Perhaps the strongest criticism of the Turing Test, however, is that made by Lanier (1995a)[33]. On his view,

> the problem with the Turing Test is that it presents a concundrum of scientific method. We presume that improvement to machines takes place, so there is a starting state in our experiments where the humanis considered `smarter', whatever that means, than the computer. We are measuring a change in human ability to discriminate between human and machine behaviour. But the human is defined as the most flexible element in the measurement loop at the start, so how do we know we aren't measuring a state change in the human, rather than in the computer ? Is there an experimental difference, in this setup, between computers getting `smarter' and humans getting `stupider' ? I don't think so. (p.79)

Although originally formulated in the context of intelligence (or mind), the Turing Test is readily applied to other phenomena: For example, Dennett (1978) describes a Turing Test for determining whether a machine has intentionality and Sober (1991) has outlined a Turing Test for artificial life. It is an implicit assumption within computationalism that the Turing Test provides an appropriate means of measuring the extent to which artificialities (or artificial phenomena) are behaviourally isomorphic with corresponding natural phenomena. For this reason, Turing Tests for various artificialities are proposed in the following sections.

4.3.7. Artificiality and Emergence

In chapter 1, possible links between computationalism and artificiality on the one hand, and between computationalism and emergence on the other were briefly introduced. A potential third link between artificiality and emergence was not explicitly considered; however, such a connection follows naturally if artificiality is identified with simulation since a relation exists between simulation and emergence in the context of non-linear dynamical systems. As Rasmussen et al. (1995) state,

> a simulation is a mechanism which interacts many state transition models of individual subsystems (i.e. system components) and thereby *generates* system dynamical phenomena. (p.2)

---

32 As Nagel (1998) has shown, the strong relativist version of this position, viz. everything is relative, is ultimately self-defeating.

33 A variant of Lanier's "dumbing down" argument gives rise to *three* possibilities for how the Turing Test can be passed: (1) *A* has 'smartened up' to the level of *B* so that *C* is unable to distinguish between them; (2) *B* has `dumbed down' to the level of *A* so that *C* is unable to distinguish between them; (3) *C* has `dumbed down' its *test criteria* such that it is unable to distinguish between *A* and *B*.

The simulation-as-artificial-as-apparent connection is implied by statements such as "a simulation is a *representational* mechanism that is distinguished by its capacity to generate relations that are not explicitly encoded [emphasis added]" (p.8) and the artificiality-emergence link is made explicit by the following assertion: "simulation is a natural method to study emergence." (p.8) However, it should not be inferred that this link only holds for *simulation*, that is, "weak" artificiality since, on computationalism, naturality is itself ontologically-computational in which case the possibility of simulation collapsing onto realization[34] is not precluded.

4.3.8. Functionalist Unification Reconsidered

In order for artificiality to serve as a unifying concept for the various artificialities (artificial analogues of natural phenomena), it was argued that it must be definable in non-teleological terms (section 4.2.5). On the basis of the investigations into artificiality made in the preceding sections, it should be possible to define artificiality in purely behavioural terms. However, the concept of multiple-realizability must be retained otherwise the artificiality as artifactuality project collapses by definition. Consequently, it is necessary to separate the concept of multiple-realizability from that of functionalism. This is possible because the entailment relation is as follows: Functionalism necessitates multiple-realizability but not vice-versa. This would seem to imply that computationalism has been undermined; however, this is not the case. Computationalism is indeed retained as the ground of artificiality; however, in a form necessarily purged of any teleological connotations. Whether this is meaningful or even possible constitutes the essence of arguments presented in Part II of this study. For the remainder of this chapter and the next, computationalism will be accepted as both ontologically correct and the basis of artificiality.

4.4. Natural and Artificial Phenomena

In the following sections, three kinds of natural phenomena (matter, life, mind) are presented and the corresponding artificial analogues (A-Physics, A-Life, AI) are briefly described. It must be appreciated at the outset that any attempt to provide *definitions* for such phenomena leads to what Kelly (1993) has referred to as the List Fallacy (section 4.3.4), the error of maintaining that a list of properties can provide a necessary unity:

> It is all too easy to adopt *A*, *B*, *C*, ... as the base properties upon which to construct an artifact and then conclude that when these properties are realized the phenomenon partially characterised by the properties is also realised. This conclusion may be justified when the construct is an artificial one defined by the properties, but is unlikely to be true when a `natural' phenomenon is at issue. (pp.65-66)

34 In fact, on the assumption of computationalism, simulation collapses onto *emulation* since both simulat*ion* and simulat*ed* are ontologically-computational.

The *definitional approach* is grounded in (at least) two principles: (i) metaphysical realism[35] and (ii) a substance-attribute (subject-predicate, object-property) ontology. In conjunction with (i) and (ii), the set of properties associated with an object is held to be *finite* and *knowable* (that is, epistemically-tractable). While appreciating the philosophical problems associated with the definitional approach (for example, the impossibility of confirming that a list of properties is exhaustive of the phenomenon being investigated), it might be argued that its use is necessary if discourse of any kind - scientific, philosophical or otherwise - is to be possible. This is a consequence of the existential fact that the world must be `cut' (chapters 1 and 6) in order to be dealt with; the alternative is to adopt a position based on Wittgenstein's concluding remark in the *Tractatus*, viz. "that of which we cannot speak, we must pass over in silence", and refrain from discourse completely[36]. It might be objected that this latter approach is only possible with respect to the `cutting' associated with *linguistic* discourse; human existence necessitates some kind of `cutting' of the world since human beings are embodied beings-in-the-world and the *body* `cuts' the world during its interaction with the latter. However, if Heidegger is right about the primacy of unreflective coping (chapter 1), then the body-world duality is a derivative way (or mode) of Being capable of being recognized as such only once it has emerged from the primordial condition of being-in-the-world. On this view, `cutting' is a necessarily Cartesian activity; beings-in-the-world do not `cut' the world (chapter 6).

Consequently, the definitional approach will be adopted in this and the following chapter subject to the caveat that the definitions presented might possibly be incomplete or even incorrect. It should also be understood at the outset that the definition problem has different implications with respect to naturality and artificiality: While it might be impossible to provide complete definitions of natural phenomena, it is possible to provide complete definitions for artificialities, particularly *if* the latter are interpreted as simulations of the former. This follows from the fact that simulations *are* complete in an *a priori* sense because they present partial views of phenomena; a candidate realization, that is, an emulation of a natural phenomenon, *may* be incomplete because

---

35 Naive realism is the metaphysical position in which an objective reality is held to exist independently of *observation* by humans and/or any other beings capable of observational acts. Stated simply, on this view, reality is held to be `out there'. Searle (1992, 1995), while defending realism, has been led to modify its basic form somewhat, viz. *external realism*, the view that although there is a reality independent of *representation*, this reality is not completely ontologically-objective in the third-person (externalistic, non-experiential) sense; this is because "some mental states, such as pains, are ontologically-subjective [that is, first-person, experiential, internalistic], but they are not representations. They are representation independent but not mind independent." (p.152)

36 This argument holds irrespective of whether an *essentialist* or *non-essentialist* position is adopted. In the former, class (or category) membership is defined in terms of *necessary* properties; in the latter, class membership is specified in terms of *sufficient* properties, that is, Wittgensteinian family resemblances (Lakoff,87). However, both positions necessarily involve `cutting' (category production) on the basis of some criteria (necessary or sufficient).

it is an open question whether or not the set of properties associated with the phenomenon can or has been listed in its entirety. (As stated earlier, this necessitates assuming metaphysical realism and a substance-attribute ontology.) In an attempt to circumvent problems associated with terminological definition, Fetzer (1990) advocates the adoption of what might be termed artificiality *criteria* as contrasted with artificiality *definitions*, "where the criterion functions as a (usually reliable, but not therefore infallible) evidential indicator for deciding, in a given case, whether or not that case is an instance of this property" (p.5). This is consistent with the abductive interpretation of the Turing Test described in section 4.3.6 and is the approach adopted herein.

In the following sections, various *criteria* for recognizing material (or physical), vital (or biological) and mental (or cognitive) phenomena are presented based on the assumption of a computationalist ontology, viz. computationalism (chapters 2 and 5). Detailed investigations of each phenomenon are, however, beyond the aim and scope of this study; for this reason, recourse has generally been made to presentations in the non-technical literature. In defending this approach against charges of excessive naivety, it is necessary (and perhaps also timely) to restate the thesis objective, viz.

> *To establish that computationalism does not provide a sufficient metaphysical basis for a unifying framework of "strong" emergent artificiality.*

What *is* being investigated is the possibility of phenomenal unification based on a computationalist ontology; what is *not* being attempted is contribution to the knowledge base associated with a particular phenomenal domain. (Arguments in favour of pursuing the former have been presented in chapter 1.) Nonetheless, while being aware of the possible problem of constructing arguments directed at `straw men', it is maintained, with Schrödinger (1944), that

> I can see no escape from this dilemma .. that some of us should venture to embark on a synthesis of facts and theories .. at the risk of making fools of ourselves. (p.vii)

4.5. Mind

In the following section, the concept of mind is briefly introduced. The presentation has three objectives: (1) to identify some of the basic features associated with mind and provide brief descriptions of these characteristics; (2) to outline how these characteristics can be interpreted in computational terms; and (3) to examine the possibility of a link between mind and life.

4.5.1. What is *Mind* ?

Describing mind is extremely difficult; *defining mind*, even more so. Irrespective of whether an essentialist or non-essentialist (that is, cluster-based) approach is adopted, problems remain. Penrose, writing in *The Emperor's New Mind* (1990), reluctantly admits that

> we shall be having enough trouble with coming to terms with `consciousness' as it stands, so I hope the reader will forgive me if I leave the problems of `mind' and `soul' essentially alone! (p.525).

Unfortunately, `definitional modesty' becomes problematic when attempting to explore possible connections between "computers, minds and the laws of physics". Can such investigations proceed in the absence of a prior conception of mind ? Arguing that a conception might *emerge* from such investigations does not solve the problem: How can mind be recognized since re-cognition necessitates cognition, that is, prior understanding of that which is being recognized, in this case, mind. Reasoning in this way leads to one of two possibilities: Either (i) mind *is* defined *a priori* or (ii) it is excluded as a subject of investigation. However, if defined, there is always the possibility that the definition will be incomplete (section 4.4). As Putnam (1967) states,

> the hypothesis that any inventory includes a list of all ultimate `building blocks' of causal processes that there are is a synthetic one and cannot be regarded as true by pure logic. (p.92)

What has not been presented is a third possibility, the *hermeneutic* approach (chapter 1), which is, in fact, the approach implicitly adopted in most investigations of phenomena (natural or otherwise): An *a priori* conception (definition) is proposed and subsequently *refined* during the course of investigation. This may result in contraction, expansion or even complete replacement of a concept by another, thereby following the general pattern of change associated with Kuhnian paradigms (chapter 1).

While appreciating the necessity of a definitional approach (in the hermeneutic sense), there is still the problem of how to `break into' the hermeneutic circle (chapter 1) and provide a concept of mind which is likely to prove fruitful, that is, conducive of refinement. In this connection, the following statement due to Harth (1982) is most pertinent:

> *Mind* is a troublesome word. We have inherited it from the Latin *mens*. However, we look in vain for its equivalent in some other languages. In German we have *Seele* (soul), *Geist* (spirit), *Verstand* (intellect), *Vernunft* (reason), *Germut* (disposition), *Gedachtnis* (memory), *Meinung* (opinion), *Absicht* (intent) - but no word has all the shades of meaning of *mind*. In general, we have no difficulty in the use of the word. The context makes clear which of the above or other meanings are intended. But trouble arises when we try to *define* mind or make pronouncements about its relationship with the brain. (p.234)

In order to maximize the possibility of definitional refinement (and thereby provide scope for further investigation), a relatively comprehensive definition of mind must be adopted. The following list due to Edney (1994) is proposed:

- feeling
- emotion
- thinking, cogitation, ratiocination, reasoning

- intelligence
- verbalization
- will, volition, decision
- memory
- imagination, foresight, vision
- intuition

Three notions will be briefly examined in what follows, viz. (i) intelligence, (ii) intentionality, and (iii) consciousness.

4.5.2. Intelligence

Moody (1993) maintains that defining intelligence is just as problematic as defining mind. The *Oxford Companion to Philosophy* (1995) offers the following definitions: "An intelligent *creature* is one capable of coping with the unexpected. An intelligent *person* is one in whom memory and the capacity to grasp relations and to solve problems with speed and originality are especially pronounced." The *Oxford Companion to the Mind* (1987) provides the following generic definition, viz. "the capacity to learn from experience, and adapt to one's environment." Neisser (1979) defines intelligence in non-essentialist terms, that is, with respect to a `prototype'. On his *existentialist* conception, intelligence is a cluster concept or Wittgensteinian family resemblance. His justification for this approach is that

> there are no definitive criteria of intelligence, just as there are none for chairness; it is a fuzzy-edged concept to which many features are relevant. Two people may both be quite intelligent and yet have very few traits in common - they resemble the prototype along different dimensions .. [Intelligence] is a resemblance betwen two individuals, one real and the other prototypical. (p.185)

However, there are (at least) two problems with this position: First, the possibility (in fact, near certainty) of *multiple* prototypes being proposed, thereby rendering the definition meaningless because of its potential expansion to all individuals; second, it fails to differentiate what people mean by `intelligence' from what `intelligence' means in itself (*an sich*)[37]. The former such definitions are relativistic and folk-psychological as opposed to realistic and objective. (Consequently, choice of approach necessarily involves metaphysical issues.) More conventional approaches involve making use of quantitative metrics such as IQ or `intelligence quotient' and *g* (or `general intelligence') factor analysis (Spearman,27). However, Kamin (1981) has contested the validity of these approaches on the grounds that they are genetically reductive, making little or no allowance for the role of learning and environmental (specifically sociological) factors. Another significant criticism against quantitative approaches involves the argument that such metrics inevitably incorporate cultural bias (Gould,81). Interestingly, Turing (1948), an early advocate of machine or artificial intelligence (section 4.5.9), maintained that "the idea of `intelligence' is itself emotional rather than mathematical" (p.2). This

---

37 Assuming, of course, that an *objective* (that is, observer-independent) definition of intelligence exists.

would seem to undermine the prospect of formulating an objective (universal) concept of intelligence and at the same time provide support for Neisser's position. This subjectivist approach finds expression in Fetzer's (1990) semiotic view of intelligence in which the latter is defined in terms of rationality and associated with notions such as ends, action, beliefs, and humanity. While it might be argued that this is a somewhat anthropocentric definition, it is useful since it makes possible the establishment of a connection between intelligence and intentionality.

4.5.3. Intentionality

The notion of intentionality was introduced in chapter 1 in connection with a discussion of Husserlian phenomenology. In summary, an anlysis of consciousness reveals it always as consciousness of something, that is, there is an `aboutness' associated with consciousness by which the mind is directed towards objects under some aspect. Thoughts, beliefs, desires, intentions would all be regarded as instances of intentional phenomena; pains, the experience of redness and other primitive sensations (`qualia') would not, although thoughts about such experiences would be regarded as intentional. The directedness of consciousness is referred to as *intentionality* and, according to Husserl, can be accounted for via reference to an abstract representational structure within consciousness, viz. the *noema*. Using a phenomenological procedure called the transcendental reduction (chapter 1), Husserl was able to reduce the subjective ego (consciousness) to what he referred to as the Transcendental Ego. What is significant about this reduction is that it involves a movement from one ontological category (subjectivity) to another (objectivity). (The Transcendental Ego must be - and *is*, according to Husserl - non-conscious since it is identified as an *objective formal structure*.) The reduction of a subjective Cartesian ego to the objective Transcendental Ego defined in representational (formal) terms provides support for the multiple-realizability thesis since anything capable of instantiating the representational structure of the Ego would be capable of consciousness. Hence, by arguing that consciousness is ultimately representational in character, Husserl laid the foundations for the computational-functionalist or cognitivist approach to the mind (chapter 1).

The significant point in the context of the present discussion is that intelligence seems to imply some form of goal-directedness or teleology (section 4.5.2) and in the context of the mind, this takes a specific form, viz. intentionality. According to the Husserlian view, however, intentionality necessitates consciousness[38].

---

[38] It could be objected that unconscious, subconscious or, more accurately, non-conscious forms of intentionality are conceivable; hence, consciousness may not be a necessary condition for intentionality. This is consistent with Heidegger's position, viz. the non-representational (non-mentalistic) intentionality associated with *Dasein* (being-in-the-world) is primordial relative to the representational (mentalistic) intentionality associated with the Cartesian subject or ego (chapter 1). However, it must not be forgotten that recognition and, moreover, *articulation* of the non-representational mode of human being necessitates a `switch' from *Dasein* to Cartesian ego; even if Heidegger's assertion that `language is the house of Being' is accepted, that is, hermeneutic

4.5.4. Consciousness

Consciousness, like mind and intelligence, is extremely difficult to define. Nagel (1979) offers the following definition, viz.

> an organism has conscious mental states if and only if there is something that it is like to *be* that organism - something it is like *for* the organism. (p.166)

Both Nagel (1979, 1986) and Searle (1984, 1992) maintain that consciousness is an intrinsically and irreducibly subjective or `first-person' *experiential* phenomenon. According to Nagel, you have to *be* the thing to know what it is *like* to be the thing, that is, epistemological identity (of experience) necessitates ontological identity (of existence) and visa-versa. This position has a number of interesting corrolaries. For example,

> the fact that we cannot expect ever to accomodate in our language a detailed description of Martian or bat phenomenology should not lead us to dismiss as meaningless the claim that bats and Martians have experiences fully comparable in richness of detail to our own. (p.170)

Furthermore,

> even if I could by gradual degrees be transformed into a bat, nothing in my present constitution enables me to imagine what the experiences of such a future stage of myself thus metamorphosed would be like. (p.169)

The conceivability of this position leads him to conclude that

> perhaps anything complex enough to behave like a person would have experiences. But that, if true, is a fact which cannot be discovered merely by analyzing the concept of experience. (p.167)

This leads to the famous problem of other minds (section 4.5.5), a modern formulation of which provides the basis for arguing against the Turing Test as a test for mind, or, more specifically, consciousness (section 4.3.6). Before examining the other minds problem, it is worthwhile examining the link between consciousness and intentionality. Penrose (1990), for example, implicitly argues for a `stronger' connection between consciousness and intentionality than that described in section 4.5.3, maintaining that consciousness necessitates intentionality:

> To be conscious, I have to be conscious *of* something. (p.525)

---

discourse is the means by which the structure of Being is articulated, it is an existential fact that this discourse is carried out by human beings in the mode of the Cartesian ego. (Derrida and other deconstructionists might contest this position, maintaining that a text (discourse) is `open' in the sense that interpretation can transcend the meaning originally intended by the author. This is why literary criticism is best described as a *hermeneutic* activity (chapter 1). However, while accepting the validity of this argument, it remains the case that criticism is carried out in the mode of the Cartesian ego. Hence, the above argument stands.)

This position is supported by Goswami (1993) who proposes an idealist interpretation of mind, defining consciousness in terms of four associated concepts, viz. (1) *awareness*, (2) *objects* (thoughts, feelings etc), (3) a *subject* (experiencer), and (4) identification of consciousness as the ground of all being. Penrose goes further to claim that consciousness is a necessary condition for intelligence, the latter a "subsidiary" phenomenon associated with the former. In (Penrose,94), the link between consciousness and intelligence is elaborated as follows:

> my own use of the term `understanding' certainly implies that a genuine possession of this quality would require some element of *awareness* to be present. Without any awareness of what some argument is all about, there can surely be no genuine understanding of that argument .. Awareness is indeed *something*, and this something may be present or absent, at least to a degree. (p.37)

Penrose maintains, therefore, that subjectivity has genuine ontological status, viz. it *is* something. Furthermore, he holds that

> `intelligence' requires `understanding' .. and `understanding' requires `awareness'. (pp.38-39)

On his view, awareness is the *passive* aspect of the phenomenon of consciousness. However, "consciousness has an *active* aspect also, namely the feeling of *free will*." (p.39) The argument for a necessary link between mind and consciousness is implicitly supported by Levin (1979) who in the context of a discussion of artificial intelligence maintains that "the proponent of artificial minds has achieved an empty victory if `mind' is so construed that a being can have a mind without being conscious: he just shifts the philosophical problem to machine consciousness." (p.186) However, Moody (1993) has contested this view, maintaining that non-sentient intelligence is readily conceivable and citing a number of examples involving unconscious problem solving in support of this contention. Moody's position is interesting because it is consistent both with (i) a Heideggerian interpretation of intentionality in terms of `intelligent' situated activity or *coping* (chapter 1) and (ii) a definition of intelligence in biological terms (section 4.5.10).

4.5.5. The Other-Minds Problem

The fact that conscious experience has an irreducibly subjective aspect, viz. "only I can know what it is like to be me", gives rise to an interesting problem: Since I can only ever have access to my own conscious experiences (via reflexive acts of self-consciousness), what grounds do I have for believing that other beings have mental experiences (sensations, thoughts etc) similar to my own or even have experiences at all ? The vast majority of proposed solutions to the other-minds problem can be placed somewhere between the following two extremes: On the one hand, *solipsism*, the view that *my* mind is the *only* mind (or thing) that exists, everything else being, in some sense, a *product* of my mind. This is the position of the Cartesian sceptic who maintains that everything beyond "*cogito ergo sum*" or "I think therefore I am" is dubitable; on the other hand, *panpsychism* (chapter 1), the view that everything in the universe is conscious. This position, which is essentially a modern form of philosophical animism[39], is often advanced on the grounds that since subjective (experiential) phenomena cannot be reduced to objective (non-experiential) phenomena (chapter 7), subjectivity must be a fundamental ontological category (Chalmers,96).

The most popular position, however, involves inductive inference of consciousness or subjective experience on the basis of *behavioural* evidence; specifically, *one* draws a widespread analogy from one's *own* behaviour *and internal mental states* to the internal states of others when their behaviour is similar. On this view, the problem of other-minds reduces to one of determining the set of behavioural criteria which are necessary and/or sufficient for the attribution of consciousness to an entity. However, it is unclear whether the link between behaviour (objective, external) and conscious experience (subjective, internal) is necessary (that is, essential) or merely contingent. Searle (1992) presents three different positions with respect to this issue: (1) $\Box B \wedge E$, (2) $\Diamond E \wedge \sim B$, and (3) $\Diamond B \wedge \sim E$, where *B* denotes behaviour, *E* denotes conscious experience, $\Box$ denotes logical necessity and $\Diamond$ denotes logical possibility. (3) or the *zombie* argument is of particular importance in the context of the artificiality debate: If consciousness is a *necessary* condition for intelligence (Penrose,94), the Turing Test for artificial intelligence must be invalid since, as a *behaviouristic* (that is, ontologically-objective, third-person, externalistic) test, it cannot provide a means by which to confirm the presence or absence of consciousness. It is crucial to note that the other-minds problem follows from the Cartesian separation of subject from object; it does not arise for non-mentalistic *Dasein* or being-in-the-world (chapter 6).

4.5.6. The Mind-Body Problem

The necessity-contingency problem with respect to the link between behaviour and conscious experience leads directly to consideration of the so-called `hard problem' (Chalmers,96) within the philosophy of mind. This is the problem of explaining how an ontologically-objective substrate can give rise to ontological-subjectivity (that is, conscious experience) and is a restatement of the famous *mind-body problem*, viz. the problem of determining the nature of the link between mind and matter. For substance-dualists such as Descartes, the mind-body relation was a relation between two kinds of stuff, *res extensa* (matter or spatially extended stuff) and *res cogitans* (mental stuff). However, substance-dualism is unacceptable on the modern scientific worldview since the latter is based on the assumption of a monistic metaphysics, viz. one kind of stuff

---

39 Pepper (1942) identifies the root metaphor (chapter 1) of animism as man; consequently, in animist thinking, the world is interpreted in anthropomorphic (or human-like) term, consciousness being a defining characteristic of human being.

which is *physical*[40]. Additionally, substance-dualism is held to be problematic because it cannot explain the link between the two kinds of stuff without invoking a `God of the gaps' (Griffin,88). Since Descartes, numerous alternatives to the dualistic position on the mind-body problem have been developed. Stapledon (1939) provides the following graphical summary of four positions (Fig 4.2):

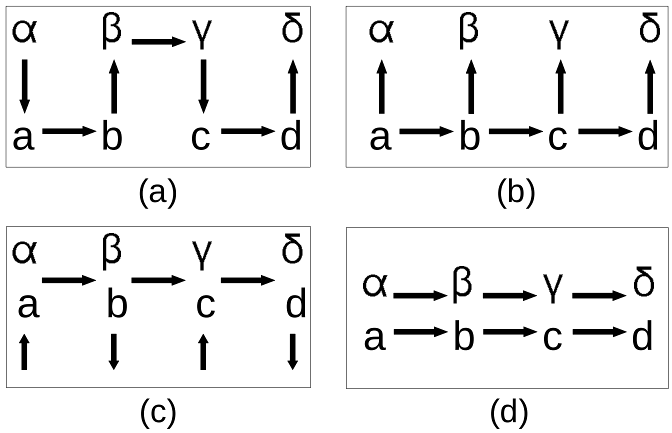

Fig 4.2 Four kinds of mind-body relation. (α, β, γ etc refer to mental events; *a*, *b*, *c* etc refer to material events.)

In Fig 4.2 above, (a) represents Cartesian substance-dualism, (b) epiphenomenalism, (c) dual-aspect theory, and (d) psychophysical parallelism.

Bunge (1977a) has identified ten kinds of behaviour-mentation relation (Table 4.2):

Philosophy of Mind	Explanation of behaviour	Explanation of mentation
Idealism, panpsychism, phenomenalism	Manifestation of the workings of a spirit (individual or worldwide); no precise laws.	Autonomous and spontaneous activity of the mind coverable by law containing only mentalist predicates.

[40] Furthermore, if the physical universe is causally-closed as is maintained, for example, by Chalmers (1996), it becomes impossible to explain non-physical interaction between the physical world and the mental world in a way that does not undermine the principle of the transitivity of causation. However, the argument for causal-closure has been contested from a number of positions including mentalism (Marres,89), panexperientialism (Griffin,98) and radical emergentism (Silberstein,98).

Neutral monism, dual-aspect theory	Manifestation of a non-mental, non-physical being; explainable with a single set of laws translatable into mentalist and physicalist terms.	Manifestation of a non-mental, non-physical being; explainable with a single set of laws translatable into mentalist and physicalist terms.
Eliminative materialism, behaviourism	Outcome of stimuli, hence describable by S-R laws (no intervention of CNS).	Mentation non-existent, hence not to be explained.
Reductive materialism	Motor outcome of physical CNS events, hence explainable in physical terms.	Physical activity of the CNS.
Emergent materialism	Motor outcome of biological CNS events, explainable with the help of biological laws, some of which contain new predicates.	Biological activity of plastic sub-systems of CNS, explainable with the help of biological laws containing new predicates.
Mutual independence of mind and body	Biological events explainable in purely physiological terms plus possibly theological ones.	Mental events explainable in purely mentalist terms plus possibly theological ones.
Psychophysical parallelism, pre-established harmony		
Epiphenomenalism	Motor outcome of CNS events.	Non-motor effect of CNS activity.
Animism	Motor outcome of mental events (e.g. intending and wishing).	Unexplainable except possibly in supernatural terms.
Interactionism	Under dual control of body and mind; only partially explainable.	Autonomous though influenced by bodily events; unexplainable by science.

Table 4.2. Various proposed solutions to the Mind-Body Problem.

Computational-functionalism (computationalism) is absent from both the above schemes. According to Boden (1987), this is because "the `new' concept of machine provided by AI, viz. the hardware-software dichotomy, largely resolves the mind-body problem". (p.4) However, it is important to examine the validity of this claim.

4.5.7. A Functionalist Conception of Mind

Sober (1991) holds that "functionalism in the philosophy of the mind is best seen as an empiricial thesis about the degree to which the psychological characteristics of a system constrain the system's physical realization." (p.754). Functionalists view mind in terms of a set of essential connections between beliefs, desires, memories, and other mental states. Mental states can be broadly classified into two kinds: (i) *experiential* states involving sensory qualities and (ii) *intentional* states involving propositional attitudes. Lacey (1988) defines `strong' functionalism as the thesis that "mental states must themselves be analyzable in such a way that eventually all mental or psychological terms are eliminated; if they cannot all be eliminated we have only weak functionalism." (p.397) Lacey further maintains that "only strong functionalism would have any hope of explaining how mind might *emerge* in a material universe." (p.397) Various functionalist theories of mind have been proposed in the literature, the majority of which are framed in computational terms: For example, Putnam (1960, 1967) defines mind in terms of programs for Turing machines, thereby supporting the hardware is to software as brain is to mind analogy presented in chapter 2. Additional support for a functionalist conception of mind is provided by interpreting thought in symbolic terms. According to Hillis (1988), "it seems likely that symbolic thought can be fruitfully studied and perhaps even recreated without worrying about the details of the emergent system that supports it." (p.180) Hence,

> [one approach to the mind is to] build a model of the emergent substrate of intelligence. This artificial substrate for thought would not need to mimic in detail the mechanisms of the biological system, but it would need to exhibit those emergent properties that are *necessary* to support the operations of thought [italics added]. (p.180)

This emergentist approach (chapter 3) assumes functionalism since it makes a claim for the multiple-realizability of mind. However, a functionalist theory of mind is not limited to a theory of intelligence. For example, Rucker (1985) implicitly assumes a functionalist position in defining personhood or `self' in formalistic terms as follows:

> Daily one eats and inhales billions of new atoms, daily one excretes, sheds, and breathes out billions of old ones. Physically, my present body has almost nothing in common with the body I had twenty years ago. Since I feel that I am still the same person, it must be that `I' am something other than the collection of atoms making up my body. `I' am not so much my atoms as I am the *pattern* in which my atoms are arranged. Some of the atom patterns in my brain code up certain memories; it is the continuity of these memories that gives me my sense of personal identity. (p.146)

Additionally, Hofstadter (1981a) maintains that "emotions are an automatic by-product of the ability to think" (p.81) and that

> the ability to think, feel, and consciousness are just different facets of one phenomenon, and no one of them can be present without the others .. consciousness has got to come from a precise pattern of organization [and] requires a certain way of mirroring the external universe internally, and the ability to respond to that external reality on the basis of the internally represented model .. what's really crucial for a conscious machine is that it should incorporate a well-developed and flexible self-model. (pp.81-82)

His commitment to "strong" functionalism leads him to an emergentist position with respect to the mind-body problem; more specifically, to a computational-emergentist solution of the `hard problem' of conscious experience (section 4.5.6), viz.

> eventually, when you put enough feelingless calculations together in a huge coordinated organization, you'll get something that has properties on another level. You can see it - in fact, you *have* to see it - not as a bunch of little calculations, but as a system of tendencies and desires and beliefs and so on. (p.84)

There are a number of problems with this position which will be examined in chapter 7 when the category problem, that is, the problem of explaining how ontological subjectivity can emerge from an ontologically-objective substrate, is addressed[41]. For the remainder of this section, it will be assumed that Hofstadter's (1979) claim to the effect that "all brain processes are derived from a computational substrate" (p.561) is valid. Acceptance of this form of functionalism leads directly to the computational theory of mind.

4.5.8. The Computational Theory of Mind (CTMi)

The computational theory of mind (CTMi), like computational theories of life (section 4.6.2), matter (section 4.7.6) and other phenomena, is based on the assumption that the ontological version of the Church-Turing Thesis (CTT) and the general form of the Physical Symbol System Hypothesis (PSSH) are both valid (chapter 2). The ontological CTT states that *all* processes[42] (natural or artificial) are effective (or mechanical) procedures and can be implemented by running suitable programs (Turing machine specifications) on a Universal Turing machine. The general PSSH states that a physical symbol system (PSS) provides the *necessary* and *sufficient* conditions for realizing any phenomenon, natural or artificial (chapter 2).

Newell (1980) presents the following list of essential (necessary) properties of mind:

- Behave as an (almost) arbitrary function of the environment
- Operate in real-time
- Exhibit rational, that is, effective adapative behaviour
- Use vast amounts of knowledge about the environment
- Behave robustly in the face of error, the unexpected, and the unknown
- Use symbols (and abstractions)
- Use (natural) language
- Exhibit self-awareness and a sense of self
- Learn from its environment
- Acquire its capabilities through development

---

41 Clearly, the category problem - as defined herein - is *almost* identical to the `hard' or mind-body problem (section 4.5.6), the distinction being that proposed solutions to the category problem must be *emergentist* in nature. This is because this study is concerned with evaluating the sufficiency of computationalism as a metaphysical basis for "strong" *emergent* artificiality.

42 Substances (that is, objects) are, on this view, processually-constituted (chapter 2).

- Arise through evolution
- Be realizable within the brain as a physical system
- Be realizable as a physical system

He maintains that "the notion of general intelligence can only be informally circumscribed, since it refers to an empirical phenomenon." (p.171) In an earlier paper, Newell and Simon (1976) maintain that "symbols lie at the root of intelligent action." (p.114) This leads Lakoff (1987) to identify what he refers to as

> *The Algorithmic Mind position*: Every cognitive process is algorithmic in nature; that is, thought is purely a matter of symbol manipulation (p.339)

According to Newell (1980), "that humans are physical symbol systems implies that there exists a physical architecture that supports that symbol system" (p.174); further, "there must exist a neural organization that is an architecture - i.e. that supports a symbol structure." (p.174) Hence, both mind and brain are defined in terms of PSSs. In (Newell,90), the following are identified as the basis of a *behaving system*:

- knowledge
- representation
- computation
- symbol-manipulation
- architecure
- intelligence
- search and problem spaces
- preparation and deliberation.

Newell asserts that "theories of human cognition are ultimately theories of physical, biological systems" (p.42) and proceeds to define the mind in Darwinian terms as follows:

> I want to take *mind* to be the control system that guides the behaving organism in its complex interactions with the dynamic real world .. The mind then is simply the name we give to the control system that has evolved within the organism to carry out the interactions to the benefit of that organism or, ultimately, for the survival of its species. (p.43)

A variant of the PSSH, qualified by the necessary condition that a PSS be a knowledge-level system, is presented as the basis of a unified theory of cognition. (A *knowledge level system* is one which contains knowledge about the goals the system is to pursue and the means by which such goals can be achieved.) This leads Newell to define intelligence in the following terms:

> intelligence is the ability to bring to bear all the knowledge that one has in the service of one's goals .. Intelligence is [therefore] relative to goals and relative to knowledge. (pp.90-91)

The issue of intentionality in the context of the CTMi is addressed by Dyer (1990) who asserts that,

> the simulation of intentionality is the same as embodying intentionality because the `simulation' of information processing on a computer is information processing. (p.312)

However, the assumption that intentionality is to simulated intentionality as information processing is to simulated information processing is problematic: The `simulation' of information processing *is* information processing because computers are information processors, that is, their ontology (nature or being) is *defined* in terms of their capacity to process information. Hence, simulated and simulator belong to the same ontological category, thereby allowing for emulation (section 4.3.3). However, it is not *necessarily* the case that intentionality (simulated) and information processing (simulator) belong to the same ontological category[43].

With respect to an explanation of consciousness in terms of the CTMi, Hofstadter et al. (1981b) begin by asserting that

> the mere fact that [consciousness] has resisted for so long all attempts to characterize it suggests that our conception of it is at fault. (p.8)

However, this position assumes that subjectivity is reducible to objectivity. McGinn (1987) describes the computationalist view of consciousness as follows:

> The idea .. is that the brain has (in addition to material and functional properties) computational properties; and it is these that `underlie' the presence and operations of consciousness. (p.285)

Exploring links between the ontological CTT, the PSSH and the computationalist view of mind, McGinn (1987) argues that a connection follows almost naturally once it is accepted that computation implies the attribution of propositional content to computational devices and propositional content is usually expressed in symbolic terms. Thus, the central concept underlying the CTT and PSSH is the notion that phenomena are ultimately reducible to symbols or representations. In order for this view to avoid the problem of what it is that the symbols *refer to* (or what it is that the representations are *re-presentations of*), it is necessary to claim ontological self-sufficiency for computationalism: On this view, symbols are self-referential; they do not represent anything or, as Baudrillard (1983) states, they are their own pure simulacrum (chapter 6).

4.5.9. Artificial Intelligence (AI)

In the previous section, the CTMi was outlined; in this section, artificial intelligence (AI)

---

43 Following Searle (1992, 1995) and Tallis (1994), it might be argued that intentionality and information processing belong to separate ontological categories since the former is an *intrinsic*, ontologically-subjective phenomenon while the latter is an *extrinsic*, ontologically-objective phenomenon, that is, an institutional fact (or *artifact*).

is briefly examined. The distinction between the CTMi and AI is essentially *methodological*: The CTMi starts from the *natural* phenomenon of mind and attempts to interpret it in computational terms; AI, on the other hand, starts from the *artificial*[44] phenomenon of computation and attempts to interpret it in mental terms. Both approaches converge under the assumption of functionalism (chapter 1).

Artificial intelligence, like natural intelligence, is notoriously difficult to define. For example, the *Handbook of Artificial Intelligence (Vol.I)* (Barr,81) defines AI as

> the part of computer science concerned with *designing* intelligent computer systems, that is, systems that *exhibit* the characteristics we associate with *intelligence* in human behaviour - understanding language, learning, reasoning, solving problems, and so on [emphasis added].

and the *Encyclopedia of Artificial Intelligence (Vol.I)* (Shapiro,92) as

> a field of science and engineering concerned with the computational understanding of what is commonly called *intelligent* behaviour, and with the *creation* of artifacts that *exhibit* such behaviour [emphasis added].

Three notions are common to the above definitions: (i) intelligence; (ii) creativity or design; and (iii) behaviour. The latter notion can itself be understood in at least three senses as previously stated (section 4.3.3). Perhaps the best indicator of the definitional problem with respect to AI is a paper by Negrotti (1991d) presenting a list of "One Hundred Definitions of AI [or artificial intelligence]" which includes the following: (1) *study* of the corresponding natural phenomenon (for example, "study of intelligence by building and analysing computer programs", "study of apparently or actually intelligent behaviour in machines" etc); (2) *simulation* of the corresponding natural phenomenon (for example, "simulation of human behaviour by a machine", "simulation of purposeful behaviour" etc); (3) *instantiation* of the corresponding natural phenomenon (for example, "cognitive mechanics", "mechanical epistemology?" etc). From the above sample of definitions it can be inferred that AI is both (i) a means by which the phenomenon of intelligence (and more broadly, mind) can be studied, and (ii) either a simulation or instantiation of the phenomenon of intelligence itself; hence, AI can be considered as both a *tool* for and the *phenomenon* of study, corresponding to "weak" and "strong" positions (section 4.3.5) respectively. Boden (1987) defines AI as "the use of programs as tools in the study of intelligent processes, tools that help in the discovery of the thinking-procedures and epistemological structures employed by intelligent creatures." (p.17) AI is explicitly identified as "the study of computer *programs* [emphasis added]." (p.3) Hence, it might be inferred that Boden is committed to `weak' or *soft* AI. However, there is implicit support for the PSSH in the assertion that "intelligence may be defined

---

44 This statement is not intended to imply at this stage in the discussion that "strong" computationalism is invalid. It merely indicates that human beings have come to know computers by constructing them. Hence, computers are human artifacts and their functionality is, therefore, artifactual.

as the ability creatively to manipulate symbols, or process information, given the requirements of the task in hand." (p.17)

Kelly (1993) maintains that

> rather than requiring a theory of intelligence, Artificial Intelligence may be viewed as *building* such a theory. (p.37)

This view is both important and problematic: The former because it identifies AI with synthesis or artifactuality, thereby drawing attention to the *poiētic* (coming-forth, becoming) aspect associated with artificiality (chapter 6); the latter because, like all synthetic approaches, it necessitates the prior existence of a definition of the phenomenon it is attempting to realize (on a `strong' interpretation of AI) while at the same time attempting to construct this very definition. (This is a restatement of the re-cognition problem discussed in section 4.5.1.)

4.5.10. Connections Between Mind and Life

Current strategy in AI has moved away from the top-down *design* of intelligent systems and towards the bottom-up *evolution* of such systems[45]. For example, Fogel et al. (1966) define intelligence in evolutionary terms as

> *the ability of any decision-making entity to achieve a degree of success in seeking a wide variety of goals under a wide range of environments*. (p.2)

A potential link between intelligence and life is established in the assertion that "intelligent behaviour is also exhibited by creatures at much lower levels in the phylogenetic series." (p.3) Evolution plays an important role within naturalistic accounts of intelligence since it provides the basis for an emergentist account of mind, and a possible reduction of mentality to the non-mental. Hence, Scriven (1953) maintains that if intelligence is to be defined and explained, this will have to be done at least partially in biological terms. However, McGinn (1987) contests the validity of this view, arguing that "being biologically alive is not a necessary condition of consciousness, but that it is necessary that a conscious being should behave like a living thing (of a certain sophistication)." (p.283) Poundstone (1986) goes a stage further in speculating about the possibility of artificial intelligence (AI) without artificial life (A-Life) in a computational context. In response, Scriven (1953) maintains that

> there is an essential connection between the capacity for complex behaviour and Consciousness; the one is a necessary condition of the other. But it is not a sufficient condition; and though we may decide that living things are Conscious from their behaviour, we cannot decide if everything is

45 It is interesting to note that such a move derives support from Simon's (1969) establishing a link between intelligence and evolution as the basis of the artificiality concept (section 4.2.1).

> Conscious from its behaviour. *Life is itself a necessary condition of Consciousness*, and though behaviour is a factor which sometimes decides the question whether a certain system is alive, it is again not the only one [emphasis added]. (p.34)

Newell et al. (1976) hold that "there is no `intelligence principle', just as there is no `vital principle' that conveys by its very nature the essence of life." (p.115) Vitalism is the view that there exists some kind of special substance, principle or `force', for example, Bergson's *élan vital* or Dreisch's *entelechia*, which is responsible for life. The inherent dualism in vitalistic accounts means that emergence in an ontologically reductive sense (chapter 3) is rejected. The vitalist holds that the matter-life relation is *discrete*; non-living physico-chemical matter cannot give rise to living biological entities. The ontologically-reductive emergentist, on the other hand, maintains that life is *continuous* with matter, reflecting a difference of degree and not of kind. The nature of the difference is often expressed in organizational or informational terms. For example, Stonier (1992) maintains that "just as *information* is a basic, physical property of the universe, so is *intelligence* a product of the evolution of information systems." (p.15) His concept of intelligence is important since it supports a continuum view between intelligence and vitality (life):

> Intelligent activity, for the most part, involves an ability of a system to analyse its environment, and then to make an intelligent response. An intelligent response may result in one of three states:
>
> 1. The system has enhanced its own *survivability*.
> 2. The system has enhanced its own *reproducibility*.
> 3. If the system is goal-oriented, it has enhanced the *achievement* of that goal. (p.15)

However, Adler (1990) has contested the continuum thesis on the following grounds:

> In the life of all other animals, mind is embodied completely. Mind is found entirely imbedded in physical organs. Mind is *in* matter. Only in man does mind rise *above* matter or *over* matter, by virtue of man's having a mind that has intellectual as well as sensitive powers, conceptual as well as perceptual thought, the power to think about what is unperceived and totally imperceptible. (pp.5-6)

A variant of this argument is briefly examined in chapter 6 in connection with Heidegger's intentionalistic account of the existential *as*-structure, that is, the human being's capacity to understand and appreciate a phenomenon *as* the phenomenon that it is[46]. However, in what follows, it will be assumed that the continuum thesis is valid.

4.6. Life

In this section, the concept of life is discussed. The presentation has three objectives: (1) identify the basic *properties* of life (assuming an essentialist position) without providing

---

46 Searle (1995) presents an alternative account of this structure, viz. *X* counts as *Y* in context *C*, from a neo-Husserlian perspective.

a detailed technical presentation of these characteristics[47]; (2) outline how these characteristics can be *interpreted* in computational terms; (3) examine the possibility of a *reduction* of life to matter. In addition, the relationship between the `natural' phenomenon of life and its artifical analogue, artificial life (AL or A-Life) is investigated. The concepts in this section are discussed in much greater detail than those associated with the phenomena of mind (section 4.5) and matter (section 4.7). This has been necessary for two reasons: First, life is assumed, on the continuum view, to be the `bridge' between matter and mind. Hence, it is likely that an investigation of life will contribute to an understanding of the other two phenomena; second, life is perhaps *the* paradigmatic example of an emergent phenomenon and hence, an examination of this concept is essential in order to evaluate the unifying framework of computationally emergent artificiality presented in chapter 5[48].

4.6.1. What is *Life* ?

According to *The Oxford Companion to Philosophy* (1995),

> this, the distinguishing feature of organisms, is best thought of as involving some kind of complex organization, giving an ability to use energy sources for self-maintenance and reproduction. Efforts to find some distinctive substance characterizing life have proven as futile as they have been heroic. The one thing which is clear is that any analysis of life must accept and appreciate that there will be many borderline instances, like viruses. Inconvenient as this may be for the lexicographer, this is precisely what evolutionary theory would lead us to expect.

A survey of various dictionaries leads to the following three definitional categories:

1. Life as an *organization* distinct from inorganic matter (with an associated list of properties)
2. Life as a certain kind of *animated behaviour*
3. Life as a special, incommensurable, quality - *vitalism*

Vitalism (section 4.5.10) is, according to mainstream biological thought, unacceptable since it posits a dualistic conception of reality, thereby conflicting with ontological monism, the ideal of science. Moreover, scientific advances in fields such as biochemistry, molecular biology, and perhaps most importantly, molecular genetics, have led to almost complete rejection of the idea of a vitalistic `ghost in the machine'.

47 Technical details are *purposely* ignored for two reasons: (i) *methodological* - simplicity of analysis, and (ii) *ontological* - abstraction is a necessary condition for multiple-realizability and functionalism.

48 As will be seen in chapter 7, this claim is highly questionable since the category problem, viz how ontological subjectivity *emerges* from an ontologically-objective substrate, constitutes a far more radical instance of emergence given an ontologically-objective interpretation of life. This is because in the latter case, the emergence of life from matter does not appear to involve *ontological* category creation.

According to the vast majority of biologists, Descartes got it *almost* right[49]; living things *are* machines. For example, Mayr (1982) rejects substance-dualism on the grounds that "the concept of consciousness cannot even approximately be defined and therefore detailed discussion is impossible." He further maintains that

> as far as the words `life' and `mind' are concerned, they merely refer to reifications of activities and have no separate existence as entities .. The avoidance of nouns that are nothing but reifications of processes greatly facilitates the analysis of the phenomena that are characteristic for biology. (pp.74-75)

In place of the substance-dualism associated with vitalism, Mayr proposes a *processual* conception (chapter 2) of life:

> Attempts have been made again and again to define `life'. These endeavors are rather futile since it is now clear that there is no special substance, object, or force that can be identified with life. *The process of living, however, can be defined*. There is no doubt that living organisms possess certain attributes that are not or *not in the same manner* found in inanimate objects. (p.53)

Mayr goes on to present the following list of characteristics by which living organisms differ from inanimate matter, viz.

- Complexity and Organization
- Chemical Uniqueness
- Quality
- Uniqueness and Variability
- Possession of a Genetic Program
- Historical Nature
- Natural Selection
- Indeterminacy

Sagan (1985) presents the following five definitions of life: (pp.985-986)

- *Physiological*: any system capable of performing a number of functions such as eating, metabolizing, excreting, breathing, moving, growing, reproducing, self-repair, responding to external stimuli.

- *Metabolic*: any system with a definite boundary, continually exchanging some of its materials with its surroundings, but without altering its general properties, at least over some period of time.

- *Biochemical*: any system that contains reproducible hereditary information coded in nucleic acid molecules and that metabolize by controlling the rate of chemical reactions using enzymes.

- *Genetic*: any system capable of evolution by natural selection.

- *Thermodynamic*: any system which is `open' in the sense of exchanging light, heat, matter etc with its surroundings (or `environment'). The Second Law of Thermodynamics states that in `closed' systems, non processes can occur which increase the net order of the system. Living systems are

49 His mistake was to assert a substantialist yet non-physicalist conception of mind.

> localized regions within the universe (a closed system) where there is a continuous increase in order (at the expense of an increase in disorder of the rest of the universe.)

Although there appears to be no single property by which something may be classified as living, according to Farmer et al. (1991), a list of generic properties is likely to include the following:

- Life is a complex pattern in space-time
- Self-reproduction
- Information storage of a self-representation
- Possession of a metabolism
- Functional interactions with the environment
- Interdependence of parts
- Stability under environmental perturbation (robustness)
- Ability to evolve
- Growth/expansion

Self-reproduction and information storage of a self-representation feature in a number of lists associated with information-theoretic definitions, for example, those of von Neumann (1966) and Orgel (1973). (In chapter 5, self-reproduction is taken to be the *defining* characteristic of life.)

Mayr (1982) adds the following to the above list:

- Emergence of new and unpredictable qualities at hierarchical levels

while Emmeche (1993) argues in favour of including

- Autonomy (with respect to human beings)

The identification of life with autonomy is supported by Polanyi (1962), viz.

> instances of morphological types and of operational principles *subordinated to a centre of individuality* .. [Furthermore,] no types, no operating principles and no individualities can ever be *defined* in terms of physics and chemistry [emphasis added]. (p.383)

Consistent with his interpretation of the concept of emergence (chapter 3), Polanyi (1968) expands upon the above statement as follows:

> if the structure of living things is a set of boundary conditions, this structure is extraneous to the laws of physics and chemistry which the organism is harnessing. Thus the morphology of living things transcends the laws of physics and chemistry. (p.1309)

Emmeche argues for the identification of autonomy as the defining feature of life on the grounds that "this criterion reflects [1] the *evolutionary* fact that life is not a predesigned but a naturally evolved phenomenon, and [2] the *ecological* fact that life is usually not dependent on us for its existence, so an artificially created organism should be able to go on living a life of its own within a *natural* environment [emphasis added]." (p.561) The latter condition, viz. naturality of environment, is extremely important since it raises the issue of whether or not computational `life-forms' are autonomous. This is because computational life-forms are embedded in `universes' (computers) which are *artifactual*. Challenging Emmeche's first claim, viz. life is an evolutionary phenomenon, in the context of artificial and possibly even *natural* life is much more controversial, involving (1) a local to global shift with respect to questions of ontology and epistemology and (2) consideration of the possible validity of enlightened creationism (chapter 6).

Assuming evolutionary theory as valid, Maynard Smith (1986) defines life in terms of two properties, viz.

1. Possession of a metabolism (biochemical, physiological)
2. Functionality of organismic parts (genetic, evolutionary)

Evolution is defined in terms of three concepts: (1) multiplication, (2) variation, and (3) heredity. Put simply, if variation in a population of entities fulfilling these three conditions differentially affects the capacity for survival and reproduction, then that population will evolve. Emmeche (1992), assuming a structuralist position, criticizes Darwinian theory for being "unable to give any satisfying account of the nature of developmental and evolutionary constraints" (p.467), and on this issue is supported by many of the contributors to (Ho,88). As Emmeche states,

> we cannot by the present theory of biology distinguish between possible and impossible forms of life .. The genetic code, for example .. might have been differently composed. However, *its presumed arbitrarity might not be due to historically frozen accidents and various external and (with respect to the living system) contingent causes*; rather some general biochemical constraints on possible forms of protein synthesis and regulation not yet understood may have acted lawfully in the process of creation of this specific code, disallowing the formation of other code tables [emphasis added]. (p.467)

This view is consistent with that of Kauffman (1995) who argues in favour of augmenting the neo-Darwinian explanation of biology with what might be described as a field-theoretic biology defined in dynamical systems terms. Kauffman's approach supports a computational conception of life, locating it within the larger context of an emergentist (or self-organizing) conception of nature; consequently, it supports the idea of computationally emergent artificiality (chapter 5).

Dennett (1995) maintains that a *necessary* condition for life is the existence of an autonomous metabolism on the basis that "it is a deep if not utterly necessary condition for the sort of complexity that is necessary to fend off the gnawing effects of the Second Law of Thermodynamics. All complex macromolecular structures tend to break down over time, so, unless a system is an *open* system, capable of taking in fresh materials and replenishing itself, it will tend to have a short career." (p.127) He also points to the existence of a more or less definite boundary distinguishing the organism (living entity) from everything else as an additional *necessary* condition; this enables self-preservation to be constrained within finite limits. This latter characteristic, which is linked to the notion of autonomy, is contestable since it assumes that the idea of a biological organism or biological *self* is well-defined. Maturana and Varela (1980) present an alternative formulation based on systems theory (chapter 3) in which the organism-environment distinction is viewed as epistemological.

The thermodynamic conception of life is particularly important in the context of this study because of the link between thermodynamics and information theory on the one hand and the link between information and computation on the other. For this reason, the thermodynamic definition of life will be examined further.

4.6.2. Towards a Computational Theory of Life (CTL)

A precedent for the thermodynamic (entropic) or informational approach to life can be found in Herbert Spencer's *First Principles* (1872). Spencer defined evolution as

> a change from a state of indefinite, incoherent homogeneity towards a state of definite, coherent heterogeneity. (p.396)

The notion of evolution as a movement from disorder to order was explored by Schrödinger (1944) who established a link between genetics and thermodynamics by maintaining that life was to be defined in `negentropic' terms, that is, in terms of a genetically-based propensity towards the maintenance of order. Following this approach, Chaitin (1970) restated the problem of life and evolution in terms of organization and complexity: According to algorithmic information theory, the complexity of a phenomenon is equal to the length of the shortest program necessary to compute the phenomenon. On this basis, Chaitin argues that the life question can be defined as the problem of the relation between wholes and parts, viz.

> if both are equally complex, the parts are independent (do not interact). If the whole is very much simpler than the sum of its parts, we have the interdependence that characterizes a living being. (p.15)

Thus, phenomenal emergence, in this case, the emergence of life from matter, implies a *reduction* in complexity and an *increase* in *organization*. According to Chaitin, complexity and organization are, therefore, in reciprocal relation. This informational or entropic conception of life finds support within biology. For example, Orgel (1973) maintains that

> it is difficult, if not impossible, to find a rigid definition of life that incorporates all of our intuitive ideas. Instead, we can make a *list* of the attributes that help us to decide whether or not a system is living - reproduction, metabolism, excitability, and so on - and agree to call an organism alive if it possesses a suitable selection of these attributes. This is a useful approach in introductory discussions of terrestrial biology, but it is not so useful when one discusses *alien* forms of life. In the latter case, it is too difficult to complete the list; it is impossible to enumerate all the types of behaviour that might characterize nonterrestrial forms of life. (pp.191-192)

Orgel lists the following as *necessary* and *sufficient* conditions that a structure must fulfil in order to qualify as `living':

1. The object is *complex* and yet *well-specified*.
2. The object is able to *reproduce* (or alternatively, the object may be the descendant of related objects that can reproduce, even if it is itself `sterile').

Furthermore, these conditions imply that

(a) the object is a product of *natural selection* (or *human technology*), and

(b) the *information needed to specify the object* is *stored in a structure that is stable* for the reproductive lifetime of the object.

Orgel maintains that "a new term for such `living' organisms, whether terrestrial or not, must now be introduced. They are Complex Information-Transforming Reproducing Objects that Evolve by Natural Selection - *CITROENS*." (p.193) He goes on to state that

> we are familiar with many products of technology that fail to be CITROENS only because they do not reproduce autonomously. There is, thus, an exception to the rule that objects of high information content must be the products of human ingenuity. This exception does not weaken the argument, since the intelligent 'creators' in this case are themselves the products of natural selection. (pp.196-197)

This leads Orgel to an informational or computational view of life. He argues that biologists should "concentrate on the *structure* and *behaviour* [function] of .. objects rather than their status as `living' or `nonliving' beings." (p.189) This approach leads to the following position, viz. "living organisms are distinguished by their *specified* complexity." (p.189) However, he maintains that

> these *vague* ideas can be made more precise by introducing the idea of *information*. Roughly speaking, the information content of a structure is the minimum *number* of instructions needed to specify the structure. One can see intuitively that many instructions are needed to specify a complex structure. On the other hand, a simple repeating structure can be specified in rather few instructions. (p.190)

Orgel's position is important since it provides implicit support for a computational functionalist (that is, computationalist) approach to life, viz.

> the structure and behaviour of an object would need to be nonrandom and reasonably complicated to interest the student of extraterrestrial [or *artificial*] life. We have already seen that a great deal of information is needed to specify the structure of a complicated nonrandom object. It may be concluded that anything that we would want to call `living' would have to have a high information content [that is, be specified by a large number of instructions]. (p.192)

A concept closely related to the computational theory of life (CTL), viz. artificial life (or A-Life), is described in section 4.6.6. However, before discussing that concept it is necessary to examine some notions associated with the concept of life.

4.6.3. The Continuum Hypothesis

Life can be interpreted on an emergentist framework in essentially two ways: (1) as a *discrete* (or binary) property which appears with a specific organization of the non-living substrate; and (2) as a property which is *continuous* with the other properties of the non-living substrate. An early advocate of a discrete-essentialist approach was Broad (1925) who maintained that "*all* bodies which would be said to be `alive' behave differently in many ways from *all* bodies which would be said not to be `alive' [emphasis added]." (p.53) This position was also advanced by von Neumann (1966) who postulated a *threshold of complexity* separating `simple' systems from `complex' systems; according to this view, the former are able to produce systems of lower complexity only, whereas the latter, by contrast, are able to create systems of higher complexity than themselves. Hence, according to von Neumann, life necessitates a certain degree of complexity. Bennett (1990) has contested this view on the grounds that it fails to differentiate between potential and actual complexity. In support of this contention, he cites the example of a corpse: It is structurally complex and, on a computationalist view, structure (form, syntax) implies function. But is it alive given that the function is not being computed, that is, actualized ? This necessitates a consideration of the distinction between programs and processes (chapter 2). On the computationalist view, the former are potentially the latter whereas the latter are actualizations of the former. The problem then becomes how to transform a program into a process, that is, how to convert something which is static into something which is dynamic. Including a physical substrate (`hardware') on which the program (`software') executes is unacceptable from a computationalist perspective since computation would then be supervenient on matter and hence, non-ontological in a primordial sense.

Levy (1992) provides the following overview of the continuum hypothesis:

> Some scientists suggest that the definition-of-life question is a red herring. Life, they say, should be gauged on a continuum, and not granted according to binary decision. A rock would certainly be low on any continuum of aliveness, and a dog, a tree, and a human being would rank highly. More ambiguous systems would fall in a middle region of semi-aliveness - somewhere below bacteria, which almost everyone agrees are alive, and somewhere above rocks. Viruses, which some biologists consider living and others do not, would reside in the upper reaches of this middle ground. Below that would would be complex systems that no one really considers to be alive but that display some behaviours consistent with living organisms - things such as the economy and automobiles. (pp.6-7)

This position is endorsed by Farmer (1991) who also maintains that life should be considered as a continuum property of organizational patterns, with some more or less alive than others. Levy further states that "there is a particular advantage in regarding life in this manner: using systems that no one would classify as truly alive, biologists could nonetheless isolate the *qualities* of life." (p.7) Godfrey-Smith (1994) describes three versions of the continuum thesis based on an analysis of Spencer and Dewey's approaches to the mind-life relation, viz.

Ontological

Weak Continuity: Anything that has a mind is alive, although not everything that is alive has a mind. Cognition is an activity of living systems.

Strong Continuity: Life and mind have a common abstract pattern or set of basic organizational properties. The functional properties characteristic of mind are an enriched version of the functional properties which are fundamental to life in general. Mind is literally life-*like*.

Methodological

Methodological Continuity: Understanding mind requires understanding the role it plays within entire living systems. Cognition should be investigated in this whole organism context. (p.83)

Weak continuity is closely related to the discrete life hypothesis as described above whereas strong continuity may be identified as the continuum hypothesis proper. Orgel's (1973) position on this issue is interesting:

> It follows immediately that any `living' system must come into existence either as a consequence of a long evolutionary process or a miracle .. At first, replicating structures are formed that have low but nonzero information content. Natural selection leads to the development of a series of structures of increasing complexity and information content, *until one is formed which we are prepared to call `living'*. (p.192)

This statement can be interpreted as asserting that life is *ontically continuous with* yet *epistemically distinct from* matter, a view which is consistent with the emergence-relative-to-a-model concept (chapter 3) described by Cariani (1991) in which ontological reductionism without epistemological reductionism is maintained. Pattee (1989) defines the problem of life in terms of the measurement problem, viz. the production of *records*. However, in addition to the `downwards' problem of explaining life in terms which are consistent with a physico-chemical ontology, there is also an `upwards problem', viz. explaining the emergence of mind, more specifically consciousness, from a non-mental substrate (chapter 7). Adoption of either thesis (continuum or discrete) necessitates describing life at two complementary levels, viz.

Level 1 - `building blocks' or components of life (physical-chemical explanation)
Level 2 - `characteristics' or properties of life (biological explanation)

The discrete-continuum issue conceals a much deeper problem, viz. the monism-pluralism issue with respect to categorial ontology (chapters 3, 5, 6 and 7). For example, Needham (1974) maintains that

> it would be correct to say that the living differs from the dead in degree and not in kind because it is on a higher plane of complexity of organization, but it would also be correct to say that it differs in kind, since the laws of this higher organization only operate there. (p.55)

The question is whether or not the laws at the higher level are ontologically reducible to those at the lower level. If not, then a form of ontological pluralism is necessary. A difference of *degree* implies an *intra*-categorial difference whereas a difference of *kind* implies an *inter*-categorial distinction. Intra-categorial difference tends towards categorial monism whereas inter-categorial distinction tends towards categorial pluralism. Although it is possible to assert an inter-categorial pluralism with respect to epistemology while maintaining an intra-categorial monism with respect to ontology and, in fact, this is the position adopted by most proponents of the concept of emergence (chapter 3), this does not solve the *category problem* (chapter 7) since the latter is ontological not epistemological. It is interesting to note that category problems (of one kind or another) are ubiquitous throughout naturality and artificiality. For example, in the context of a discussion of artificial life (section 4.6.6), Bedau (1991) assumes a Cartesian position with respect to the life-mind relation in maintaining that

> the simulation-or-reality debate seems more tractable in artificial life than in artificial intelligence because ALife can sidestep some of AI's sharpest thorns - life need not involve subjectivity and self-consciousness. (p.498)

However, Bedau's assertion that

> progress on ALife's simulation-or-reality debate might even help break the impasse in the analogous debate in AI (p.498)

is problematic because of the other-minds problem (section 4.5.5) and the possibility of ontological pluralism with respect to the categories of subjectivity and objectivity, viz. ontological subjectivity cannot be reduced to ontological objectivity (chapter 7).

4.6.4. Life and Functionalism

A functionalist approach to life is implicit within Orgel's (1973) information-theoretic formulation of the concept (section 4.6.2). This move, involving a dualistic separation of form from matter and the identification of *necessary* vital properties with the former, is supported by Simons (1983) who maintains that

> living systems can be recognized according to how they process information and energy, how they are structured, how they behave, and so on, r*ather than by the specific chemistries by which they accomplish their tasks* [emphasis added]. (p.6)

According to Simons, "the idea that life can be recognized independently of the substance out of which it is constructed derives support from modern functionalism .. Functionalism is largely concerned with mental phenomena but some of its elements are equally applicable, *mutatis mutandis*, to an identification of the characteristics whereby life is to be recognized." (pp.7-8) In support of this position, Sober (1991) states that

> recent philosophers of biology have made the [functionalist] point by arguing that an organism's fitness is the upshot of its physical properties even though fitness itself is not a physical property.

> What do a fit cockroach and a fit zebra have in common ? Not any physical property, any more than a wood and wire mousetrap must have something physical in common with a human mouse catcher. *Fitness is multiply realizable*. (p.753)

Sober is led to assert that "behaviourism is a mistake in psychology, but it may be the right view to take about many biological properties." (p.759) This view is based on the assumption that the causal mechanism involved in, for example, photosynthesis within plant cells is both (1) objectively describable (epistemological assumption) and (2) objective in nature (ontological assumption). This ontic and, as a corollary, epistemic objectivity makes Turing Tests for life (section 4.6.8) a realistic possibility; furthermore, the separation of form from matter and the association of behaviour with the former provides support for the multiple-realizability thesis, and thereby, the CTL. As Simon (1983) states,

> if a system can reproduce and also handle energy and information in appropriate ways then the system has a claim to be regarded as living. A corollary is that the genesis of the system is irrelevant. A mechanically *assembled* system [or artifactual system] may reasonably be regarded as living if its internal functions and behaviour in the world fulfil the necessary criteria. (p.7)

It should be noted that functionalist, essentialist accounts of life are ahistorical with respect to *poiēsis* (coming-forth or becoming). However, it may well be the case, as will be argued in chapter 6, that the nature of *poiēsis* places constraints on Being such that (i) the Being (that is, ontology) of naturals and the being of artificials (as artifactuals) are necessarily distinct ontological modalities and (ii) the possibilities for emergence in the former are essentially different from those associated with the latter[50].

Prior to examining vitalistic functionalism in the context of A-Life (section 4.6.6), it is worthwhile briefly examining a case study.

4.6.5. Bedau's Concept of Life: A Case Study

Bedau (1996) begins by asserting that "we can only search for life if we have a prior conception of what life is." (p.333) This view is consistent with the hermeneutic solution to the recognition problem described in section 4.5.1. He considers three alternative conceptions of life, viz.

1. life as a loose cluster of properties (or Wittgensteinian family resemblances)
2. life as a specific set of properties (the essentialist thesis)
3. life as metabolization

He regards (1) as "a fall-back position that can be justified only after all candidate unified views have failed." (p.335) After discussing various essentialist conceptions of

50 Specifically, that naturals support *open* ontological emergence whereas artificials (as artifactuals) support *closed* (or bounded) emergence (chapters 6 and 7).

life (2), Bedau concludes that what is missing from all such accounts is a *unifying* cause, that is, a single cause which is responsible for generating the diverse range of properties that can be associated with life. With respect to (3), Bedau maintains that "any possible form of life that persists in the face of the second law of thermodynamics apparently must have a metabolization .. [Thus], metabolization is at least a necessary condition of all physical forms of life." (p.337) However, metabolization is not a *sufficient* condition since this would entail viewing candle flames and vortices as living entities, a position which conflicts with scientific intuition. Furthermore, it is doubtful whether metabolization can account for the other properties characteristic of life.

For this reason, Bedau proposes a conception of life defined in terms of the evolutionary process of adaptation, referring to "the *suppleness* of the adaptive process and its unending capacity to produce novel solutions to unanticipated changes in the problems of surviving, reproducing, or, more generally, flourishing [emphasis added]." (p.338) He further maintains that "natural selection will yield supple adaptation only if the criteria for selection change as the system evolves." (p.339) These views in combination lead Bedau to assert that

> the entity that is living in the *primary* sense of that term is the supplely adapting system itself. Other entities that are living are living in a *secondary* sense by virtue of bearing an appropriate relationship to a supplely adapting system. (p.339)

On this basis, a *definition* of life is proposed, viz.

A x is living *iff* x is living_1 or x is living_2.

B x is living_1 *iff* x is a system undergoing supple adaptation.

C x is living_2 *iff* there is some living_1 system y such that either (1) x meets condition A_1 and y meets condition B_1 and x bears relation C_1 to y or (2) x meets condition A_2 and y meets condition B_2 and x bears relation C_2 to y or ... or (n) x meets condition A_n and y meets condition B_n and x bears relation C_n to y.

Bedau states that "whereas some might refer only to a system's *capacity* [or *potentiality*] to undergo supple adaptation, I hold that life involves the *exercise* [or *actualization*] of this capacity. For me the key is not supple adaptability but supple adaptation." (p.340) Hence, Bedau's scheme meet's Bennett's criticism of von Neumann with respect to potential and actual complexity (section 4.6.3).

Bedau's definition is important because it is able to handle anomalous cases such as viruses and mules, the latter of which are clearly living yet non-reproductive. As Bedau states, "these infertile organisms exist only because of their connections with other, fertile organisms which do play an active role in a biosphere that undergoes supple adaptation." (p.340). (On Bedau's scheme, viruses and mules would be classified as living_2.) However, Davidge (1992) has contested the validity of this `populational' approach, arguing that "life does not occur at the general level, it occurs at the individual level; it is not populations that are alive, but individuals." (p.450) In response, Bedau states that,

> one might worry that it is a category mistake to think that an evolving system could be alive. This worry originates with the idea that individual organisms are the entities that are alive and concludes that life cannot be a population undergoing supple adaptation, since the whole evolving population of organisms is of a different logical category from an individual organism. However, this objection has no force for those who are seeking the fundamental explanation of the diversity of living phenomena. Supple adaptation would provide this explanation even though an individual living organism is itself only a small and transitory part of the whole adapting population." (p.340)

As stated previously, Bedau identifies supple adaptation with the open-ended evolution of adaptive traits, viz. "if we continually see (on a relatively long time scale) new clusters of traits that are persistently used (on a relatively short time scale) significantly more than would be expected in the absence of adaptation, then we have positive evidence for the occurrence of the process of supple adaptation." (p.346) A quantitative metric for measuring supple adaptation based on trait usage is defined which leads him to assert that "a system's level of vitality *V(t)* reflects the extent to which new significant adaptations are arising and persisting. So if we view life as supple adaptation, we can use a system's vitality *V(t)* to define the degree to which it is living or involves life. By this sort of means, a system's usage distribution *U(t,u)* and vitality *V(t)* could figure centrally in the explanation of the system's supple adaptation, and perhaps even the explanation of the extent to which the system involves life." (p.354)

The concept of supple adaptation is a development of the idea of *intrinsic* adaptation introduced in (Packard,89) and subsequently refined in (Bedau,91). In *intrinsic*-adaptation, teleology is an *a posteriori* or teleonomically-emergent property of the system while in *extrinsic*-adaptation, systemic teleology is specified *a priori*. Naturalistic evolution is held to be the paradigmatic instance of the former while genetic algorithms and learning in supervised artificial neurals nets may be identified as examples of the latter. However, there is a problem with both supple adaptation and intrinsic adaptation. Davidge (1992) maintains that living systems must only be teleological in an *a posteriori* sense and Bedau (1991) supports this position. However, while the teleology at the *global* (systemic, populational) level is assumed to be emergent or *a posteriori*, it is in fact parasitic on the *a priori* teleology of components (that is, individuals) at the local level. In the computational models investigated by Bedau and Packard, the components are `bugs' whose behaviour (ethology) is limited to foraging. However, foraging *is* a teleological activity. Hence, the global *a posteriori* teleology of the bug population is *supervenient* on the local *a priori* teleology of individual bugs. Hence, contrary to claims made by Bedau (1991, 1994), teleology - at least in this case - is not emergent in the sense of an inter-categorial distinction or difference of kind (section 4.6.3). What is required is the emergence of teleology in a non-teleological substrate via some form of self-organizing process (chapter 3). Although Bedau and Packard's models do not provide such a substrate, since their model tacitly assumes a teleological substrate (bug ethology) as ontologically primitive, other models have been proposed which do provide, at least in theory, the necessary framework. These are briefly examined in chapter 5. One potential candidate is *Tierra* (Ray,91), an A-Life ecosystem in which organisms are identified with self-replicating code segments within the `virtual machine' universe (chapters 2 and 5) provided by the CPU and RAM of a computer. It is often claimed by computationalists that teleological behaviour in *Tierra* is *a posteriori* or emergent. However, upon closer inspection this assertion can be shown to be false. For example, competition for resources is explicitly coded into the *Tierran* universe. As Ray (1994) himself states,

> evolving digital organisms will compete for access to the limited resources of memory space and CPU time, and evolution will generate adaptations for the more agile access to and the more efficient use of these resources. (p.14)

However, competition *is* a teleological activity. Furthermore, and perhaps more importantly, Tierran organisms are not *primordially*-emergent, that is, they do not emerge from the computational substrate in which they are embedded via self-organization; rather the Tierran `world' has to first be `seeded' with a primitive population of replicators which, under Darwinian evolution, evolves to generate organisms of increasing diversity and complexity. As Ray (1991) admits,

> while the origin of life [that is, primordial biotic emergence] is generally recognized as an event of the first order, there is another event in the history of life that is less well known but of comparable significance: the origin of biological diversity and macroscopic multicellular life during the Cambrian explosion 600 million years ago .. The work presented here aims to parallel the second major event in the history of life, the origin of diversity. *Rather than atempting to create prebiotic conditions from which life may emerge, this approach involves engineering over the early history of life to design complex evolvable organisms, and then attempting to create the conditions that will set off a spontaneous evolutionary process of increasing diversity and complexity of organisms* .. From *a single rudimentary ancestral creature* [or `seed'] containing *only* the code for self-replication, interactions such as parasitism .. hyer-parasitism, sociality, and cheating have emerged spontaneously [emphasis added]. (p.373)

The crucial point to appreciate in the context of the present discussion is the non-emergence of teleology in the categorial sense. Hence, rather than demonstrating how teleology may be reduced to the non-teleological, such studies in fact support the hypothesis that teleology is a categorial primitive, that is, teleology is ontological. This is consistent with the Aristotelian claim that final causation constitutes a primitive, non-reducible *type* of causation (chapter 6).

4.6.6. Artificial Life (A-Life)

Langton (1989b) defines artificial life (A-Life) as

> the study of man-made [artifactual] systems that exhibit behavious characteristic of natural living systems. It complements the traditional biological sciences concerned with the *analysis* of living organisms by attempting to *synthesize* life-like behaviours within computers and other artificial media. By extending the empirical foundations upon which biology is based beyond the carbon-chain life that has evolved on Earth, Artificial Life can contribute to theoretical biology by locating *life-as-*

> *we-know-it* within the larger picture of *life-as-it-could-be*. (p.1)

While criticizing A-Life researchers for their naivety with respect to adherence to technical detail, Miller (1995) asserts that "`life as it could be', logically and extra-terrestrially" (p.17) constitutes a valid research programme for A-Life. However, Miller goes on to state that "A-Life, despite its pretenses to expanding the scope of possible biologies, has been tied far too tightly to real biology." (p.20) This statement refers to the problem of distinguishing *necessary* from *contingent* properties and its solution will determine the extent to which life in general is coupled to the specificity of terrestrial life. Belew (1991) maintains that "ALife's goal is to abstract the `logical form' of life, independent of the particulars of the carbon-based biological life (BLife) forms that arose on this planet and with which biology is almost exclusively concerned." (p.8) This position is supported by Langton (1989b) who asserts that

> life is a property of *form*, not *matter*, a result of the organization of matter rather than something that inheres in matter itself .. It is effects, not things, upon which life is based - life is a kind of behaviour, not a kind of stuff - and as such, it is constituted of simpler behaviours, not simpler stuff. (p.41)

The A-Life approach can be summarized as follows:

> Whereas biology has largely concerned itself with the material basis of life, Artificial Life is concerned with the formal basis of life .. [It] starts at the bottom, viewing an organism as a large population of *simple* machines, and works upwards *synthetically* from there - constructing large aggregates of simple, rule-governed objects which interact with one another nonlinearly in the support of life-like, global dynamics. The `key' concept in AL is *emergent* behaviour. (p.2)

This leads to the following methodology within A-Life research (Langton,89b):

- bottom-up rather than top-down modelling
- local rather than global control
- simple rather than complex specifications
- emergent rather than prespecified behaviour
- population rather than individual simulation

The above methodology is based on three assumptions, viz.

1. "that the `logical form' of an organism can be separated from its material basis of construction, and that `aliveness' will be found to be a property of the former, not of the latter." (Langton,89b;p.11)

2. "that the essential nature of the fundamental principles of life can be captured in relatively simple models." (Bedau,92;p.494)

3. The continuum hypothesis (section 4.6.3) : "the ALife-AI claim is, `The smartest dumb thing you can do is stay alive.' That is, ALife represents a lower bound for AI." (Belew,91;p.9)

Emmeche (1993) presents five different conceptions of life: (p.559)

- *GOFBO*: Good Old Fashioned Biological Organisms
- *MOMACE*: Modern Macromolecular-based Cells
- *ABLI*: Abstract Life (in a biochemical medium or formal/symbolic space)
- *ROLI*: Robotic Life (animats, nanobots, neo-cybernetic systems etc)
- *CYBERLIFE*: life-like structures in virtual realities

Wheeler (1996) defines an *animat* as an artificial animal or artificial autonomous agent, whereby the latter is meant

> any adaptive system which, while in continuous long-term interaction with its environment, actively behaves so as to achieve certain goals. (p.210)

The animat approach has been investigated by Steels (1994) in the context of an exploration of assumption (3), viz. the continuum hypothesis. However, it will not be examined further in the context of the present discussion since it is a non-computational[51] form of A-Life. Various problems associated with life have been studied and the results of such investigations presented in the proceedings of the A-Life conferences (Langton,89a) (Langton,91a) (Varela,92) (Langton,93) (Brooks,94b) (Moran,95): For example, self-organization, the origin of life, evolutionary dynamics (punctuated equilibria, coevolution, Lamarckism etc), learning and communication, cultural evolution, and philosophical issues such as necessary-contingency and necessity-sufficiency problems with respect to the properties of life, matter-form relations and the simulation-realization issue. Most such investigations take place `in silico', that is, in computational media (or *substrates*). According to Langton (1989b),

> computers provide an alternative medium [to `wet' carbon-chain chemistry] within which to attempt to synthesize life. Modern computer technology has resulted in machinery with tremendous potential for the creation of life *in silico*. (p.39)

Langton goes on to list the following properties of computational (or *in silico*) A-Life:

1. They consist of simple programs or specifications.

2. There is no single program that directs all of the other programs (ie: distributed control).

3. Each program details the way in which a simple entity reacts to local situations in its environment, including encounters with other entities.

4. There are no rules in the system that dictate global behaviour (This is not strictly correct; the 'physics' of the computational substrate is a global unifying property of the system).

5. Any behaviour at levels higher than the individual programs is therefore emergent.

Support for the *in silico* approach to A-Life is provided by Orgel (1973) who maintains,

---

51 Non-computational because not *completely* computational in ontology: Robotic A-Life may contain formal, computational elements; however, the former cannot be *abstractly* specified completely in terms of the latter.

in the context of a discussion of the genetic program, that

> nothing comparable to it exists in the inanimate world, except for manmade computers. (p.55)

Additional support for the computational approach to A-Life is implicit in one of the foundational concepts underlying computationalism and conventional AI, viz. the physical symbol system hypothesis or PSSH (chapter 2). Statements by Newell et al. (1976) in which computers are referred to as "our organism, the machine" (p.113) and assertions such as "the machine - not just the hardware, but the programmed, living machine - is the organism we study" (p.113) establish a clear link between the PSSH, the CTL and A-Life.

4.6.7."Strong" and "Weak" A-Life

"Strong" and "weak" positions corresponding to those defined for AI can be established for A-Life. Belew (1991) defines the "strong" position as follows:

> ALife simulations are, or at least can become, first-class examples of living systems. (pp.12-13)

However, a more precise formulation, distinguishing "strong" from "weak" A-Life, is presented by Kawata et al. (1994):

> In the weak approaches, Artificial Life programs are made to simulate the life of known, existing organisms in order to understand the processes of real organisms. In the strong approach, researchers try to create Artificial Life, and search for the nature of life that may or may not be found on Earth. (p.417)

Miller (1995) distinguishes between "strong" and "weak" A-Life as follows, viz.

> `strong A-Life' (computer processes as realizations of living systems) versus `weak A-Life' (computer processes as simulations of living systems). (p.21)

Emmeche (1991) identifies "strong" (realization) and "weak" (simulation) theses associated with three forms of A-Life, viz. (i) computational (`software'), (ii) robotic (`hardware') and (iii) chemical (`wetware'). Additionally, two multiple-realizability positions are defined, viz. (1) "the thesis of medium-dependent life in multiple possible media" (p.83), and (2) the Platonic/formalist thesis of medium-independent life connected with notions of either self-organization and emergence or a qualitative set of life criteria. On the first view, physicality (that is, material embodiment) is a *necessary* condition for life while on the second view it is a merely contingent property.

Before examining a Turing Test for life, it is worthwhile reconsidering a problem which was introduced earlier in the context of mind and which has a direct bearing on the simulation-realization issue. In section 4.5.8, the idea that the simulation of intentionality is identical to intentionality was briefly investigated and shown to be based on the assumption that intentionality and information processing belong to the same ontological category. A similar claim can be made in the context of the CTL or A-Life. For example, Taylor et al. (1989) maintain that in *RAM*, an A-Life ecosystem simulator, it is observed that

> the life of an organism is in many ways similar to the execution of a program and .. the global (emergent) behaviour of a population of interacting organisms is best *emulated* by the behaviour of a corresponding population of co-executing programs [emphasis added]. (p.275)

The same argument originally presented in section 4.5.8 in the context of the CTMi can be applied here, viz. emulation (or realization) can only be applied intra-categorially and it is precisely whether or not life and computation belong to the same category that is to be established[52].

4.6.8. A Turing Test for Life

The Turing Test was described in section 4.3.6. In terms of the presentation introduced there, a Turing Test for life would involve replacing the computer, *A*, with a computational A-Lifeform and replacing the human subject, *B*, with a living organism. *A* would then attempt to replicate the behaviour of *B*. The Turing Test for life assumes with Sober (section 4.6.4) that life is an objectively definable *physical* phenomenon. Consequently, the partition separating *A* and *B* from the interrogator, *C*, must be relocated. (The function of the partition is to conceal the *ontology* of the participants from the interrogator such that immediate classification on the basis of non-essential evidence is prevented.) Defining the partition in a *computational* A-Life context is problematic: If *A* and *B* must be observed *directly*, it will be obvious which is the natural life-form and which the A-Lifeform. A possible solution to this problem involves adopting a definition of life such as that presented by Bedau (section 4.6.5) and comparing *indirect* evidence of lifelike behaviour, for example, trait usage (that is, allele expression) statistics in a population of organisms. The problem with this approach is that it may not always be possible to gain access to details regarding the corresponding parameter in the natural organism.

There are other problems with the Turing Test for life. As Moreno et al. (1994) state, "we still do not know what is contingent and what is essential - universal - in the basic mechanisms of living phenomena." (p.406) However, it should be recognized that the mere statement of the necessary-contingency problem *as a problem* assumes

---

52 However, Keeley (1993) contests this position on the grounds that the first-person problems associated with mind and consciousness do not extend to life which, on the conventional biological view, is a third-person (or ontologically-objective) phenomenon in the same way that information-processing or computation is a third-person phenomenon. This position may, in turn, be contested following Searle (1992, 1995), Tallis (1994) and Lanier (1995b), who maintains that computation *as distinct from intrinsic natural phenomena such as life* is an *extrinsic* phenomenon that is ontologically-dependent on human beings.

functionalistic-essentialism *a priori*; that is, the very possibility of defining a Turing Test for life necessitates holding life to be definable in functionalistically-essentialist terms and not to be a cluster concept. Additionally, as Emmeche (1992) states,

> it is not clear what kind of criteria can be used to evaluate theories and models of `life as it could be' in a non-trivial subset of possible worlds .. As anything is possible in pure imagination, AL has to take recourse to the earthly biology to see if a particular instance of an artificially constructed model of life has a plausible behaviour. (p.468)

This point is extremely important since it draws attention to the necessarily *geocentric* (or earth-centred) nature of Turing Tests for life and, by implication, to the fact that a categorial `cut' (chapter 6) must be made in order to classify an A-Lifeform as a simulation or realization (instantiation) of life.

On the basis of such criticisms, Davidge (1992) maintains that "there is no [currently existing] test equivalent to the Turing Test of AI" (p.448). However, he offers the following as an element of a possible Turing Test for life: "AL systems can be judged on the correspondence between their proposed level of biological analogy and their implementation primitives." (p.453) In this connection, the following table listing analogies between hardware and wetware is presented:

Natural Life	*Artificial Life*
energy	energy
atom	electron
carbon	silicon
molecule	transistor
biochemical process	logic gate
organelle	ALU, registers
cell	processor
multicellular	multiprocessor

Table 4.3 Vitalistic analogies between wetware and hardware.

However, this approach is incompatible with an *in silico* or computationalist - that is, *software* - approach to A-Life[53]. For this reason, other formulations of the Turing Test for life must be considered. Pattee (1989) maintains that A-Lifeforms should (in order to be testable against empirical real-world evidence) only be delimited by adherence to universal physical laws and natural selection. He holds that "strong" emergence (realization as opposed to simulation) could form the basis of a Turing Test for life if there was consensus on how to recognize emergent behaviour. However, identifying life

53 Strictly-speaking, this criticism is incorrect since computationalism supports the construction of virtual machine hierarchies (chapter 2 and 5) and hence, the *substitution* of hardware by software and visa-versa; in short, computational (or software) analogues of hardware analogues of wetware *are* possible, in which case Davidge's definition of the Turing Test for A-Life may be valid.

with emergence is problematic since although emergence is a *necessary* condition for life (assuming an evolutionary perspective), it is clearly not a *sufficient* condition: For example, NaCl is emergent from Na and Cl, yet is not viewed as living (chapter 3). Pattee clarifies his position by identifying a particular concept of emergence, viz. *measurement*, as characteristic of living systems. On his view,

> new measurements can be considered as one of the more fundamental test cases for emergent behaviour in artificial life models. For this purpose, we may define a generalized measurement as a record stored in the organism of some type of classification of the environment. This classification must be realized by a measuring device constructed by the organism. The survival of the organism depends on the choice and quality of these classifications, and the evolution of the organism [or species] will depend on its continuing invention of *new* devices that realize *new* classifications. (pp.73-74)

Consequently, Pattee is led to maintain that

> by formalization of life, one may be throwing out the whole problem, that is the problem of the relation of symbol to matter (p.69).

He presents three possibilities for the symbol-matter mapping, viz.

1. we can simulate everything by universal symbol systems (PSSs).
2. we can realize universal symbol systems with material constructions.
3. we can realize endless types of structures and behaviours by symbolic constraints on matter. (p.70)

On this framework, it appears that "strong" computationalism - that is, *realization* as opposed to *simulation* - is impossible. However, it is an open question whether or not matter can be reduced to computation (in which case the symbol-matter distinction becomes merely epistemic). In short, if a computational interpretation of matter *is* possible, Pattee's classification of computational A-Lifeforms as *necessarily* simulations ("weak" A-Life) must be invalid. A computational conception of matter is described in section 4.7 and chapter 5.

Finally, it is worthwhile examining a claim by Davidge (1992) to the effect that "intelligence and life are inseparable. If you recognize one, you will recognize the other." (p.448) If Penrose's position, viz. intelligence is inseparable from awareness and consciousness (section 4.5.4), is accepted, then life is seen both to imply and be implied by consciousness. This view leads to a form of panpsychism (chapter 1) with respect to the phenomena of life and mind. However, according to Keeley (1993), life is an ontologically-objective phenomenon whereas consciousness is an ontologically-subjective phenomenon; consequently, it is possible to pass a Turing Test for life on the basis of objective behavioural evidence and yet fail a Turing Test for mind (more precisely, consciousness) because of the other-minds problem (section 4.5.5)[54].

4.6.9. Problems with the CTL

There are a number of issues associated with the computational theory of life (CTL). In this section, five related problem areas are identified and briefly discussed, viz. (i) abstaction, (ii) physicality, (iii) embeddedness, (iv) hermeneutics, and (v) morphization.

4.6.9.1. Abstraction

Rosen (1993) maintains that "understanding how something works also tells you how to build it, and conversely." (p.93) According to Rosen, in a *machine*, epistemology and ontology coincide (a position which is explored in detail in chapters 6 and 7). A machine is "fractionable" or decomposable (chapter 3), that is, can be partitioned into components which can be analyzed separately and then synthesized to produce the original machine. This allows for the separation of structure from function and supports the possibility of multiple realization. However, Rosen argues that life is non-fractionable: For example, the engine and airfoil of a bird (a natural) is its wing. In this case, two functions are located within one physical structure; in an aeroplane, by contrast, engine and airfoil are separate structures. Taking the bird apart kills the bird whereas taking the plane apart does not prevent it from being put back together again[55].

Emmeche (1992), while accepting the possibility of system decomposition *in principle*, questions whether or not

> life [is] a multi-media-realizable phenomenon because it is *intrinsically computational*, or because the form of movement of *any* specific natural phenomenon (that can be described by an algorithm) can be realized by a computational setup. (p.467)

The CTL is based on the assumption that what is essential to life is a pattern of movement, a dynamic processual organization which is formalizable in computational terms. However, this view assumes *a priori* (1) that the matter-form distinction is valid and that form can be *objectively* abstracted from matter and (2) that matter itself, ultimately reduces to form (chapter 2 and section 4.7.6). This commitment to realism within the context of the CTL is criticized by Emmeche who provides implicit support for *observational-relativism* (chapter 3) in his adoption of a semiotic perspective, viz.

---

54 In defense of Davidge, Keeley might argue that it is not the former's position that is incorrect, but Penrose's assertion of a necessary link between intelligence and consciousness. On this view, emergentism is retained while panpsychism is rejected.

55 Interestingly, Simon (1969) cites the same example, viz. bird and aeroplane, in arguing the case in favour of functional decomposition. However, his argument is at a purely *conceptual* level, and hence, does not conflict with Rosen's position.

> our *models* of the logic of living systems are not necessarily instances of the true logic inherent in the very systems themselves .. The physical/chemical processes within an organism are of a different kind, described by a different set of theories, than the processes within a computer running some programme. Their functions may be similar on some level of description, but the inherent logic of the processes, on the physical/chemical level (and probably on higher levels as well), is likely to be different. (p.469)

While contesting realism in the context of computationalism, the matter-form distinction is held to be valid; consequently, he accepts the possibility that life may be a computational phenomenon although the *concept* of computation (chapter 2) would have to be redefined in a vital (that is, biological) context. However, if computation is an artifactual concept and if life is a computational phenomenon, it follows that life must be artifactual. This view is unacceptable on a naturalistic evolutionary (specifically Neo-Darwinian) framework. There are two possible solutions to this problem: Either (1) computation is a *natural* phenomenon or (2) life is a *non-computational* phenomenon[56].

Emmeche (1992) lists three objections to functionalism and the multiple-realizability thesis, viz.

1. It does not guarantee that our formalization of specific systems - whether mental, biological, or physical - can catch all the essential factors that govern such a system. There might even be aspects of the system that are in principle unformalizable.

2. The construction (of any material kind) that implements the formal structure .. is still in need of our interpretation in order to give any meaning .. semantics is not intrinsic to syntax.

3. The functioning of a construction implementing some formal structure may well be functionally equivalent to other implementations (or realizations) on one chosen level of description, while on another level it may show dissimilar properties that from a biological point of view may seriously affect its chances of survival in a realistic environment (pp.471-472).

Three issues are identified: (i) the possibly irreducible role of *matter* in life, (ii) the importance of *hermeneutics* or interpretation in the context of formal-computational systems, and (iii) *abstraction*, which has been dealt with already. However, what is most significant from the point of view of this study and yet which is not included in the above list is the assertion that "one cannot separate cognition from volition and emotion [since] these `psychical' properties are features of genuine biological processes. As the `psyche' of man or animal in this sense is medium-dependent, so is a living organism's teleonomic orientation and relation to its environment." (p.472) This point, which is central to the critique of computationalism presented in this thesis, will be examined separately in chapter 7. In the remainder of this section, two problems associated with embodiment (physicality and embeddedness) and two problems associated with interpretation

---

56 In chapters 6 and 7, explicit arguments will be presented against (1) and implicit arguments in favour of (2).

(hermeneutics and morphization) will be examined[57].

4.6.9.2. Physicality

In defense of the view that embodiment or physicality is an important, perhaps *necessary*, condition for vitality, Emmeche (1992) presents instances of biological phenomena[58] involving processes which are *apparently*[59] incapable of being described in purely computational terms. In support of this position, Moreno et al. (1994) maintain that a strict formalist approach to modelling life will ultimately prove unsuccessful because information implicit in the dynamics associated with matter cannot be completely captured in a formal representation. They argue that

> if it were the case that the relations among components arise from their material properties, the complex organization of living systems could not be fully understood except by recourse to the properties of living matter. (p.407)

In support of this contention, they cite an example, the problem of self-reference, maintaining that "it is not very clear that a self-reference understood as a syntactic connectivity of components is enough to explain other living phenomena such as self-reproduction or evolution. Those seem to require some sort of semantics that appears to be very much in-built in the specific materials that take part in the living organization." (p.408) Thus, consideration of the embodiment issue leads directly to two related problems, viz. embeddedness and hermeneutics. The link between the two is the notion of truth: Computationalism assumes an axiomatic variant of the coherence theory of truth whereas *hylomorphic* (Cariani,91) or hybrid computationalist-physicalist schemes assume some variant of the correspondence theory (chapter 3). Moreno et al. (1994) also hold that

> any formal or computational model has to code directly or indirectly all the information that specifies the behaviour of the system. Therefore, if a given material structure has in-built information we should go down into the lower level of specification until the properties only derive from the computational primitives present in the model. (p.408)

However, is this *possible* ? What level of analysis is necessary or/and sufficient for capturing the *behavioural* dynamics of a living system ? If Emmeche (1992) is correct, any such attempts at decomposition and synthesis will of necessity be observationally-relativistic (chapter 3). Alternatively, it might be the case that *all* levels are important,

---

57 It should be appreciated, however, that issues associated with embodiment and interpretation are also linked.

58 For example, metabolic pathways within cells, generation of three-dimensional structures during DNA and protein folding etc.

59 It is important to include this condition because it has not yet been established whether or not matter can be redefined in computational terms (section 4.7.6).

that is, the organization must be `correct' all the way down, in which case functionalism would be incorrect. Randall (1996) has argued that on a Platonistic view, matter ultimately reduces to a duality of form and void; hence, any properties associated with matter must ultimately derive from a particular type of form or organization (chapter 2). Thus, a formalist decomposition is logically possible assuming Platonism; however, this approach can be contested on (at least) two grounds: (i) *epistemic* - it is impossible to ensure that a lowest level of analysis has been attained. That is, it is impossible to be certain that the behavioural atoms or primitives in a system *are* primitive; (ii) *ontic* - forms are *static* and hence, cannot account for the dynamics of living systems (Zeno's paradoxes of motion). A formalist decomposition is problematic since it necessitates a non-formalist substrate in which to be realized (or actualized)[60]. Software requires hardware on which to execute; hence, the elements in the software-hardware duality, although *functionally*-interchangeable, cannot be reduced to a monism of either component. Emmeche (1991) supports this critique maintaining that

> the `reproduced' entities [in computational systems] do not really as intended contain all the information needed for determining the process of reproduction. From a purely formal point of view this may be the case, but *the physical machine that realizes the process, and which is not reproduced, supports the embedding universe of the reproducing automata and acts as a co-determiner of the process, but is itself not determined by it* [emphasis added]. (p.85)

Appeals to virtual machine hierarchies (chapters 2 and 5) do not resolve the problem, according to Emmeche, since "there is still an additional external machine whose determination does not depend on the process of reproduction." (p.85) As Moreno et al. (1994) state, "(1) there is a deep entanglement between logical form or `software' and material structure or `hardware' (part of the information is implicit in the structure of the components) and (2) an independence between organization and structure would require to make explicit the information that specifies the organization." (p.409) On this basis, Emmeche (1991) maintains that "in an autonomous living system, we cannot make the distinction between the entity being reproduced and an ultimate machine whose properties do not depend on the process of reproduction and which is not reproduced itself." (p.85) However, this view is problematic since a distinction *is* made between the dynamic reproducing entity and the (assumed) static physico-chemical natural laws which provide the ontological substrate in which reproduction is realized. In short, if the natural laws can be interpreted computationally, this objection to computationalism is untenable. This possibility is examined briefly in section 4.7 and more fully in chapter 5.

4.6.9.3. Embeddedness

Emmeche (1991) maintains that "it is the intrinsic and causal property of the biosemiotics [or biological signifying capacities] of the cell that explains why real self-

60 This argument will be explored in further detail in chapter 7.

reproduction is complete, while modeled self-reproduction involves external sign-relations between the observer and the system modeled." (pp.85-86) A distinction between the two kinds of self-reproduction can be made because the human being stands in a different set of ontic and epistemic relations with respect to each type of system, viz. physical and computational. In the case of computational artifacts, a `God's Eye' or exosystemic view is possible; that is, a correspondence relation can be established between artificial (as artifactual) and natural phenomena. However, if the interpreter-producer is *embedded* within the system (Rasmussen,91), it becomes impossible to establish a relation of correspondence between artificial (as artifactual) and natural phenomena since the *only* phenomena *accessible* to the observer on this view are `natural' phenomena, that is, phenomena embedded within the observer's world. Hence, it can be argued that self-reproduction in a computational universe *is* complete *from an embedded (or endosystemic) perspective.* However, as stated in section 4.3.4, the necessary and sufficient conditions for endosystemicity are a matter of dispute.

Davidge (1992) maintains that the organism-environment distinction is a necessary property of living systems. According to Emmeche (1994), however,

> it is not clear what an environment of a `program-organism' is. The interface between a cell and its environment is *spatially* well-defined; this is not so for the abstract life in a computational model [emphasis added]. (p.4)

Yet this criticism of computational A-Life is problematic: For example, if an organism is defined as a dynamic pattern of cells in particular states in a cellular automaton (CA), then the environment can be defined as the states of all other cells which are not included in the pattern. For example, if a glider (chapters 2 and 5) is identified as an `organism'[61], the pattern produced by the states of all cells not included in the glider pattern could be taken as constituting its `environment'. It is important to appreciate that on this view, it is not the cells in the CA which are differentiable into organism and environment but the dynamic pattern of cell states. Cariani (1989, 1991) has criticized this view on the grounds that the distinction is `in the eye of the beholder', that is, observationally-relativistic; epistemological as opposed to ontical. However, if, as Fredkin (1990) maintains, computational atomism is correct with respect to the ontology of the world, then the organism-environment duality is ultimately grounded in a discrete, yet connected substrate; consequently, the organism-environment duality would indeed be merely epistemological, a consequence of an arbitrary `cut' of the dynamics of the computational substrate into patterns classified as organisms and environments. This position is supported by Helmreich (1992) who maintains, on the basis of an *autopoietic* conception of living systems (chapter 6), that "organisms are always already part of their environment." (p.389) However, Helmreich (1994) is critical of computational A-Life because invariably

---

61 This is a simplification for the purposes of argument; gliders do not self-replicate and hence, are not living on an essentialist account of life.

> the individual in this formulation is seen as ontologically prior to its environment, prior to its resources, and prior to other individuals. (p.22)

This leads directly to a consideration of the hermeneutic contextuality of organisms.

4.6.9.4. Hermeneutics

Helmreich (1992) maintains that "our abstractions of `life's formal properties' are inescapably grounded in an historical, epistemological, and cultural context. In particular, I maintain that religious, philosophical, and political systems of knowledge strongly guide the way we think about life, artificial or otherwise." (p.385) Consequently, he is extremely critical of the computationalist approach to A-Life, viz. "language can never give us an unmediated description of the world, and attempts to formalize language as a kind of calculus neglect its essentially hermeneutic nature." (p.386) Helmreich (1994) contests the validity of the functionalist assumption implied in Langton's locating *life-as-we-know-it* within a larger conception of *life-as-it-could-be* which involves (i) a movement towards the adoption of a modal logic-based view of life, viz. functionalistic-essentialism (section 4.6.8), and (ii) the bifurcation of Being as *physis* into potentiality and actuality (chapters 1 and 6). He maintains that

> a theory that pretends to such a transcendent position cannot help but start with a biology that we already `know' through socially conditioned language and traditions. (pp.2-3)

> The belief that organisms are usefully thought of as programs (said more carefully, that genotypes are programs and phenotypes are the processes set into motion by those programs) makes it plausible to think of programs as varieties of organisms. And once organismic identity is flattened out like this, so is the definition of life, such that digital organisms suddenly come alive. (p.12)

This position is supported by Emmeche (1994) from a semiotic perspective, viz.

> one fundamental problem common for all criteria when used in the context of computational `strong A-life' is, that *they are really not criteria for life in the usual biological sense,* but that *they already represent another concept of life, namely life as an abstract, non-material phenomenon*, and thus their relevance as a kind of `conceptual anchor cable' to the physical world of known plants and animals is dubious [emphasis added]. (p.5)

Consequently, he is led to assert that

> one could be tempted to say that what is being studied in Artificial life .. is not even life as an abstract phenomenon, it is the life of abstract concepts ascribed to a specific interpretation of formal computational structures. (p.5)

However, there is a problem with both Helmreich's and Emmeche's positions in that their criticisms against computationalism can be directed against *systemic* (more specifically, cybernetic) approaches in general; thus, Maturana and Varela's (1980) autopoietic system conception, which Helmreich identifies as an *alternative* to computationalism, may itself suffer from problems similar to those associated with the latter (chapter 6).

4.6.9.5. Morphization

If functionalism is problematic for the above stated reasons, then why does it continue to exert such a pull on the scientific and philosophical imagination ? Why is computationalism so readily embraced ? Perhaps an example can help to answer this question. According to Simons (1983),

> a system may be said to be `feeding' when it systematically extracts energy from an appropriate source in order to support its internal life processes and its behaviour in the world. (p.6)

On the basis of this definition, W.Grey Walter's (1950) electronic `tortoises', whose behaviour was defined as (1) move towards dim light, (2) move away from bright light, (3) recharge from a brightly lit `kennel' when batteries run low, would be defined as living. However, automatically-steered cars capable of extracting fuel from gas stations (and perhaps other cars) would also have to be identified as living according to this definition. But would *we* be prepared to define cars as living ? Refusal to do so could be criticized on the grounds that the tortoises had been accepted because of *anthropomorphism* while the cars had been rejected because of *anthropocentrism*. However, the *a posteriori* nature of car teleology with respect to human beings (chapter 7) blocks their interpretation in vital terms; whether or not functionalism in the context of *in silico* or computational life is valid is an open issue since it is unclear at this stage whether computational systems can generate teleology through some kind of autonomous self-organizing processes.

Functionalism supports *morphization* (chapter 3) or abstraction across categories such that entities belonging to different classes can be *re-identified* as members or instances (tokens) of a common subsuming class or category (type). What is often overlooked, however, is *who* it is that is doing the abstraction; the human being (or group of human beings) responsible for morphization is (are) almost invariably left out of the picture, thereby leading to a realist-objectivist position with respect to the categories. As an example of such a position, consider the following statement due to Simons (1983), viz. "computers do not need to be carbon-based in order to do sums or to take decisions." (p.9) The implication is that if arithmetic is multiply-realizable and life is also multiply-realizable, there is no reason in principle why life cannot be realized in a computational substrate since arithmetic can be realized in a computational substrate. The problem with this view is that it is not clear whether computation *is* multiply-realizable. It may well be the case, as Searle (1992), Kelly (1993), Tallis (1994) and Lanier (1995b) have argued, that computation necessitates intentionality and that syntactic information processing is non-intentional *if the human interpreter is excluded from the process*. On this basis, computational life will be an instance of life only if the human interpreter reckons it as such[62].

4.6.10. Connections Between Life and Matter

Throughout the discussion in section 4.6, it has been maintained that in order for life to be explained, this must be done with reference to matter. Simon (1971) offers the following remarks in support of an emergentist position with respect to life:

> We discern two separate tendencies among biologists with respect to the concept of life. These are revealed in the difference between speaking of the 'characteristics' of life, and speaking of the 'building-blocks' of life. [However,] if life is to be defined or explained, as opposed to being described, it will have to be in terms of what is not alive; otherwise it must be regarded as an unexplained primitive. (p.193)
>
> The expression 'living matter' is to be understood as signifying, not a special type of matter, or even ordinary matter specially arranged, but matter which happens to be situated in such a way as to bear a certain relation to a living organism. It is in this sense, and this sense alone, that it can be said that the organism confers life on matter, rather than the other way around. (p.195)

In the following section, the nature of matter, which on the emergentist framework provides the ontological substrate for life, will be investigated.

4.7. Matter

In this section, the concept of matter is investigated. The presentation has two main objectives: (1) identify the main differences between the classical and modern conceptions of matter; (2) outline a computational theory of matter (CTMa) which incorporates the defining characteristics associated with the information-theoretical approach to physics. The treatment in this section is necessarily brief and based on key texts such as Sir Arthur Stanley Eddington's *The Nature of the Physical World* (1928) and various works by P.C.W.Davies (1987, 1989, 1991). A more precise formulation of the computational theory of matter (CTMa) will be presented in chapter 5 when an implementation of computationalism based on the cellular automaton (CA) formalism is described in the context of a unifying framework of emergent artificiality.

4.7.1. What is *Matter* ?

As with mind and life, the concept of matter is somewhat difficult to define. Collingwood (1945) presents a historical overview of the concept in connection with a broader analysis of the the idea of nature and the subject is dealt with explicitly in (Toulmin,62). In many of the philosophical dictionaries, matter is defined negatively by

---

62 On this view, the ontological status of computational life is observationally-relativistic, that is, epistemically-grounded.

placing it in opposition to the concept of mind. This Cartesian *dualist* approach is based on the assertion that there are two types of `stuff' in the world, viz. (1) *res extensa* (spatially-extended stuff) and (2) *res cogitans* (mental stuff). However, this definitional approach becomes problematic on a *materialist* metaphysics since substance-dualism must be rejected and hence, the mind *reduced* to matter. An alternative formulation in *monistic* terms can be traced to the pre-Socratic Greeks who defined matter in terms of a *conservation principle*, viz. matter is the fundamental stuff of existence which is *preserved* during any process of physical change. (This idea finds modern expression in the First Law of Thermodynamics, viz. in a closed system, matter-energy cannot be created or destroyed.) The Greeks were also responsible for the idea that all matter is built out of a small number of basic units, the original *atoms* (chapter 2). However, in this century, as a consequence of the work of Rutherford, it has been empirically demonstrated that atoms are not elementary particles at all, but composite structures with internal parts. Further investigation along this line of inquiry has revealed numerous levels of intra-atomic structure leading to the discovery of a `particle zoo'. For example, atoms have been decomposed into three broad species of particles: (1) *bosons* which include gluons, photons, W (weak) particles, Z particles and the Higgs boson; (2) *quarks* (up, down, charm, strange, top, bottom); and (3) *leptons* (electrons, electron-neutrinos, muons, muon-neutrinos, tauons, tauon-neutrinos).

Following Einstein's discovery of the interconvertibility of matter and energy and diffraction experiments which revealed wave-particle complementarity, matter has come to be discussed in terms of energy, particles, waves, and fields of force. The interrelation between these concepts is dense: For example, Davies (1987) maintains that "the higher the energy, the less structure and differentiation there is both in subatomic matter itself and the forces that act upon it." (p.124) Consequently, matter is able to assume a variety of forms. As temperatures are increased, matter passes through the familiar solid, liquid, gaseous and plasma phases. In the last of these phases even atoms lose their structure, becoming dissociated into electrons and ions. However, even this picture may be incomplete. As Davies states,

> if some very recent ideas are to be believed, as the temperature reaches the so-called Planck value of 10^{32} degrees, all matter is dissolved into its most primitive constituents, which may be simply a sea of identical strings existing in a ten-dimensional spacetime. Moreover, under these extreme conditions, even the distinction between spacetime and matter becomes nebulous. (p.125)

Kaku (1997) describes the essence of *superstring theory* as follows:

> In superstring theory, the subatomic particles we see in nature are nothing more than different resonances of the vibrating superstrings, in the same way that different musical notes emanate from the different modes of vibration of a violin string .. Likewise, the laws of physics - the forces between charged particles, for example - are the harmonies of the strings; the Universe is a symphony of vibrating strings. And when strings move in 10-dimensional space-time, they warp the space-time surrounding them in precisely the way predicted by general relativity. So strings simply and elegantly unify the quantum theory of particles and general relativity. (p.34)

According to Eddington (1928), "the whole trend of modern scientific views is to break down the separate categories of `things', `influences', `forms', etc., and to substitute a common background of all experience." (p.7) This breakdown is facilitated via the *measurement* process and, as Eddington states, on the scientific view "measures themselves afford no ground for a classification by categories" (p.7) although the prior existence of an external world accessible via the measures is a necessity[63].

4.7.2. Classical Physics

The classical or Newtonian concept of matter is grounded in two metaphysical assumptions, viz. atomism and determinism, which together provide the basis for the idea of *mechanism* (chapter 2). According to Davies (1983), atomism is characterized by the belief that there exist "a small number of truly elementary particles which have no internal parts and which are the building blocks of all matter." (p.48) Moreover, as Davies and Gribbin (1992) state,

> like the Greek Atomists before him, Newton treated matter as passive and inert. Indeed, *inertia* played a central role in his theory of the world. If a material body is at rest, then according to Newton's laws it will remain forever at rest unless acted upon by an external force. Similarly, if the body is moving, it will continue to move with the same speed and in the same direction unless a force acts to change it. Thus matter is entirely passive. (p.5)

Newtonian physics assumes a *substantialist* metaphysics in contrast to a processual ontology (chapter 2). The notion of a `substance' has been referred to repeatedly throughout this thesis, specifically in connection with the essentialist interpretation of phenomena. But what *is* a substance ? According to Cobb (1988),

> a *substance* is that which depends on nothing else for its existence. It is a thing that remains fundamentally the same regardless of its relations. An atom was defined by the Greeks to be a unit of substance. Modern mechanism is built on this notion. Everything that is not an atom is nothing but a structure of atoms. The atoms are not affected by the structures in which they are arranged. The structures behave like machines and are not inherently affected by their relations to other things. They can be externally affected by other things by having some of their parts separated from others, but the character of the separated parts is not affected by this separation. (p.107)

Bohm (1988) maintains that "although the more recent physics has dissolved the mechanistic view [it] is still the dominant view as far as effectiveness is concerned." (p.60) He identifies three postulates associated with mechanism, viz.

1. The world is reducible to a set of basic elements.
2. These elements are external to each other.
3. Interactions do not affect the internal nature of the elements.

---

63 In this sense, science (specifically, physics) is committed to some version of metaphysical realism.

Postulate (3) leads to a consideration of the interaction dynamics of classical systems. Newtonian mechanics is based on a specific form of determinism, viz. "billiard-ball" Laplacianism in which interactions between elements are local and take the form of pushes and pulls - *action by contact* (chapter 2). Furthermore, as Prigogine et al. (1984) maintain, Newtonian science is based on the assumption that "at some level *the world is simple* and is governed by time-reversible fundamental laws." (p.7) This view has been contested as follows:

> the *artificial* may be deterministic and reversible. The *natural* contains essential elements of randomness and irreversibility. This leads to a new view of matter in which matter is no longer the passive substance described in the mechanistic worldview but is associated with spontaneous activity [emphasis added]. (p.9)

However, under computationalism it may be possible to reconcile these positions since certain types of non-linear dynamical systems (chapter 2) whose behaviour at the local (component) level is deterministic are capable of generating `chaos', that is, unpredictable or non-deterministic behaviour at the global (systemic) level. Although the macroscopic behaviour can be generated deterministically, this behaviour is *non-computable*[64] (chapter 2). Because it is possible for a system to be both deterministic (locally) and non-computable (globally), it is possible to reconcile chance and necessity. Consequently, `deterministic randomness' (Davies,87) could provide the conceptual bridge between classical (reversible) and non-classical (irreversible) phenomena. This possibility, which is embedded in the connection between thermodynamics, information theory and computation theory, is implicit in the computational theory of matter presented in 4.7.6.

In summary, the defining concepts of classical physics (mechanism) are atomism and reversibility as described above; additionally, there is the postulate of *absolute* space and time. Many ideas associated with classical physics have been retained in modern physics, for example, notions such as *particles* and *laws*[65], as evidenced by the discussion in section 4.7.1. However, a number of recent developments within physics, in particular those associated with the relativity and quantum theories, have necessitated reconsidering the basic assumptions underlying the concept of matter.

4.7.3. The Theory of Relativity

The first major challenge to classical physics during this century was associated with the introduction of the special and general theories of relativity: The former was responsible for replacing absolute space and time with the unified relativistic notion of four-dimensional spacetime while the latter provided an interpretation of gravity in

[64] Hence, the crucial distinction between *computationality* (ontology) and *computability* (epistemology).

[65] Nagel (1961) provides a detailed account of the logical structure of physical laws.

geometrical terms, specifically in terms of the geometry of the spacetime manifold.

4.7.3.1. The Special Theory of Relativity

According to Bohm (1980), the principle difference between the relativistic and the Newtonian conceptions of the world is that

> relativity introduces new notions concerning the order and measure of time. These are no longer *absolute*, as was the case in Newtonian theory. Rather, they are now *relative* to the speed of a coordinate frame. (p.123)

Spacetime frames are not intrinsic to the universe, but to observers *within* the universe (Eddington,28); consequently, from an *endophysical* perspective, there is no such thing as *the* spacetime framework of the universe. On Einstein's model, the three spatial dimensions and single temporal dimension of the world are unified into a single structure, the four-dimensional spacetime *block universe* which has the status of a property-less extension until a *metric* is imposed on it. Rucker (1985) describes the distinction between the Newtonian model and the Einsteinian model as follows: In the former, "space is made up of *locations* [while in the latter] spacetime is made up of *events*. An `event' is just what it sounds like: a given place at a given time." (p.137) Consequently,

> in the block universe there is no objectively existing `Now' [since relativity theory implies that it is impossible to permanently mark a given space location from within the block universe]. Nothing is moving in the block universe .. (p.149)

Wolf (1991) describes the special theory of relativity as

> a set of rules that enable an observer to calculate what another observer sees when he is moving at a fixed velocity past the first observer (p.327)

and Davies (1991) presents a similar formulation, viz. "the simultaneity of events that are separated in space is *relative*. Different observers in different states of motion measure different durations [and distances] between the same pair of events." (p.70) Significantly, although all observation and interaction in the world is relativistic, the underlying laws governing such interactions are absolute. As Rucker (1985) states,

> the laws by which the states of physical systems undergo change are not affected, whether these changes of state be referred to the one or the other of two systems of coordinates in uniform translatory motion. (p.150)

The special theory of relativity is based on two assumptions:

1. The speed of light is constant.
2. Absolute motion is undetectable.

The speed of light defines an upper bound for the speed of propagation of a signal, whereby a *signal* is meant a physical, information bearing entity such as an electron or photon. This is readily apparent when the notion of a *light cone*, viz. the volume traced out in spacetime from a source of light, is investigated: Two spacetime points which lie within each other's light cone are causally connected because they can exchange signals and experience each other's influences. Thus, the speed of light constrains the causal relations of physical systems. As Bohm (1980) states,

> if we went faster than light, then, as a simple calculation shows, the electromagnetic fields that hold our atoms together would be left behind us (as the waves produced by an airplane are left behind it when it goes faster than sound). As a result, our atoms would disperse, and we would fall apart. So it would make no sense to suppose that we could go faster than light. (p.122)

According to Eddington (1928), the special theory of relativity is significant because it necessitates distinguishing between two events, viz. "the original event, somewhere out in the external world and .. a second event, viz. the *seeing* by us of the first event." (p.53) However, it has both epistemological and ontological implications: On a materialist (or physicalist) framework, the speed of light is a restriction on both knowing *and* doing, that is observation *and* action, since both are ultimately reducible to the motions of particles (or waves) and conversion of energy from one form to another.

Alexander (1920), whose metaphysics provides the basis for the unified framework of computationally emergent artificiality described in chapter 5, hypothesized an absolute Space-Time maintaining that the implications of relativity theory were not "in any way inconsistent with their being pure events or point-instants which have their `absolute' position in Space-Time." (Vol.I, p.89) He further clarified this position by stating that

> whatever modifications it introduces into the Newtonian mechanics it leaves Time and Space and Motion in their ancient reality, or rather it leaves us still with Space-Time in itself as a total from which perspectives are selections; and therefore in that sense absolute and independent of observers. (Vol.I, p.91)

However, this view is correct only with respect to *exophysical* observation, that is, with respect to an observer capable of a `God's Eye' view of the universe.

4.7.3.2. The General Theory of Relativity

Wolf (1991) describes general relativity as

> the theory of the universe that explains the presence of gravity as the distortion of space and time together. If a spacetime distortion is present, there must be matter. (p.327)

The general theory of relativity incorporates a theory of gravity entailing the view that

1. matter and energy distort space
2. distortions of space affect the motions of matter and energy.

As Rucker (1985) states, "space .. serves as the medium for transmitting gravitational effects. Mass affects space, space affects mass." (p.79) Force-fields are both continuous and local on Einstein's general theory of relativity. As Whittaker states in the introductory note to Eddington's *The Nature of The Physical World* (1928), "gravity is not a force acting at a distance, but an effect due to a modification of space in the immediate neighbourhood: secondly, it is propogated from point to point of space, being ultimately connected with the presence of material bodies." (p.viii) Fields of force are represented as structural properties of four-dimensional space-time; moreover, fields self-organize into material particles (Harris,65). Four kinds of force, viz. electromagnetic, gravitational, strong nuclear and weak nuclear, have been identified. However, at high energies, the electromagnetic and weak nuclear forces combine into the `electroweak' force; thus the four forces are in fact reducible to three forces and it is conjectured that at higher energies unification of the electroweak with the strong nuclear force and ultimately with the gravitational force may be possible, leading to a unified force field.

4.7.3.3. Beyond Relativity Theory

Bohm (1988) maintains that relativity theory transcends classical mechanism in that

> instead of having separate little particles as the constituents of matter, Einstein thought of a field spread through all space, which would have strong and weak regions. Some strong regions, which are stable, represent particles. (p.62)

However, a retroactive link between relativity theory and mechanism is identified, viz.

> relativity theory retains certain essential features of mechanism, in that the fields at different points in space were thought to exist separately and not to be internally related. The separate existence of these basic elements was emphasized by the idea that they were only *locally connected*, that the field at one point could affect a field only infinitesimally nearby [emphasis added]. (p.63)

Furthermore, although radically post-Newtonian in rejecting the idea of an objective, observer-independent reality, the theory of relativity retains a Newtonian perspective with regard to the nature of material objects themselves. As Casti (1992) states,

> on matters pertaining to the static and dynamic attributes of Newton's particles - e.g. mass, electric charge, velocity, spin - relativity theory is silent or, more accurately, tacitly accepts the Newtonian precepts .. Instead Einstein's theories focus upon the other half of the Newtonian doublet, the unexplained forces (particularly gravity) (p.419)

4.7.4. Quantum Theory

Bohm (1988) identifies four postulates as defining quantum theory, viz.

1. All action or motion is traceable to a discrete indivisible unit called a quantum.

2. Matter and energy have a dual nature; depending on context, things can manifest as waves or particles.

3. Non-local connection between things.

4. Holistic organization of parts.

On the quantum theoretical view, every physical situation is characterized by a *wave function*. However, as Bohm (1980) states, "this is not directly related to the *actual* properties of an individual object, event, or process. Rather, it has to be thought of as a description of the *potentialities* within the physical situation." (pp.128-129). This leads Davies et al. (1992) to assert that

> we live not in a cosmic clockwork, but in a cosmic network, a network of forces and fields, of nonlocal quantum connections and nonlinear, creative matter. (p.11)

However, "there remains a sense in which quantum mechanics is still a deterministic theory. Although the outcome of a particular quantum process might be undetermined, the *relative probabilities* of different outcomes evolve in a deterministic manner. [Hence,] as a *statistical* theory, quantum mechanics remains deterministic .. Quantum mechanics builds chance into the very fabric of reality, but a vestige of the Newtonian-Laplacian world view remains." (p.27) According to Penrose (1989),

> probabilities do *not* arise at the minute quantum level of particles, atoms, or molecules - those evolve *deterministically* - but, seemingly, via some mysterious larger-scale action connected with the emergence of a classical world that we can consciously perceive. (p.292)

Classical systems evolve in a *locally* deterministic and reversible manner; quantum systems, on the other hand, evolve in a way that is *globally* (statistically) deterministic and reversible. Irreversibility, that is, an `arrow of time' (section 4.7.5), arises in both systems as a consequence of *measurement* (although the nature of the measurement operation is different in each case). The essential distinctions between relativity theory and quantum theory are summarized in Table 4.4:

Relativity Theory	*Quantum Theory*
continuous	discrete
deterministic	nondeterministic
local	non-local

Table 4.4 Basic distinctions between relativistic and quantum systems.

A number of interesting properties are associated with the behaviour of systems at the quantum level, a few of which are briefly discussed in the following sections.

4.7.4.1.The Double-Slit Experiment

In Young's double-slit experiment, light from a point source passing through two nearby slits produces an interference pattern on a screen thereby demonstrating the wavelike nature of light. If one of the slits is blocked, the interference pattern disappears. However, if the source is modified so that only a single photon (unit of light energy) passes through the slit system, the pattern is repeated. Given that a single photon (particle) can only pass through one slit, it must somehow `know' of the existence of the other slit such that it can `decide' where it belongs in the interference pattern that is built up from the millions of particles passing through the slits which hit the screen. Davies (1992) maintains that Young's experiment provides "evidence for the holistic nature of quantum systems, with the behaviour of individual particles being shaped into a pattern by something which cannot be explained in terms of the Newtonian reductionist paradigm." (p.205)

4.7.4.2. The Uncertainty Principle

The Heisenberg uncertainty principle states that all observable microscopic quantities are subject to random fluctuations in their values such that these can only be described probabilistically. According to Davies (1992),

> you cannot know, at any instant, *both* the position and the momentum of a quantum particle. Indeed, [the principle] goes deeper - it says that a quantum particle *does not possess* both a definite momentum and a definite position simultaneously. (p.213)

Heisenberg's uncertainty principle involves the production of epistemological constraint as a consequence of ontology: It does not merely state that human observational capacities are limited; rather it states that these capacities are limited *because* reality is intrinsically indeterministic. This led Bohr to formulate his famous *complementarity principle*. Davies describes the link between the Heisenberg uncertainty principle and Bohr's notion of complementarity as follows:

> the trade-off between position and momentum [which are conjugate attributes of a particle] is another example of quantum complementarity at work. It turns out to bear a close relation to the wave-particle complmentarity. The wave associated with an electron is, by its very nature, a spread-out thing, and does not have a definite position, although it does encode information about the electron's momentum. By contrast, the particle associated with an electron is, by its very nature, something with a well-defined position; but a wave collapsed to a point carries no information about the momentum of the electron. Measure the position of an electron, and you do not know (*nor does the electron know*) how it is moving; measure the momentum of an electron, and neither you nor the electron know where it is located. (p.214)

Furthermore,

> the fact that electron waves are waves of *probability* is a vital component of quantum mechanics and in the quantum nature of reality. It implies that we cannot be *certain* what any given electron will do.

> Only the betting odds can be given. This fundamental limitation represents a breakdown of determinism in nature. It means that identical electrons in identical experiments may do different things. There is thus an intrinsic uncertainty in the subatomic world. (p.202)

4.7.4.3. The Einstein-Podolsky-Rosen (EPR) Experiment

Penrose (1989) briefly describes a variant of the EPR experiment as follows:

> Suppose that two spin-one-half particles - which I shall call an *electron* and a *positron* (i.e. an *anti-electron*) - are produced by the decay of a single spin-zero particle at some central point, and that the two move directly outwards in opposite directions. By conservation of angular momentum, the spins of the electron and positron must add up to zero, since that was the angular momentum of the initial central particle. This has the implication that when we measure the spin of the electron in some direction, whatever direction we choose, the positron now spins in the *opposite* direction! The two particles could be miles or even light-years apart, yet that very *choice* of measurement on one particle seems *instantaneously* to have fixed the axis of spin of the other! (p.365).

The results of the experiment are interpreted by Davies (1992) as follows:

> assuming one rules out faster-than-light signalling, it implies that once two particles have interacted with one another they remain linked in some way, effectively parts of the same indivisible system. This property of `nonlocality' has sweeping implications. We can think of the Universe as a vast network of interacting particles, and each linkage binds the participating particles into a single quantum system. In some sense the entire Universe can be regarded as a single quantum system. (p.217)

4.7.4.4. The Measurement Problem and The Interpretation Problem

Casti (1989) defines the *measurement problem* as "the question of how and when the act of measurement `collapses' the wave function" (p.440) and the *interpretation problem* as "determination of the nature of a quantum object when it is in its unmeasured state" (p.440). Various solutions to both problems have been proposed, a number of which are summarized in Table 4.5:

School	*Wave function collapse (measurement)*	*Unmeasured Attributes (interpretation)*
No objective reality exists		
Copenhagen (Bohr)	by measuring device	do not exist
Consciousness (Schrodinger, Von Neumann, Wigner)	by conscious mind	do not exist
Austin (Wheeler)	from communication	created by meter option
Duplex (Heisenberg)	from measurement act	only phenomena are real

Many Worlds (Everett, Deutsch)	no collapse	all possibilities are real
Objective reality exists		
Naive Realist (Einstein)	no	always exist
Quantum Logic (Von Neumann)	(not addressed)	always exist
Quantum Potential (Bohm)	no	always exist
Transactional (Cramer)	yes	always exist

Table 4.5 Solutions to the quantum measurement and interpretation problems.

4.7.4.5. The `Hidden Variables' Interpretation

This is based on postulating the existence of variables hidden from observation which take the form of quantum probabilities. The values of these variables are unknown prior to measurement, values which, if known, would account for measurement uncertainty. Penrose (1989) maintains that a hidden-variable theory would be consistent with all the observational facts of quantum physics *if* the theory supported *non-locality*: Hidden parameters must be able to instantaneously affect parts of the system in arbitrarily distant regions (as happens in the EPR experiment). However, implementing non-locality leads to problems with special relativity since field interactions are local on the latter. (Relativity theory and quantum theory are, as a result, apparently incommensurable physical frameworks.) The hidden variables or `naive realist' position is important because, on certain interpretations, it can be shown to be consistent with classical mechanics in postulating *local* determinism, viz. "God does not play dice". For this reason, it becomes the natural interpretation of quantum theory in the context of the realization of quantum phenomena in deterministic formalisms such as cellular automata (chapter 5). However, an alternative, the *observer-participant* interpretation, has also been advanced in the context of computationalism (section 4.7.6).

4.7.5. Thermodynamics

Although the major developments in twentieth century physics, viz. relativity theory and quantum theory, are extremely important, it is to the nineteenth century laws of thermodynamics that the origin of the computational theory of matter (CTMa) can be traced. Thermodynamics becomes important when the temporal evolution of a system is considered. The existence of an `arrow of time' is a basic fact of everyday experience. However, its explanation is another thing entirely. For example, Eddington (1928) considers the possibility that "we might appeal to consciousness to suffuse the whole - to turn *existence* into *happening*, *being* into *becoming*." (p.76) However, he maintains that "without any mystic appeal to consciousness it is possible to find a direction of time on the four-dimensional map by a study of organization." (p.76) This leads directly to the second law of thermodynamics and the mathematical concept of entropy.

Briefly, the Second Law of Thermodynamics states that the *entropy* or degree of disorder (randomness) in a closed system always increases with time. The law is statistical in the sense that it applies to a group of individuals (for example, physical particles) as contrasted with more basic physical laws (for example, Newtonian mechanics) which apply to the individuals themselves. While the Second Law necessitates the reversibility (time-symmetry) of the underlying laws (irrespective of whether they are classical, relativistic or quantum), it itself introduces an irreversibility or `arrow of time' into the description of a system at the group or statistical level. This follows directly from (1) the act of *measurement* in which a `coarse grain' statistical view of the system at the global (or group) level is produced and (2) the *non-linear* nature of the interaction dynamics at the local (or individual) level. Eddington (1928) linked the Second Law of Thermodynamics to the study of *organization* and this connection, under its modern information-theoretical interpretation, has been used in formulating computational theories of natural phenomena. (The application of thermodynamic and information-theoretical concepts to biology was discussed in section 4.6.2.)

Heisenberg (1979) described thermodynamics as a `bridge' between classical and quantum physics by virtue of its concern with observation and measurement, viz.

> thermodynamics leaves classical physics and goes into the region of quantum theory, for it speaks about situations of observation; it does not speak about the system as it is, but about the system in a certain state of being observed, namely in the state of temperature equilibrium. (pp.11-12)

It is important to note that the Laws of Thermodynamics are instances of what Eddington has called `secondary laws'. As he states,

> I have called the laws controlling the behaviour of single individuals `primary laws', implying that the second law of thermodynamics, although a recognized law of Nature, is in some sense a secondary law .. Some things never happen in the physical world because they are impossible; others because they are too improbable. The laws which forbid the first are the primary laws; the laws which forbid the second are the secondary laws. (p.82)

He further maintains that,

> secondary law is not in conflict with primary law, nor can we regard it as essential to complete a scheme of law already complete in itself. It results from a different (and rather more practical) conception of the aim of our traffic with the secrets of Nature. (p.83)

4.7.6. A Computational Theory of Matter (CTMa)

Davies and Barrow (1992) maintain that the classical materialistic conception of nature has given way to an informational or computational conception. On their view,

> matter as such has been demoted from its central role, to be replaced by concepts such as organization, complexity and information. (p.9)

The link between thermodynamics and the computational theory of matter (CTMa) is the mathematical concept of *information* due to Shannon and Weaver (1949). The information content or *negative entropy* associated with a message is proportional to the amount of prior uncertainty it resolves; uncertainty can itself be quantified in relation to the number of possible states of affairs that might be the case. The information content (in binary digits or bits) of a discrete process A with K states, each state having probability p_i is defined as follows:

$$H(A) = -\sum_{i=1}^{K} p_i \log_2 p_i \qquad (4.1)$$

The amount of information carried by any message will be determined by the number of possible alternatives that have been selected from and the relative prior probabilities of the different messages. However, the above formulation of information in probabilistic terms does not map easily onto the idea of information as conceived in in a computational context. As Zurek (1989) states,

> the information which is being processed by the computer is a concrete `record', a definite sequence of symbols. Its information content cannot be represented adequately in terms of Shannon's probabilistic definition of information. One must instead quantify the information content of the specific, well-known `record' in the memory of the computer - and not its probability of frequency of occurrence, as Shannon's formalism would demand. (p.ix)

This leads to the notion of *algorithmic* information content as independently developed by Kolmogorov, Solomonoff and Chaitin (1990). The algorithmic information content of a physical entity is given by the size, in bits, of the most concise message (for example, the shortest Turing machine program that can execute on a UTM) which describes that entity with the requisite accuracy. Regular systems can be specified by means of concise descriptions; hence, algorithmic information content can be regarded as a measure of disorder (or irregularity). Bennett (1982) has shown that the average algorithmic entropy of a thermodynamic ensemble has the same value as its statistical (ensemble) entropy. Consequently, it is at least conceivable that a consistent thermodynamics could be built on an algorithmic foundation. Furthermore, and most importantly, attempts have been made to map the relativity and quantum theories - Eddington's `primary laws' - onto a computational framework grounded in algorithmic information theory (Zurek,89) which has clear links to secondary laws (thermodynamics) as stated above.

4.7.6.1. Thermodynamics and Self-Organization

One of the major problems with the above approach is reconciling thermodynamic laws describing the increase in disorder or randomness of a system (entropy) with what Davies (1987) has referred to as "laws of organization" which describe an increase in information (decrease in entropy). On his view,

> organization is a quality that is most distinctive when it refers to a process rather than a structure .. It might be said that order refers to the *quantity* of information (i.e. negative entropy) in a system, whereas organization refers to the *quality* of information. (pp.75-76)

As Davies (1989) states,

> whereas entropy is a measure of information loss, organization (or depth) refers instead to the *quality* of information. Entropy and depth are not each other's negatives. (p.62)

Depth refers to the amount of "work" or information processing a system has to do in order to reach a particular state. It is crucial to appreciate that both sets of laws are mutually consistent: For example, evolution is marked by an increase in the complexity of living systems[66]. This fact can be reconciled with the Second Law because can living systems are *open* systems exchanging matter and energy with their environments; in their most encompassing sense, these environments are identical with the universe which is held to be a *closed* system. Consequently, *local* evolutionary increases in complexity are offset by *global* increases in the entropy of the universe. However, there is a problem, viz. explaining how non-reversible systems can emerge from reversible systems, an issue which will be discussed further in chapter 5.

4.7.6.2. The Physical Church Turing Thesis (PCTT)

Wolfram (1985) maintains that

> one expects the fact that computers are as powerful in their computational capacities as any physically realisable system can be, so that they can simulate any physical system. (p.785)

If the Physical Church Turing Thesis (PCTT) (chapter 2) is true, then the above statement implies that it will be impossible to distinguish a simulated universe from the genuine article; as Davies (1992) states, the universe becomes its own simulation. This position is advanced by Fredkin (1990) in connection with the "digital mechanics" concept, viz. the universe as a computer, specifically a giant cellular automaton (chapters 2 and 5). Tipler (1994) develops this idea is some detail:

> a perfect simulation [that is, an *emulation*] exists if the physical universe can be put into one-to-one correspondence with some mutually consistent subcollection of mathematical concepts. In this sense of `simulation' the universe can certainly be simulated, because `simulation' then amounts to saying that the universe can be exhaustively `described' in a logically consistent way. Note that `described' does not require that we or any other finite (or infinite) intelligent being can actually find the description. It may be that the actual universe expands into an infinite hierarchy of levels whenever

[66] However, Chaitin (1970) maintains that this is only true when considering the *phylogenetic* hierarchy: In terms of the *ontogeny* of individual organisms, the reverse actually holds, that is, the complexity of the organism (whole) is less than the complexity associated with the organism considered as a simple sum of its parts. Chaitin maintains that this is due to the multifaceted nature of the structural and functional coupling between parts in organisms, a position that is supported by Rosen (1993).

> one tries to to describe it exhaustively. In such a case, it would be impossible to find a Theory of Everything. Nevertheless, it would still be true that a `simulation' in the more general sense existed if each level were in one-to-one correspondence with some mathematical object, and if all levels were mutually consistent (`consistency' meaning that, in the case of a disagreement between levels, there is a rule - itself a mathematical object - for deciding which level is correct). The crucial point of this generalization is to establish that *the actual physical universe is something in the collection of all mathematical objects*. This follows because the universe has a perfect simulation, and we agree to identify the universe with its perfect simulation, that is, with its emulation. Thus, *at the most basic ontological level, the physical universe is a concept* [emphasis added]. (p.209)

Further support for this view is provided by Davies (1989) who maintains that

> the laws of physics define the allowed mechanical operations that occur in the physical universe, and thence the possible activities of a Turing machine. These mechanical operations thus determine which mathematical operations are computable and define for us what might be called simple solvable mathematics (like addition). For some reason, those same laws of physics can be expressed in terms of this simple mathematics. There is thus a self-consistency in that the laws generate the very mathematics that make those laws both computable and simple. (p.66)

He summarizes this position in the following diagram (Fig 4.3):

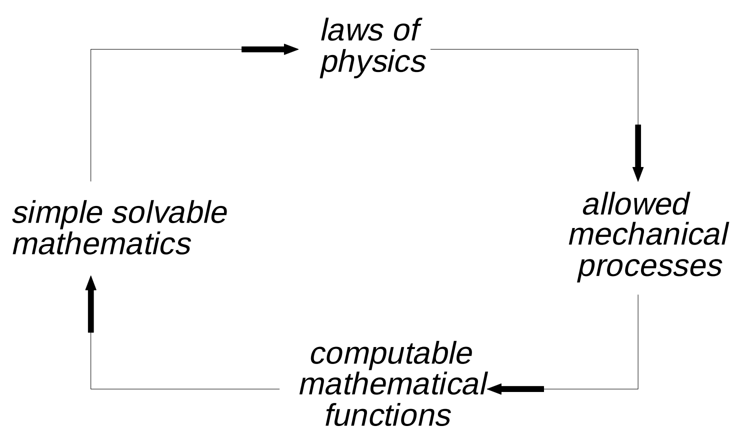

Fig 4.3 Closure of physics and computation.

If this self-consistent loop is to provide the basis for "strong" computationalism, viz. the *ontological* Church Turing Thesis or OCTT (chapter 2), it must account for non-physical phenomena such as mind. The *observer-participant* position proposed by Wheeler in connection with the quantum interpretation and measurement problems (section 4.7.4) provides a suitable framework within which to develop such a scheme.

4.7.6.3. Wheeler's `Meaning Circuit'

D'Espagnat (1981) defines matter as that which is both (1) *conserved* in change and (2) *concrete* in experience. This definition is consistent with a conception of physics in which the act of observation (and thereby, the existence of observers) is fundamental, that is, irreducible. The idea is developed most thoroughly in the *meaning circuit* concept: In Einstein's universe, space, matter (or energy), and fields of force are geometrically defined; for example, gravity is the curvature of spacetime caused by massive centres of attraction. However, Wheeler (1979) maintains the need for a pre-geometrical physics grounded in mathematical logic on self-referential grounds, viz.

> logic is the only branch of mathematics that has the power to think about itself. This magic feature may be the indication that in logic we must look for the branch of mathematics out of which, in some as yet unconceived way, the physical world is somehow constructed (p.57)

This leads Wheeler to propose the `meaning circuit', a *bootstrap* theory in which the universe is conceived as a closed system supporting the evolution of observers responsible for collapsing the quantum wave function describing the universe that brought them into existence. Penrose (1989) describes the meaning circuit as follows:

> the evolution of conscious life on this planet is due to appropriate mutations having taken place at various times. These, presumably, are quantum events, so they would exist only in linearly superposed form until they finally led to the evolution of a conscious being - whose very existence depends upon all the right mutations having `actually' taken place! It is our own presence which, on this view, conjures our past into existence. (p.381)

Wheeler (1989) describes the meaning circuit as a "vision of the world as self-synthesized" in the sense of a "self-referential deductive axiomatic system" (p.9). He elaborates the nature of this scheme as follows:

> No structure, no plan of organization, no framework of ideas underlaid by another structure or level of ideas, underlaid by yet another level, and yet another, *ad infinitum*, down to bottomless blackness. To endlessness no alternative is evident but a loop such as: Physics gives rise to observer-participancy; observer-participancy gives rise to information; and information gives rise to physics. (p.8)

As to the notion of an observer-participant, Wheeler defines it as "one who operates an observing device and participates in the making of meaning." (p.13) With respect to the possible role of consciousness in observation, he offers the following caveat:

> we .. steer clear of the issues connected with `*consciousness*'. The line between the unconscious and the conscious begins to fade in our day as computers evolve and develop - as mathematics has - level upon level upon level of logical structure. We may someday have to enlarge the scope of what we mean by a `who'. (p.15)

On this basis, Wheeler presents a CTMa grounded in the following premises, viz.

1. The world cannot be a giant machine, ruled by any pre-established continuum law.

2. There is no such thing at the microscopic level as space or time or spacetime continuum.

3. The familiar probability function or functional, and wave equation or functional wave equation, of standard quantum theory provide mere continuum idealizations and by reason of this circumstance conceal the information-theoretic source from which they derive.

4. No element in the description of physics shows itself as closer to primordial than the elementary quantum phenomenon, that is, the elementary device-intermediated act of posing a yes-no physical question and eliciting an answer or, in brief, the elementary act of observer-participancy. Otherwise stated, every physical quantity, every it, derives its ultimate significance from bits, binary yes-or-no indications (p.3).

Wheeler summarizes this idea in the phrase "*it from bit*" and maintains that "every item of the physical world has at bottom - at a very deep bottom, in most instances - an immaterial source and explanation." (p.5)

Laszlo (1993) maintains that proponents of such `bootstrap' theories

> contest [the conventional atomistic view] that the physical world is built of identifiable building blocks. There are no basic particles; everything is built of everything else. Particles are made of other particles by binding forces that are themselves created by the exchange of particles among particles - the observed world lifts itself into existence by its own bootstraps. In Heisenberg's [and Wheeler's] view this bootstrapping world is built as a mathematical structure; thus there is no use asking to what beyond themselves, the formulas of physics would refer. (p.34)

The meaning circuit concept is important because it supports a closed universe view (chapters 2 and 3) and the "strong" artificiality thesis (section 4.3.5). This is because, by adopting a variant of the coherence theory of truth (chapter 3), it provides a self-grounding interpretation which, according to Wheeler (1989), is the *only*[67] alternative to a `tower of turtles', that is, an infinite regress. For this reason, Rasmussen (1991b) has adopted the participatory interpretation in defining the necessary and sufficient conditions for "strong" A-Life, viz.

Postulate 1: *A universal computer at the Turing machine level can simulate any physical process (Physical Church-Turing thesis).*

Postulate 2: *Life is a physical process.*

Postulate 3: *There exist criteria by which we are able to distinguish living from non-living objects.*

Postulate 4: *An artificial organism must perceive a reality R_2, which, for it, is just as real as our "real" reality, R_1, is for us (R_1 and R_2 may be the same).*

Postulate 5: *R_1 and R_2 have the same ontological status.*

Postulate 6: *It is possible to learn something about the fundamental properties of realities in general,*

---

67 This statement must be qualified because, as will be shown in chapter 6, alternative frameworks such as supernaturalism and Heideggerian ontology allow the universe to be grounded without necessitating an infinite regress of levels. (Supernaturalism grounds the universe in God and Heideggerian ontology in Being.)

and of R_1 in particular, by studying the details of different R_2's. An example of such a property is the physics of a reality.

Postulates (4-6) are grounded in an ontological version of the coherence theory of truth (chapter 3). This is significant because coherence theories assume a specific set of ontic (productive) and epistemic (interpretative) relations between a human being (anthropic component) and its world and, as will be shown in chapter 7, epistemic relations are intimately bound up with ontic relations under the concept of *poiēsis* (coming-forth, bringing-forth). If it can be shown that the mode of *poiēsis* is different in naturals and artificials (chapter 6) then the above postulates and as a corollary, the "strong" computational artificiality thesis, will be undermined.

4.7.6.4. Artificial Physics (A-Physics)

Artificial physics may be defined as the study of artifactual systems that exhibit behaviours characteristic of natural physical systems. A two-level explanatory framework can be proposed in which the `characteristics' level (2) includes concepts such as matter, energy, force, velocity etc and the `building block' level (1) is defined in terms of computation. A computational concept of matter based on the cellular automaton formalism is described in detail in chapter 5 in connection with a unifying framework of emergent artificiality. However, it is worthwhile briefly considering two opposing approaches (top-down and bottom-up) to artificial physics (or A-Physics): Hayes (1979) applies the top-down approach characteristic of conventional symbolic AI to A-Physics, describing a formalization of common-sense knowledge about the everyday physical world of objects, shape, space, movement, substances (solids and liquids), time etc. Although this is the standard approach in artificial reality (A/VR) development as described in section 4.8, it does not map readily onto an emergentist framework. Rasmussen et al. (1991a), on the other hand, present a computational framework for self-programmable matter based on the notion of a modified von Neumann machine and discuss how this framework can be translated to alternative schemes capable of supporting universal computation, for example cellular automata (chapters 2 and 5). According to Rasmussen et al.,

> the term *programmable* indicates that the dynamics of such systems have a clear computational interpretation and that functional properties can be programmed into the system via the elements. The term *matter* indicates that the dynamics is defined through the interactions of the fine-grains of the system, e.g. at the level that defines the `physics' of the system. Thus, *self-programmable matter* is a dynamical system of interacting elements, with associated functional properties, which through their autonomous dynamics develop new compositions of elements with new associated functional properties. Such systems are characterized by *an ability to construct novel elements within themselves*. (p.213)

4.7.7. Matter and Intentionality

Mind and life are readily interpreted in teleological (goal-directed, intentional) terms;

moreover, the (almost) natural attribution of teleology to such phenomena facilitates the establishment of a link between human intelligence and biological adaptation, a link which is fundamental to Simon's concept of artificiality (section 4.2.1). Applying a teleological interpretation to matter, however, would generally be considered to entail a *category error* although there is clear historical precedent for such a move with Aristotle and neo-Aristotelian process philosophers such as Whitehead (1978). According to Serres (1981),

> the word `matter' is derived from the Latin *materia* denoting something substantial and massive, whereas the Greek word doesn't correspond to that at all. The Greek *ule* has a very practical meaning referring to the basic material used by a craftsman, for instance by a carpenter. [Thus, Matter has a concrete or experiential meaning which indicates] the work that is put in, the *transformation* of things. (p.180)

In both Plato and Aristotle, matter is defined in opposition to *form*. The transformation of matter on the Aristotelian scheme is interpreted in terms of the doctrine of the four causes, viz. material, efficient, formal and final causation (chapter 6). The idea of a transformation of matter from an indeterminate to a determinate condition via the imposition of form leads naturally to a teleological or intentionalistic view of matter. Serres (1981) maintains that since the beginning of the twentieth century, "physicists have talked about atoms, molecules, particles, quantum fields and many other things like that, but they hardly ever used the word `matter'. *It is only used any longer by metaphysicists*: for instance by materialists." (p.183)

A link between matter and intentionality is proposed in a modern context by Pattee (1995b), who presents the following definition of matter based on the distinction between physical laws and symbols, viz.

> by matter and energy [are meant] those aspects of our experience that are normally associated with physical laws. These laws describe those events that are as independent of the observer as possible, i.e. independent of initial conditions. The laws themselves are moot until we provide the initial conditions by a process of measurement. Laws and measurements are necessarily distinct categories. Laws do not make measurements, individuals make measurements. *Measurement is an intentional act* that has local significance and hence involves symbolic aspects usually in the form of a numerical record [emphasis added]. (p.11)

Thus, on Pattee's view, intentionality enters into the discussion at the `interface' between matter and symbol during the measurement process.

4.8. Reality

In this section, the concept of artificial and virtual reality is briefly examined. This concept is important because it encapsulates the notion of artificiality in a totalistic sense, viz. reality as the aggregate of all phenomena (material, vital, mental etc).

4.8.1. What is *Reality* ?

This is a difficult concept to define, one which is intimately tied up with other related notions such as truth and appearance[68]. For present purposes, reality is identified with the phenomenal world, that is, with (1) the world as experienced through the senses and (2) the domain within which agents - that is, entities which are capable of initiating action - are situated. Defining reality in these simple terms is important because it leads directly to the concept of artificial (or virtual) reality or, to paraphrase Langton (1989b), from reality-as-it-is to reality-as-it-could-be.

4.8.2. Artificial (or Virtual) Reality (A/VR)

Baudrillard (1983) captures the essence of virtual reality (or, in his terms, the *hyperreality* that is the simulacrum) in the following statement:

> This is a completely imaginary contact world of sensorial mimetics and tactile mysticism; it is essentially an entire ecology that is grafted on this universe of operational simulation, multisimulation and multiresponse. (p.140)

Spring (1991) and (Heim,93) define `virtual reality' via a composition of entries from *Webster's Dictionary*, viz. "a fact or real event that is such in essence, but not in fact". Hence, virtual reality is committed to functionalistic-essentialism: A virtual reality (VR) is not required to duplicate the contingency of the natural world (naturality) but its essential characteristics. Foley (1987) identifies three components as essential in artificial realities (ARs), viz. imagery, behaviour and interaction, while Burdea (1994) defines virtual reality in terms of immersion, interaction and imagination. Krueger (1991) supports the functionalist conception of ARs maintaining that what is important about an AR is the capacities it provides for creating synthetic realities for which there are no antecedents. This view is endorsed by Heim (1993) and also by Laurel (1993) who asserts that "the primary concern of VR is not constructing a better illusion of the world; it is learning to think better about the world, and about ourselves." (p.214)

Walser (1991) defines virtual reality in terms of the related concept of cyberspace:

> Cyberspace as a phenomenon is analogous to physical space. Just as physical space is filled with real stuff (so we normally suppose), cyberspace is filled with virtual stuff. Cyberspace, the medium, enables humans to gather in virtual spaces. It is a type of interactive simulation, called a *cybernetic simulation*, which gives every user a sense that he or she, personally, has a body in a virtual space. Just as cybernetic simulation is a special kind of interactive simulation, a *cyberspace*, the phenomenon, is a special kind of virtual space, one that is populated by people with virtual bodies. [Crucially,] *a cyberspace must have at least one human player* (since a cyberspace emerges from a cybernetic simulation, which embodies a person), but the other players can be AI programs running on decks [i.e. physical entry points into cyberspace] that are not being used by humans. (p.58)

68 A Heideggerian interpretation of these concepts is presented in chapter 6.

Cyberspace, as defined above, is consistent with Wheeler's meaning circuit, more specifically, with the irreducibility of observer-participation. This position is endorsed by Pimentel et al. (1993) who maintain that an essential component of VR ontology is support for *designed* (or participative) experience, viz. “every sensory detail is a design decision.” (p.147) As Pimentel et al. state, virtual reality is “the experience of *being in another world*, a world governed by selected laws, and inhabited by objects (and actors) with whatever properties the creator chooses to design.” (p.16) Heim (1993), following Heidegger, defines a *world* as "a total environment for human involvement"; *the* World is identified as the "horizon or totality of all involvements".[69] The notion of involvement is very closely connected to the idea of *presence*. Sheridan (1992) defines the sense of `virtual presence' as "feeling as if you are present in the environment generated by the computer" and proposes three principal determinants for the sense of presence:

1. extent of sensory information
2. control of relation of sensors to environment
3. ability to modify physical environment

According to this view, two necessary conditions for involvement are: (1) a sense of immersion and (2) the capacity for interaction. These provide the basis upon which a Turing Test for artificial or virtual reality may be constructed. However, Pimentel et al. (1993) have contested the first condition maintaining that the experience of involvement appears to require triggering of the user's *imagination* rather than stimulation of his or her *senses*; consequently, imagination is of higher priority than immersion if the latter is interpreted in the sense of "degree of similarity with reality". As they state, "neither the designer nor the user actually believes the actions on the stage or in the computer are real, but they agree to pretend *as if* they are real [emphasis added]." (p.154) Consequently, they argue for a redefinition of immersion in essentialist terms, that is, in terms which are independent of the contingency associated with sensations occurring during human interactions with the real world.

4.8.3. A Turing Test for Reality

Adopting the terminology associated with the description of the Turing Test in section 4.3.6, a Turing Test for reality, would involve replacing *B* (the human candidate) by the world considered in various ways, viz. physically, experientially, socially etc and determining whether or not a computer-generated A/VR supports involvement or virtual presence, that is, the experience of presence within an environment by means of a communication medium (Steuer,92). Such a medium would need to provide a sense of immersion and the possibility for navigation and interaction within the artificial reality such that the interrogator would not be able to use sensory cues to determine whether or

[69] Consequently, *the* world incorporates the virtual worlds created by human artificers.

not his or her current environment was real or virtual[70] (Shapiro,92). In discussing the requirements for a virtual reality to pass the Turing Test for AR, Walser (1991) notes that "what matters is the extent to which players are willing to suspend disbelief in the illusion that they inhabit bodies apart from their physical bodies." (p.56) This position is supported by Heim (1991) who argues for a distinction between reality and artificial reality, viz. "cyberspace should evoke imagination, not repeat the world." (p.33) Heim further states that,

> Cyberspace can contain many alternate worlds, but the alternateness of an alternate world resides in its capacity to evoke *in us* alternate thoughts and alternate feelings [emphasis added]. (p.33)

Walser (1991) supports this position, viz. "the art .. is not in what the spacemaker [or creator of the virtual reality] constructs, but in the communication of insight which the spacemaker cannot construct (that is, some aspect of a deeper truth or higher reality)." (p.57) This leads Heim (1991) to list three features by which reality might be differentiated from A/VR and which must be embodied in artificial worlds if such alternate realities are to pass the Turing Test for reality: (1) mortality/natality; (2) temporality or continuity of events from the past into the future; and (3) care or concern. The latter criterion is interesting because it establishes a link between A/VR and intentionality (chapters 1 and 6). However, there is problem with defining immersion in terms of involvement such that experiential isomorphism with the real world is regarded as contingent: It negates the possibility of a Turing Test for matter since the latter necessitates that the essential features of the physical world *as perceived by the senses* are incorporated as test criteria (for example, the relativistic, quantum and thermodynamic properties of physical systems).

4.8.4. The Ontological Status of Virtual Worlds

One of the principle distinctions between conventional approaches to A/VR and the approach to reality construction described in this thesis, viz. computationally emergent artificiality (chapter 5), is in methodology: Virtual worlds are usually constructed top-down whereas an autonomous artificial universe is constructed bottom-up via emergent processes (chapter 3). This reflects in the choice of ontological primitives for the virtual worlds: Conventional approaches to A/VR construction assume pluralism (heterogeneous primitives) whereas computationally emergent artificiality is based on a monism, viz. computationalism (chapter 2). For example, in NPSNET, a workstation-based, 3D visual simulator for virtual world exploration and experimentation (Zyda,92), the physics of the artificial world is constructed via interaction with the system user under a Newtonian (object, force) framework:

---

70 Interestingly, Dennett (1991), a proponent of functionalism and supporter of the "strong" AI programme, adopts what might be regarded as a "weak" AR position in asserting that hallucinations indistinguishable from reality are probably impossible.

> Objects are defined with the *defobject* token .. This structure is flexible and useful for building complex objects from simpler sub-objects. An object can be given many physical properties using the *defphysics* token. These properties include the object's initial location and location constraints in the environment, initial orientation and orientation constraints, initial linear and angular velocities and constraints on each, the object's mass and centre of mass, the object's ability to absorb forces (elasticity), the dimensions of a bounding volume and a local viewpoint for the object.
>
> Forces are defined and added to an object's force list with the *defforce* token. Two types of force are supported: deforming and non-deforming. Deforming forces are used for object explosions and bending. Non-deforming forces are used to alter an objects linear and angular velocities. Forces can be specified as awake or asleep. This allows the selective application of previously defined forces. (p.150)

However, there have been attempts to move away from classical top-down conceptions of the world towards approaches incorporating self-organization. For example, Walser (1991) maintains that

> under the classical scientific view there is no need to give a place to the human body in any account of human reason because the classical view presupposes the existence of an objective reality with a rational structure. Reason is treated as a purely abstract system for converging step by step on the one correct description of the world. Under the new view, however, the world is not assumed to have a rational structure, and there is no sense in trying to find one. Instead, there are many possible worlds, as many as sentient beings can invent and experience. (pp.53-54)

Walser argues for a *relativistic* conception of reality, viz. "a *virtual reality* is a consensual reality that emerges from an interactive simulation .. in contrast to a consensual reality that emerges from the ordinary physical world. By *consensual reality* I mean the world, or a simulation of the world, as viewed and comprehended by a society" (p.55) and further, "a virtual reality is `consensual' in that its players [that is, the humans involved in the interaction] have agreed, explicitly or implicitly (by virtue of their participation), to relate to it in the same way, to `play fair'. But the reality is constructed through an organic process of give and take among the players, whether through cooperation, conflict, negotiation, compromise, agreement, force, abstention, or whatever." (p.55)

In contrast to this `evolutionary-emergent' interpretation, Laurel (1991a, 1991b) emphasizes an `intentionalistic-design' interpretation of virtual reality based on the paradigm of dramatic interaction. For Laurel (1991a), the issue is "whether virtual worlds and the experiences people have in them are or are not *designed*." (p.95) Laurel maintains that the Aristotelian conception of dramatic action involves the actualisation of potentiality via a progression through probability culminating in necessity which, in turn, involves the imposition of *constraints* by agents, that is, those with the capacity to initiate action. On her view, "the course of the action and the outcome can be variable, but only within the universe of possibilities created by the elements of environment, situation and character." (p.96) Thus, Laurel proposes that "virtual worlds should, in some sense, be *designed*. By `designed' I mean that a world and the experiences that one can have in it are consciously shaped. The fact is that by their very nature virtual worlds *are* designed, whether we admit to it or not." (p.97) Hence, virtual or artificial realities are ontically *a posteriori* with respect to the human being (anthropic component), that is, they are *artifacts* (chapter 7). This position is clarified by the following assertion in (Laurel,91b), viz. "an obvious but easily overlooked element of situation-building is the fact that all of the relevant aspects of the situation-building must be successfully represented." This is the conventional top-down symbolic view of world *construction*. However, in the postscript to the 1993 edition of *Computers as Theatre*, Laurel appears to reverse the position argued in (Laurel,91a) and (Laurel,91b) in maintaining that

> as an activity becomes less artifactual (like painting or literature) and more ephemeral (like conversation or dancing), sensory immediacy and the prosody of experience gain primacy over structural elegance in the realtime stream of events. In shared virtual worlds, structural elegance becomes much less about the progression of events and more about facilitating the *emergence* of patterns and relationships [emphasis added]. (p.208).

Furthermore, "rather than figuring out how to provide structure with pleasing emotional textures, the problem becomes one of creating an environment that *evokes* robust projective construction [emphasis added]." (p.209) She concludes by stating that "as long as designers see themselves as authors of one-to-many experiences, all of us will only be bottom-feeding on the fringes of fundamentally non-interactive forms." (p.212) Can the two positions, viz. A/VR as designed *vs.* A/VR as emergent be reconciled ? The debate over the ontological status of artificial (or virtual) worlds has been defined in terms of whether the artificiality is "strong" in the sense of realization (emulation) or "weak" in the sense of simulation. However, it will be argued in chapter 6 that this *epistemological* approach to the problem of artificiality (which is grounded in the Kantian appearance-reality distinction) obscures a more primordial issue, viz. the ontological question concerning the distinction in *poiēsis* (coming-forth or becoming) of artificials (or artifactuals) as contrasted with naturals.