ON THE NATURE OF CAUSATION IN COMPLEX SYSTEMS

George F R Ellis[1]

Abstract: *A reliable understanding of the nature or causation is the core feature of science. In this paper the concept of top-down causation in the hierarchy of structure and causation is examined in depth. Five different classes of top-down causation are identified and illustrated with real-world examples. They are, 1. mechanical top-down causation; 2. top-down causation via non-adaptive information control; 3. top-down causation via adaptive selection; 4. top-down causation via adaptive information control; and 5. intelligent top-down causation (i.e. the effect of the human mind on the physical world). Recognising these forms of causation implies that other kinds of causes than physical and chemical interactions are effective in the real world. Because of the existence of random processes at the bottom, there is sufficient causal slack at the physical level to allow all these kinds of causation to occur without violation of physical causation. That they do indeed occur is indicated by many kinds of evidence.*

1: CAUSATION AS THE CORE OF SCIENCE

The nature of causation is a core issue for science, which can be regarded as the move from a demon-centered world to a world based on reliable cause and effect, tested by experimental verification (Sagan 1996). Causes are separated from effects[2] by searching for correlations between phenomena such that manipulation of one (`the cause) can be shown, in a specific context, to reliably result in specific changes in the other (`the effect') at a later time. One has to search for this correlation in the midst of internal and environmental noise (Pearl 2000). Laboratory tests of isolated systems allow an understanding of the elements of causation, which are interactions between the particles that underlie all physical existence; these interactions in multiple combinations underlie the emergence of complex phenomena such as life.

Thus physics is the basic science, characterized by mathematical descriptions (Penrose 2005) that allow predictions of physical behavior to astonishing accuracy and underlies the other sciences (Feynman 1995, Chapter 4; Hewitt et al, 2007). The key question is whether other forms of causation such as those investigated in biology, psychology, and the social sciences are genuinely effective, or are they rather all epiphenomena grounded in purely physical causation? The latter view is suggested by strong reductionist views based in the fact that all physical entities we see around us, including ourselves, are based in the same chemical elements (Emsley 2003) composed from the same kinds of elementary particles, interacting with each other only through the four fundamental physical forces (Penrose 2005, Oerter 2006). How can there then be room for any other type of causation?

---

[1] Mathematics Department, University of Cape Town, Rondebosch, Cape Town 7701, South Africa
Email: George.Ellis@uct.ac.za
[2] http://en.wikipedia.org/wiki/Causality.

I will claim here that there are indeed other types of causation at work in the real world, described quite well by Aristotle's four types of causes (discussed in Section 6 below). There are of course many contexts in which different kinds of causality are experienced: in physics and chemistry, where particles and forces interact in a way described by variational principles and symmetries; in biochemistry and cell biology, where information is important and adaptation takes place; in zoology, where purpose, planning, and anticipation are important; and in psychology and sociology, where analytic reflection, symbolic understanding, values and meaning all are causally effective.

How does it all fit together? The overall framework for understanding these forms of causation and their interaction is the hierarchy of complexity (see Figure 1),[3] ranging from particle physics and nuclear physics to astronomy and cosmology on the one hand, and to psychology and sociology on the other, with coarse-graining and consequent loss of detailed information relating each of the higher levels to lower levels. This structuring leads to the emergence of effective (phenomenological) laws at each of the higher levels, with apparent autonomy from the lower levels (Anderson 1972). It is this independence from the details of lower level causation that allows phenomenological laws to be good effective theories of higher level interactions (for they are levels of stable constitutive relationships); thus for example motor mechanics and neurosurgeons do not have to understand particle physics or nuclear physics in order to ply their trade. Thus the context of the discussion is the modular hierarchical structures underlying complexity (Flood and Carson 1990; Simon 1992, Chapter 7). The key idea I will pursue is that as well as bottom-up causation, top-down causation takes place in these structures (Campbell 1974, van Gulick 1995), due in particular to the crucial role of context in determining the outcomes of lower level causation (Bishop and Atmanspacher 2006). This takes place in four overall physical situations that certainly occur: namely in the natural world of inanimate objects, in plants and animals; in intelligent beings, and in manufactured artefacts[4].

I suggest there are at least five different types of top-down causation that can take place, depending on the context: namely, mechanical top-down causation; top-down causation via non-adaptive information control; top-down causation via adaptive selection; top-down causation via adaptive information control; and intelligent top-down causation. There could be others, but I claim that these can all be regarded as well-established. This paper will explore these forms of top-down causation and the associated implications for our understanding of the nature of causality, and hence for the nature of the scientific endeavour. In brief: the conclusion is that *there are other forms of causation than those encompassed by physics and physical chemistry*. A full scientific view of the world must recognise this fact, or else it will ignore important aspects of causation in the real world, and so will give a causally incomplete view of things (Ellis 2005, 2006a, 2006c). This has obvious implications for reductionism, but I will not explicitly enter the complex debate about that issue here.

---

[3] http://en.wikipedia.org/wiki/Hierarchy.
[4] See Simon (1992, Chapter 1) for discussion of the domain of manufactured objects (artefacts). They are certainly part of what exists, and so need to be included in an examination of causality in the real world.

2: FUNCTIONAL CONTEXT: MODULAR HIERARCHICAL STRUCTURES

The context of the emergence of complexity is the hierarchical structure of matter and causal relations, characterized both by scale and by an appropriate classification and language of description for the entities that are recognized at each scale. It is a hierarchy of whole-part relations, which at the bottom levels can be seen as physical (one entity is physically a part of a larger one) but at the higher levels is a causal hierarchy (one entity provides the causal context for the other). Examples are,

Example 2.1: The physical hierarchy overall, see Figure 1a.[5]

Example 2.2: Hierarchical structure in the life sciences, including intelligent life (Peacocke 1989, Campbell and Reece 2005), see Figure 1b.

Example 2.3: Hierarchical structure in digital computers (Tanenbaum 1990).

Example 2.4: The hierarchical structure of the brain (Scott 1995).

To enable true complexity to emerge, there will be numerous quasi-independent modules at each level of the hierarchy (West-Eberhard 2003, Schlosser and Wagner 2004), interacting with each other in a network which can be represented by causal diagrams (Greenland and Pearle 2006), and enabling encapsulation, information hiding, abstraction, and inheritance (Booch 1994).An example is metabolic networks (Ravasz et al, 2002)[6]. We will not develop these aspects of complexity further in this paper, except to note first that the modules at each level will generally constitute the next lower level in the hierarchy, and second that the network structure is an irreducible higher-level characteristic. In addition to the properties of the units themselves, it is the set of relations between units, for example large scale topological relations as well as local causal motifs (Vazquez et al 2004, Barabási and Oltvai 2004, Alon 2007) that is crucial in building up complexity. These aspects cannot be reduced to lower level variables.

A hierarchical structure will be described by a corresponding hierarchy of variables appropriate to describing the different levels of the hierarchy. A *high level variable* is a quantity that characterizes the state of the system in terms of a description using high level concepts and language – it cannot be stated in terms of low level variables. Use of such variables involves information hiding, for they are the relevant variables for the higher level description, e.g. the pressure, temperature and density of a gas, without including unnecessary lower level details (such as all molecular positions and velocities). The higher levels of structure and causation cannot be reduced to lower level terms, as the relevant concepts lie outside those that can be described in terms of lower level

---

[5] A much more detailed description is given in http://www.mth.uct.ac.za/~ellis/cos0.html. The hierarchy is implicit rather than explicit in most text books on physics (e.g. Hewitt 1998), and is omitted even in texts on integrated science (e.g. Hewitt et al 2007) and science literacy (e.g. Rutherford and Ahlgren, 1990).
[6] Other examples are neural networks, gene regulatory networks, signal transduction networks, cell signaling networks, and protein-protein interaction networks.

concepts. As it is causal relations that count at the higher levels rather than physical nature or scale, high level entities that occur in the life sciences hierarchy need not have a material nature (ethical values are an example, see below). They nevertheless have a clear place in the causal hierarchy, which can be thought of as bifurcating into a natural sciences and a life sciences branch with a single trunk (Murphy and Ellis, 1995).[7]

Some high level variables can be obtained from low level variables by coarse graining, but some cannot. This is a crucial distinction. We can refer to them as *coarse-grained high level variables* and *irreducible high level variables* respectively. Gas pressure is a coarse grained high level variable. Tertiary (folding) structure of bio-molecules is an irreducible high level variable. The latter *supervenes on*[8] the lower level variables, rather than being obtained by coarse graining (which typically involves an integral over a set of lower level variables). Coarse graining is a key idea on the natural sciences side. It is an open question as to how useful it is on the biological side.

Equivalence classes,[9] based in equivalence relations,[10] are a key concept in terms of understanding top-down causation (Auletta et al, 2007). An equivalence class identifies all lower level states where the corresponding higher level variables are equivalent as far as the higher level behavior is concerned; that is, they form good predictors of higher level behavior. In general, they can be based in many different lower level states.

Example 2.5: An equivalence class of molecular states can give the same pressure, density and temperature of a gas (Goodstein 1985); it is these higher level variables that are relevant for understanding and predicting gas behavior, as for example in the ideal gas law[11]: $P\,V = nRT$. Entropy[12] is a measure of the amount of variation at the lower levels that gives the same state at the higher level (Penrose 1989, 309-314); it characterizes how many values of hidden variables can underlie the same higher level description (they are integrated out or averaged over to attain the high level description).

Example 2.6: Structural equation models can be observationally equivalent (Pearl 1998, Section 2.3; Pearl 2000).

3 BOTTOM-UP CAUSATION

Bottom-up causation is the ability of lower levels of reality to have a causal power over higher levels, in some cases uniquely determining what happens at the higher levels (Figure 2a). As the lower level dynamics proceeds, for example diffusion of molecules

---

[7] Some characterise the hierarchy as one of *increasing complexity*. While this is true for some parts of the hierarchy, it is not always true; for example the universe on the largest visible scales is much simpler than many of its constituents (an amoeba is much more complex than the universe). In any case this proposal is problematic in that there is no good definition of complexity that applies broadly; for example is a society, an ecosystem, or a human being more complex?

[8] http://en.wikipedia.org/wiki/Supervenience.

[9] http://en.wikipedia.org/wiki/Equivalence_class.

[10] http://en.wikipedia.org/wiki/Equivalence_relation.

[11] http://en.wikipedia.org/wiki/Ideal_gas_law.

[12] http://en.wikipedia.org/wiki/Entropy.

through a gas, the corresponding coarse-grained higher level variables will change as a consequence of the lower level change, for example a non-uniform temperature will change to a uniform temperature. Examples are:

Example 3.1: Gases. Determination of gas properties through the kinetic theory of gases, seen as bottom-up causation from molecular motions to gas properties (Goodstein, 1985).

Example 3.2: *Solids.* Determination of metallic properties such as conductivity through the quantum theory of solids, based on the study of electron movements in a lattice (Goodstein 1985, Chaikin and Lubensky 2000).

Example 3.3: The Chemical Bond. The explanation of the chemical bond in terms of physical processes involving electron orbitals (Laidler et al 2003, Buyana 1997).

Example 3.4: Genetics. Animal development as based in the reading of genes entails bottom-up causation from DNA to the phenotype (Watson et al 2003, Campbell and Reece 2005).

Example 3.5: *Action Potentials*. The understanding of neuronal processes in terms of ion diffusion and the Hodgkin-Huxley equation (Scott 1995, Kandell et al 2000, Greenspan 2007).

Example 3.6: *The brain,* understood as a neural network built up from interacting neurons (Scott 1995, Kandell et al 2000, LeDoux 2002, Koch 2004).

The core of the strong reductionist view of science is that all can be explained by such bottom-up mechanisms based in the laws of physics, with no remainder. This project is very successful in many cases (see e.g. Chandrasekhar 1998). Bottom-up causation allows a certain degree of complexity to be spontaneously built up in non-equilibrium situations, often demonstrating symmetries and broken symmetries, without higher level guidance (Clayton 2006, Chapter 3). Self-assembly and self-structuring can lead to emergence both of simple structures such as those associated with dynamical system attractors,[13] for example stars and galaxies, and to more complex patterns such a Benard Cells,[14] patterns associated with the reaction-diffusion equation,[15] those occurring in the Game of Life,[16] and biological examples such as ant hills and flocks of birds.[17]

These however do not extend to truly complex systems such as a single living cell[18] (see Maynard Smith and Szathmary 1997, Morowitz 2002, Luisi 2006 for how they came into being). It seems that developing very complex systems such as those occurring in biology requires various forms of top-down causation. Even simpler examples of emergence such

---

[13] http://en.wikipedia.org/wiki/Attractor.
[14] http://en.wikipedia.org/wiki/Benard_cells.
[15] http://en.wikipedia.org/wiki/Reaction_diffusion.
[16] http://en.wikipedia.org/wiki/Conway%27s_Game_of_Life.
[17] http://en.wikipedia.org/wiki/Emergence.
[18] http://en.wikipedia.org/wiki/Cell_%28biology%29.

as those mentioned above depend to some degree on the boundary conditions; this is a form of top-down causation, see Section 4.1 below.

4: TOP-DOWN CAUSATION

Top-down causation (Campbell 1974, Luisi 2002, van Gulick 1995)[19] is the ability of higher levels of reality to have a causal power over lower levels (Figure 2b). Dynamic effects take place at some time, and the outcome would be different if the higher level context were different. Altering the high-level context alters lower level actions, which is what identifies the effect as top-down causation, where the high level context variables are not describable in lower level terms, which is what identifies them as context variables.

One should note here, following Auletta's terminology (see Auletta et al 2007), that there is a significant difference between having *causal power* (to concur in producing a certain effect) and having *causal effectiveness* (which, in ideal situations, would suffice to bring about an effect, given the other conditions). Formal causes, for example constraints, only have causal power but not causal effectiveness. For example,[20] the structure of a forest (the way the trees are disposed) together with other environmental items (rocks, plants, rivers, and so on) will affect (have causal power on) the way natural agents (like wind or fire) will propagate. For instance, wind may be more canalized in some parts and blocked in other ones. However, what is here causally effective is the wind (or the fire), not the structure of the forest, which would remain completely ineffective (not operative) and unable to concur in any causal process without an effective causal agent. All formal constraints have this character. Top down causation as considered here means having causal power over lower levels, channeling causal effectiveness at those levels.

Top-down causation is ubiquitous in physics, chemistry, and biology, because the outcome of lower level interactions is always determined by context. For example, the motion of the Moon round the Earth causes tides locally on Earth, a top-down influence from a scale of 384,000 km to a scale of meters, which then influences the lives of crabs; fluctuations in the interior of the Sun cause radiation changes that alter conditions in ecosystems on Earth, influencing the distribution of micro-organisms; and so on. These top-down effects result firstly from the fact that theoretical physics is based in partial differential equations whose solutions depend on the boundary conditions, and secondly because geometrical and structural relations exist in complex systems that dictate which components can interact with which others through which physical effects. For example the wiring in a computer channels electrons from one specific component to another and thus enables logical computations to be performed. The kind of computation performed and resultant output, and hence the detailed switching of transistors at the micro level, depends both on the component connectivity and on the kind of programme loaded into the computer (word processor, music, or graphics for example) – a high level concept. These are constraints on the lower level dynamics and so have causal power (Juarrero 2002). They are causally effective only when such formal constraints from above are

---

[19] Some prefer the term *whole-part causation*.
[20] I thank Gennaro Auletta for this example.

combined with efficient or circular causes at the same ontological level. Note that we can consider such same level causation at each level as ontologically real (if this were not so we could not know what the ontologically real causation was, as we do not know what the fundamental bottom level of physics is).

Top-down action is the special case of top-down causation where higher level variables are manipulated so as to cause changes in lower level variables. An example is turning a light switch on. This causes electrons to flow in the wire so that the bulb lights up. The switch is a higher level entity: it cannot be described in lower level terms (albeit it is constructed from lower level parts). There is a same level explanation in terms of phenomenological variables at the level of the switch and wiring (operating the switch closes the circuit). This is always what happens when reliable behaviour occurs at any level: it is described by an effective theory[21] (Hartmann 2001) at that level. This same level theory is however enabled by the lower-level action of electrons flowing or not flowing. Whether they flow or not depends on higher level variables (in this case, the switch state) which simply cannot be meaningfully described in terms of lower level variables per se (actually because it is equal to an equivalence class of lower level variables). There is indeed a bottom-up explanation, when we are given the higher level context described by the higher level variables. Without that higher level context, there is no lower level explanation.How do you demonstrate this top-down causation, in this case? - turn on the light switch, and measure the resulting change in current**.** How do you demonstrate that it is due to higher level variables? Take the circuit apart, and the effect will vanish (changing the switch will have no effect). All the lower level variables will still be there; it is only their higher level relationships that will have changed

Effective same-level action occurs when top-down causation combined with bottom-up causation leads to a resulting high-level outcome that depends only on the initial high-level state. In that case, the low level dynamics commutes with coarse graining for all low level states that correspond to each of the high level states (Figure 3b), and a coherent high level dynamics emerges from the lower level dynamics (Ellis 2006c); the light switch is an example. In this case each set of lower level states corresponding to a single higher level state forms an equivalence class as far as the higher level dynamics is concerned. The resulting same-level action allows a phenomenological description of the higher level action that is independent of the particular lower level states that realize this action. This is the basis of the independence of higher level descriptions from lower level details and the reason that we can consider same level causation at each level as ontologically real, expressed in terms of viable *effective theories* for the dynamics at that level (Hartmann 2001). When the lower level dynamics does not commute with coarse graining in this way, no coherent higher level dynamics emerges (Figure 3a); for example this occurs in chaotic dynamics.[22]

Thus the possibility of coherent higher level action emerging from the lower level dynamics is based in the ***principle of equivalence of classes***: the same top level state must lead to the same top level outcome, independent of which lower level states realize

[21] http://en.wikipedia.org/wiki/Effective_field_theory.
[22] http://en.wikipedia.org/wiki/Chaos_theory.

the top level state (Auletta et al 2007). The high level outcome is then the same for the whole equivalence class of lower level variable values, no matter which particular one instantiates the high level state. If different outcomes result from different lower level realizations of one top level state, we do not have reliable same level action resulting from top-down causation.

How do you demonstrate top-down causation? You show that a change in high level variables results in a demonstrable change in lower level variables in a reliable way, after you have altered the high level variable.[23] It is the reliable nature of the change that characterises it as causation and not just a random change; and this is also what leads to predictability (the result is repeatable and thus testable). Thus you just have to show that altering the high-level context alters the outcome in a way depending only on the top level state, where the context variables are not describable in lower level terms.

Can we distinguish different kinds of top-down causation? Yes we can.While all top-down causation can be characterized as due to higher level variables setting the context for lower level action, there are clearly distinguishable ways this can happen. I suggest there exist at least the five different kinds of top-down causation indicated in Figure 4, themselves forming a hierarchy. They might not represent all the forms of top-down causation, but I believe these can all be regarded as well-established. Note that we do not have to explain in detail how these various classes of causation work, in order to determine both that they do indeed represent top-down causation, and that they are distinct from each other. Clarification of how each works, and how to verify this experimentally, is an ongoing research program.

In the following subsections I discuss the nature of each of these five classes of top-down causation in turn, considering them in inverse order.

4.1 MECHANICAL TOP-DOWN CAUSATION.
Mechanical top-down causation occurs when high-level variables have causal power over lower level dynamics through system structuring and boundary conditions, so that the outcome depends on these higher level variables. The resulting high level relations are then an inevitable consequence of the low level interactions, given both the high level context and the low level dynamics (based in physics); all the high-level variables concerned are coarse-grained low level variables or their representations. It is the physical structuring and equations of state that determine the outcome resulting from particular boundary and initial conditions (Shearer et al 1971), also perhaps being influenced by the existence of attractors in phase space, described by the mathematics of dynamical systems theory (Glendinning 1996). The set of all lower-level variable values that correspond to the same values for a set of higher-level characterising variables form an equivalence class relative to the higher level action, and this is the underlying basis of reliable higher level phenomenological laws. This is the kind of causation envisaged in the physicalist reductionist world view, and occurs in all physical and natural systems. Here are some examples:

---

[23] This time sequencing is crucial, even if the intervening time elapsed is very small.

Example 4.1.1: ***Diesel engines:*** Compressing a gas mixture in a cylinder can result in ignition of the gas in a predictable way.[24] A cylinder is a high level concept, as are the pressure and temperature of the gas; low level concepts are molecules of $C_{12}H_{23}$ and O_2, and the chemical and physical reactions between them.

Example 4.1.2: Nucleosynthesis in the early universe (Silk 2001, Dodelson 2003). The nuclear reaction rates are controlled by cosmological variables (average gas density and temperature). The nuclear reactions that take place in the early universe, and hence the elements produced in nucleosynthesis then, therefore depend on the rate of expansion of the universe, determined by macroscopic cosmological variables. The gravitational equations determine the large scale metric and density evolution (global variables) in a bottom-up manner; this then determines the local nuclear reaction rates in a top-down manner (by setting the environment for those reactions), which reactions finally determine the resulting nuclear fractions (global variables again) in a bottom-up manner. Hence the resulting nuclear abundances can be used to determine the average density of baryons in the universe – a key cosmological parameter.

Example 4.1.3: ***Phase Transitions***: Changes in higher level variables cause a phase transition[25], representing a discontinuous alteration in the mode by which lower level interactions lead to higher level behaviour and consequent high-level equation of state. The variables that cause the phase transitions are higher level (coarse-grained) variables. You cannot describe phase transitions without them. In a laboratory situation, they are manipulated by the experimenter to cause the phase transition.

Example 4.1.4: ***Fluid convection*** has been examined in detail as a model for downward causation in classical mechanics (Bishop 2006).

Example 4.1.5: The arrow of time. One cannot tell how a macro-system will behave in the future on the basis of the laws of fundamental physics and the properties of the particles that make up the system alone, because time-reversible micro-physics equally allows two solutions - one the time reverse of the other; but only entropy-increasing solutions in one direction of time occur at the macro-level. This does not follow from the micro-physical laws alone (Zee 1989; Penrose 1989, 2005). Physically, the only solution to this arrow of time problem[26] seems to be that there is top-down causation by the universe as a whole, effective through boundary conditions at beginning and end of space-time, which allows the one solution and disallows the other[27]. This makes a fundamentally important difference to local physical behavior, and so is a crucial form of top-down causation from the whole universe to local systems whose behaviour is not determined on the basis of local physical laws alone (Penrose 1989, 2005; Ellis 2002).[28]

---

[24] See http://en.wikipedia.org/wiki/Diesel_engine .
[25] See http://en.wikipedia.org/wiki/Phase_transition.
[26] Actually there are various arrow-of-time problems: see http://en.wikipedia.org/wiki/Arrow_of_time.
[27] Boltzmann's H-theorem (http://en.wikipedia.org/wiki/H-theorem) does not solve the issue, because it applies equally well in both directions of time (Penrose 1989, 2005).
[28] It should be noted that this is a theoretical explanation that has not been, and indeed cannot be, experimentally verified. Nevertheless there are good reasons for believing this is the case.

Example 4.1.6: Algorithmic computational procedures in a digital computer *that proceed on the basis of initial data only*; the algorithms[29] (stored in high-level computer programs) determine the machine code that then determines the (low-level) switching of transistors (Tanenbaum 1990). This represents top-down causation, as different stored programs will employ different algorithms and so result in different transistor switching. This may eventually become realizable at the molecular level (Conrad 1985). Such machine-like processes controlled by algorithms may entail purpose (a desired outcome) without embodying goals (no feedback control system can be implemented in the algorithm, because in the stated context it cannot utilize updated information). Then they will run tending to produce the desired outcome, but will be vulnerable to disturbances that they are unable to respond to. Thus in real world situations they will be unreliable. Examples are a stock control system that is not related to checks of the actual physical stock from time to time, and an aircraft autopilot that is not fed updated information on position and winds.

Example 4.1.7: An isolated system. Paradoxically, an isolated system is a special case of top-down causation: it is the case where there is no interference with the system by the outside environment, enabling the system to evolve under the influence only of its own internal dynamics. This is the case aimed for in laboratory physics, and is only attainable on Earth for limited time intervals and through human intervention – a strong case of top-down action! A key point here is that an isolated system is only possible in a subset of cosmologies where tidal gravitational forces and gravitational waves are small. In most cosmologies this will not be true, and the universe will interfere with local systems (Ellis 2002); in those cases laboratory determination of the laws of physics will be difficult if not impossible, and it is unlikely that life (and so an experimenter) will come into existence. Thus lack of such interference is a particularly fortuitous form of top–down causation from the global context.[30]

In real-world practice, systems are never isolated, which is why separation of the internal dynamics from external interference is so difficult (Pearl 2000). This interference is a form of top-down causation from the environment to the system, related to the quantum theory idea of decoherence (Blanchard et al 2000, Zurek 2003).

4.2: NON-ADAPTIVE INFORMATION CONTROL.
In non-adaptive information control, higher level entities influence lower level entities so as to attain specific fixed goals through the existence of feedback control loops, whereby information on the difference between the system actual state and desired state is used to lessen this discrepancy (Beer 1966; Simon 1992, Chapter 1).

Ensuring the goal is attained through feedback control functions by comparison of the system state and the goal by a controller; information on the difference is fed back to the activator, see Figure 5. Thus feedback control systems depend essentially on information

---

[29] http://en.wikipedia.org/wiki/Algorithm.
[30] "No causal effect" is of course a special case of generic causal effects; c.f. Sherlock Holmes and the curious incident of the dog in the night (http://en.wikipedia.org/wiki/The_Adventure_of_Silver_Blaze.).

flows,[31] plus an evaluation of that information relative to the chosen goals. Their dynamics can be described by block diagrams and transfer functions (Harrison and Bollinger 1969). Note that in this non-adaptive case (in contrast to the adaptive case considered in Section 4.4), the goals are embodied in the system structure and so do not change with time; there may however be some associated form of information storage and retrieval, and perhaps even implicit or explicit information processing. Thus although the goals are fixed, complex information processing and modeling may take place in the attempt to attain those goals.

These control circuits are higher level entities, as they are based in higher level concepts (the high level system state and the goal). The goals are intrinsic higher-level properties of the system considered, and determine the outcome (unlike the usual physics case, where the initial state plus boundary conditions determine the outcome). A goal is an equivalence class of lower level states: it does not matter which particular lower level state occurs, as long as the corresponding higher level variables are in the desired range. It is this equivalence class (an irreducible high level variable) that is causally effective and determines the outcome in these systems. This underlies all homeostatic systems.

This is top-down causation because the goals are usually[32] only expressible in higher level terms, and are implemented by higher level networks; these cannot be reduced to lower level entities, precisely because it is the relations between the parts that make the network into a feedback control system. Taking the system apart destroys those relations. Note that I use the term `goal' only when there is an identifiable feedback control system that leads to its reliable realization; attractors in a dynamical system are therefore not goals, in the sense I use the term, as they do not involve control information, which is central to feedback control.[33]

Non-adaptive information control only occurs in living beings or manufactured objects; it does not spontaneously arise in the natural world (in astronomy, geology, oceanography, or atmospheric physics).[34] In the biological case, the goals are genetically determined and are the same across a species and independent of experience and learning.

Example 4.2.1: *Steam engine governor*: The classic case is the control of the speed of a steam engine through a centrifugal governor.[35] The controller is a high level device (if you disassemble it into molecules, it will no longer function); the outcome is a desired range of molecular densities and speeds in steam in a pipe (at the lower level) resulting in a desired rate of rotation of a wheel (a high level control variable).

---

[31] This is made explicit in Forrester's `systems dynamics' diagrams (Forrester 1961; Roberts 1981), where information and material flows are represented differently.

[32] The exceptions are some inter-level loops where lower level variables are the object of control.

[33] Corning defines *control information* as "the capacity (know-how) to control the acquisition, disposition, and ultilisation of matter/energy in purposive (teleonomic) processes" (Corning 2005, p. 367).

[34] Any stable state of a physical system will have some of the attributes of a goal, in that perturbations away from the state will (by definition of stability) lead to a reliable return to that state; but this will be because it is a dynamical system attractor, rather than being due to feedback control loops and information flows..

[35] http://en.wikipedia.org/wiki/Centrifugal_governor .

Example 4.2.2: *Thermostat*: A commonplace example is control of temperature of water in a hot water cylinder by a thermostat controlling the water heater.[36] How do you demonstrate this is top-down causation? - change the temperature setting of the thermostat, and then measure the resulting temperature changes in the water and reduced speed of motion of water molecules. This effective action has been verified in daily life many millions of times.

Example 4.2.3: *Homeostatic systems*[37] maintain constant conditions in the internal environment in the human body: "essentially all the organs and tissues of the body perform functions that help to maintain these constant conditions" (Guyton 1977), for example controlling blood pressure and temperature. How do you demonstrate this is top-down causation? - disable the relevant feedback control loops, and measure the resulting temperature changes and alteration in cell conditions.

This is a core feature of physiology; indeed, "this integrative concept [homeostasis] is the key to physiological research, which typically begins with a descriptive observation at the organismal level and leads to a search for underlying cellular and molecular mechanisms" (Rhoades and Pflanzer, 1989, p.22).

Example 4.2.4: Energy homeostasis. In *Drosophila,* information on energy shortage is relayed by lipolytic hormones, enabling metabolic regulation (Leopold and Perrimon 2007). The fat body buffers nutritional information for the entire animal.

Example 4.2.5: *Neuronal Axons*: Sodium and potassium levels in neurons are controlled by voltage gated ion channels[38] deployed in feedback control loops that return the membrane potential of an axon to its resting state, the resting potential[39] (Kandel et al, 2000; Greenspan 2007). *Paramecium* has multiple systems for returning its membrane potential to the resting state (Greenspan 2007).

This is local homeostasis, as opposed to the non-local homeostatic control systems mentioned in Example 4.2.3; such local control systems underlie and enable the more global ones.

Example 4.2.6: *Size and shape in morphogenesis*: In *Drosophila*, "a surveillance mechanism orchestrates proper tissue size and shape, and involves reciprocal interaction between the cell and tissue scales" (Lecuit and Le Goff, 2007).

Example 4.2.7: Microbiology: *Ribonuclease action*.[40] Causal effectiveness of equivalence classes in information control in microbiology has been demonstrated in the case of structural type A and B bacterial ribonuclease P (Rnase P) RNAs (Wegscheid et al 2006). These RNAs can fully replace each other in vivo despite many reported

---

[36] http://en.wikipedia.org/wiki/Thermostat .
[37] http://en.wikipedia.org/wiki/Homeostasis .
[38] http://en.wikipedia.org/wiki/Voltage-gated_ion_channel.
[39] http://en.wikipedia.org/wiki/Action_potential.
[40] http://en.wikipedia.org/wiki/Ribonuclease.

differences in their biogenesis, biochemical/biophysical properties, and enzyme function in vitro. This suggests top-down causation from these equivalence classes of RNA in vivo under standard growth conditions (Auletta et al 2007). If we assume that all the bacteria within the population are perfect clones to one another, then the experiment is clearly seen as non-adaptive (for the adaptive case, see Example 4.4.4).

4.3: ADAPTIVE SELECTION

Adaptive processes take place when many entities interact, for example the cells in a body or the individuals in a population, and variation takes place in the properties of these entities, followed by selection of preferred entities that are better suited to their environment or context (Figure 6). Higher level environments provide niches that are either favorable or unfavorable to particular kinds of lower level entities; those variations that are better suited to the niche are preserved and the others decay away. Criteria of suitability in terms of fitting the niche can be thought of as fitness criteria[41] guiding adaptive selection. On this basis a *selection agent* or *selector* (the active element of the system) accepts one of the states and rejects the rest; this selected state is then the current system state that forms the starting basis for the next round of selection. Thus this can be thought of as a generalised feedback loop with a *meta-purpose* provided by fitness criteria classifying what kinds of outcomes are desirable and which not (these are higher level purposes that are not directly attained as are the goals in a feedback control system, but still effectively guide what happens by selecting preferable outcomes). In some cases the fitness criteria may be implicit rather than explicit, being built into the way the selection agent functions rather than being a separate function. The way they are concretely realized depends on the specific context, as illustrated in the examples below.

Thus this is top-down causation from the context to the system. An equivalence class of lower level variables will be favored by a particular niche structure in association with a specific fitness criteria. Unlike feedback control, this process does not attain pre-selected internal goals by a specific set of mechanisms or systems; rather it creates systems that favor the meta-goals embodied in the fitness criteria. This is an adaptive process rather than a control process. It is the way new information is generated that was not present before (Roederer 2005), and enables emergence of complexity without dynamical attractors or specific goals guiding the process, but with an increase of complexity and embodied information, for the process searches the possible solution space in a way that is not pre-ordained and adapts to the context. The outcome is usually not predictable either from the initial conditions or the meta-goals, because of the random element involved, although both clearly influence the outcome. This underlies all life, including cells and plants and animals. It probably only occurs in living beings (Campbell andReece 2005) and `intelligent' manufactured objects (Simon 1992, Chapter 1), and does not spontaneously occur in the natural world.[42]

---

[41] Gerald Edelman uses the term "value system" (Edelman 1989). This name does not mean it refers to ethical values, or that some personality enters the game and assesses values. It simply identifies the criteria by which selection takes place, either implicitly or explicitly.

[42] Although it has been suggested it could happen in cosmology, see Smolin (1997).

This process usually takes place in the context of a network of interactions. A pressure of selection applied to a higher order network can constrain the evolution and adaptation of lower level constitutive networks. The idea is that the pressure of selection applied to the higher order network would lead to outcomes at the lower level constitutive networks that would be different than those observed by having the same pressure of selection applied to the lower level constitutive networks outside the context of the higher order network. In other words, the higher order network affects the way the lower level networks adapt through evolution. This is top-down causation by adaptive selection.

The more functionally complex the higher order network is, the more constraints enter into consideration and affect the adaptation of lower level networks. Indeed, they have to adapt in a cooperative fashion by taking into consideration not only one or two functional parameters pertaining to a single lower level network, but many different ones. For instance, in a biological context, the pressure of selection will be first on the organism as a whole rather than on the individual constitutive parts. In other words, it is for the benefit of the whole that some constitutive parts will adapt and not for their own benefit alone. This is testable by looking at the way RNA sequences evolve in vivo or in a test tube. Using RNA molecules as model systems, the type of underlying evolutionary mechanisms probably take advantage of "walks" within neutral sequence networks. The walk is constrained by the nature of the higher level that generates some biases. At the level of the RNA biopolymer sequences, that will be indicative of top-down causation by adaptive selection.[43]

Example 4.3.1*: **Darwinian evolution*** results in DNA structuring via adaptive selection over geological timescales (Williams 1992, West-Eberhard 2003, Campbell and Reece 2005), with the meta-goal - the higher level `purpose' that guides the dynamics - being survival of the organism (which is a higher level concept). The development of DNA codings (the particular sequence of base pairs in the DNA) occurs through an evolutionary process which results in adaptation of an organism to its ecological niche; the selector is birth and death, the implicit fitness criterion is survival. A different niche structure results in a different set of genes. As a specific example: a polar bear *Ursus maritimus* has genes for white fur in order to adapt to the polar environment, whereas a black bear *Ursus americanus* has genes for black fur in order to be adapted to the North American forest. The detailed DNA coding differs in the two cases because of the different environments in which the respective animals live.

This is a classic case of top-down causation from the environment to detailed biological microstructure (Campbell 1974) - through the process of evolutionary adaptation, the environment (along with other causal factors) fixes the specific DNA coding. There is no way you could predict or explain this coding on the basis of biochemistry or microphysics alone. The survival of the organism is the fitness criterion, leading to existence of all those detailed conditions which must be fulfilled for survival to be assured. This meta-goal is generally the same for every organism because it is what leads to the existence of populations that fit environmental niches better than competitors.

---

[43] I thank Luc Jaeger for these remarks.

Note that the claim is not that the environment is the *only* relevant factor; rather it is that it is *one* of the causally effective factors. There will always be multiple causal factors, some bottom-up and some top-down; the final result comes from the confluence of these effects (see Section 6 below). How do you demonstrate this is top-down causation? - Change the niche structure, and a different population will adapt to it. Humanity is at present engaged in this experiment, through our causation of global warming. It happened in the past when bacteria changed the Earth's atmosphere to oxygen and nitrogen, and new species adapted to this change.

*Example 4.3.2***:** *Neural Networks and Genetic Algorithms:* Training of artificial neural nets[44] to perform a specific task (say letter recognition) determines the interaction weights in the network (Bishop 1999). The niche is a particular set of letters to be recognised. The fitness criterion is correct pattern recognition, and the adaptive process is the training of the neural network (Carpenter and Grossberg 1991). This is a form of top-down causation from the pattern to be recognized (a high-level concept, as it is defined in terms of the relation between the elements) to the low-level property of network weights. Different sets of weights can perform the same function, so the acceptable set of weights is an equivalence class. Decision making is a property of the network rather than of any single cell (Greenspan 2007).

Genetic algorithms[45] (implemented on digital computers) are specifically designed to solve problems in an adaptive way (Mitchell 1998). A `fitness function' is defined over the genetic representation and measures the quality of the represented solution, thus providing the needed fitness criteria in this case.

*Example 4.3.3***:** *Developmental biology*: The reading of DNA codings in cells is not just a mechanical process, but rather is adaptive at all stages, responding to the environment as development takes place (Fox Keller 2000) and so allowing developmental plasticity (West-Eberhard 2003). Transcription factors modulate the transcription rates of specific genes in response to signals conveying information about the external environment (Alon 2007), enabling an adaptive response to changing environmental conditions. For example, "light is a major environmental factor that impacts many aspects of plant development, including germination and seedling growth" (Weivogel and Dean, 2002).

*Example 4.3.4***:** *Folding of single stranded DNA sequences following reverse mutations*. The selection of native nucleic acid folding (an irreducible higher level variable) is an epigenetic effect, with broad implications in the evolution of plants and their viruses. The folding structure (a higher level variable) corresponds to an equivalence class of lower level sequences, and is the biologically relevant variable determining the selection that occurs. How do you demonstrate this top-down causation? - this has been experimentally shown in detail by Shepherd et al (2006).

Example 4.3.5: Negative selection. It has been demonstrated in *Drosophila* that there can be functional genomic elements that despite having undergone many random

---

[44] http://en.wikipedia.org/wiki/Neural_networks.
[45] http://en.wikipedia.org/wiki/Genetic_algorithm.

mutational events, have not changed in function (Birney 20007). This means these variations form an equivalence class as regards genetic function, which is why any one of the class can be selected equally. In particular, Hox genes do not require clustering to maintain proper function (Drosophila Consortium, 2007).

Example 4.3.6: *The adaptive immune system*[46]. This functions as an adaptive system able to deal with infections never before encountered (Burnet 1959).

One should note here that (as illustrated by the above examples) adaptive selection is the prime way adapted complex structures can be built up on *all three biological timescales:* evolutionary, developmental, and functional. It is central to biological functioning as well as to Darwinian evolution. Thus it occurs in both phylogeny[47] and ontogeny[48].

Finally, it is possible that adaptive selection effects occur in the purely physics context of quantum decoherence (Zurek 2003). This has still to be clarified.

4.4: ADAPTIVE INFORMATION CONTROL

Adaptive information control takes place when there is adaptive selection of goals in a feedback control system, thus combining both feedback control (Section 4.2) and adaptive selection (Section 4.3). The goals of the feedback control system are irreducible higher level variables determining the outcome, but are not fixed as in the case of non-adaptive feedback control; they can be adaptively changed in response to experience and information received. The overall process is guided by fitness criteria for selection of goals. This allows great flexibility of response to different environments, indeed in conjunction with memory it enables learning and anticipation (Simon 1992, Chapter 4) and underlies effective purposeful action as it enables the organism to adapt its behaviour in response to the environment in the light of past experience, and hence to build up complex levels of behaviour.

This requires reliable ways of recognizing and classifying information. It enables goals to be specific to individuals and to vary with time and experience. It underlies animal intelligence and action, and only occurs in living organisms or intelligent artefacts (computer controlled systems).

Example 4.4.1: *Associative learning* in animals, such as Pavlovian conditioning[49]: animal response to a stimulus such as a sound, which is taken as a sign of something else and causes physical reactions implemented by motor neurons. The training is causally effective by top-down action from the brain to cells in muscles. The fitness criterion is avoidance of negative stimuli. How do you demonstrate this top-down causation? - change the conditioning, and the response is different. This occurs in higher animals (dogs for example) and in the snail *Aplysia* (Beer et al 2007, pp. 763-771).

---

[46] http://en.wikipedia.org/wiki/Adaptive_immune_system,
[47] http://en.wikipedia.org/wiki/Phylogeny.
[48] http://en.wikipedia.org/wiki/Ontogeny.
[49] http://en.wikipedia.org/wiki/Classical_conditioning.

Example 4.4.2: Bee responses: Bees respond to abstract classes of patterns, e.g. the dances of other bees, that signify specific actions they must perform,; and they can respond to symmetric and asymmetric patterns of marks in a maze (Zhang and Srinivasan 2004). The latter is learned behaviour in a particular artificial environment, reflecting top-down causation from an irreducible higher level abstract variable (a class of symmetries).

Example 4.4.3: ***The brain***: Neuromodulation allows patterns of neural activity to adapt to new conditions (Greenspan 2007). A form of adaptive selection that has been called "Neural Darwinism" by Gerald Edelman (1998)[50] takes place in the brain, refining neuronal connections on the basis of higher-level fitness criteria provided by a `value system' that guides brain plasticity in response to environmental interactions, and is made effective by neurotransmitters diffused to the cortex from the limbic system [51]. This system evaluating salience is nothing other than the hard wired primary emotions (Ellis and Toronchuk 2005). Thus a key aspect of its implementation in all higher animals is emotions, which are causally effective internal variables[52] in feedback control loops (Damasio 2000). A particular case is *habituation*, that is learning to ignore a stimulus that lacks meaning (Beer et al 2007, pp.763-767).

Example 4.4.4: Microbiology: ***Ribonuclease action***. Causal effectiveness of equivalence classes in information control in microbiology has been demonstrated in the case of structural type A and B bacterial ribonuclease P (Rnase P) RNAs (Wegscheid et al 2006), see Example 4.2.7. If we consider the bacteria organism as a population of variants, then the initial process of transformation necessary to perform the experiment of complementation can eventually be seen as a adaptive selection with the pressure of selection being the maintenance of the feedback control loop assuring the proper functioning of the whole.[53]

4.5: INTELLIGENT TOP-DOWN CAUSATION

Intelligent top-down causation occurs through the action of the human mind on the internal milieu and thence on the external world.This top-down causation from the mind to the world underlies all human planning and action. When a human being has a plan in mind (say a proposal for a bridge being built) and this is implemented, then enormous numbers of micro-particles (comprising the protons, neutrons, and electrons in the sand, concrete, bricks, etc. that become the bridge) are moved around as a consequence of this plan and in conformity with it. Thus in the real world, the detailed micro-configurations of many objects (which electrons and protons go where) is determined by the plans humans have for what will happen, and the way they implement them. Human choices based in self-reflective intelligence (whether free or not) and imagination are thus

---

[50] http://en.wikipedia.org/wiki/Neural_darwinism .
[51] This is in contrast to simple Hebbian processes, which by themselves have no value system indicating salience (other than mere usage). However such a system is provided for Hebbian processes by a context of associative learning.
[52] In analogy to internal representations in neural networks: "hidden units can be used to construct internal representations of the external world which solve difficult problems" (Elman et al, 1998, p. 65).
[53] I thank Luc Jaeger for this remark.

causally effective at many levels,[54] allowing anticipation and intelligent planning and design, based on abstract models of reality (Simon 1992, Chapters 5 and 6).

This is a form of feedback control with adaptive choice of goals as in Section 4.4, but now additionally involves sufficient computing power to support the causal efficacy of abstract entities such as action plans, the theory of the laser, and the value of money, represented symbolically. Indeed the key feature of this higher level of causation, distinguishing it from the general case of adaptive control systems, is its use of *language and abstract symbolism* (Deacon 1997). These are irreducible higher level variables (of an abstract nature: they form equivalence classes of representations) but are causally effective by top-down action from the brain to cells in our muscles (Murphy 2006). This enables information to be stored and retrieved, classified and selected as relevant or discarded, processed in the light of other information, and used to make qualitative and quantitative projections of outcomes and plan future actions in a rational way (Higgins 1976, Simon 1992 Chapter 4), altering goals according to an intelligent understanding of past experiences and future expectations. Intentional action (Juarrero 2002) then enables one to implement the resulting plans, and so change the physical world. In doing so, one should recognise the causal power of images (Boulding 1961) and formal and informal causal models of the natural and social worlds, ranging from mental images of what might happen to elaborate quantitative models of physical entities and societies.[55] These abstract entities (which are shared among many minds) play a large part in formulating our understandings and consequent actions, and hence are causally effective in the real world as they help us attain our goals. Imaginative creations of the mind involve accessing, selecting and organizing information from the environments we inhabit in new and more effective and insightful ways. But where does that information come from in the first place? From the natural, psychological, sociological and intellectual environments we are constantly interacting with. And how is that information generated? Somehow through the gradual emergence, differentiation and development of the systems which constitute social reality, under the pressure of the regularities, processes and relationships influencing their behaviour, interaction and evolution at different levels (Berger 1963).[56]

As far as we know, the only animals on Earth in which this occurs are human beings (although one should keep an open mind: some animals such as bonobos show a remarkable degree of language comprehension and problem solving). However symbolic systems can be embodied in rational artefacts (Simon 1992, Chapter 1).

Example 4.5.1: ***Aircraft Design***: Plans for a Jumbo Jet aircraft result in billions of atoms being deployed to create the aircraft in accordance with those plans**.** This is a non-trivial example: it costs a great deal of money to employ experts in aerodynamics, structures, materials, fuels, lubrication, controls, etc. to design and then to manufacture the aircraft in accordance with those plans. The plan itself is not equivalent to any single person's

---

[54] See for example Philbin (2003).
[55] One should also recognise the role of emotions in shaping behavior (Damasio 2000, Ellis and Toronchuk 2005), as well as the roles of faith, hope, and values (see Ellis 2008). Values are discussed below.
[56] I thank Bill Stoeger for these comments.

brain state: it is an abstract hierarchically structured equivalence class of representations (spoken, drawn, in computers, in brains, etc.) that together comprise the design. It is clearly causally effective (the aircraft would not exist without it).

Example 4.5.2: *A timetable for an airline* determines when aircraft fly in a more or less reliable way. It results in an aircraft flying on a particular path at a particular time, resulting in particular patterns of atmospheric pollution through specific molecules. How do you demonstrate top-down causation? - change the timetable and different patterns of pollution will result. Physics can describe the material out of which the timetable is made and the ink markings on the paper; it cannot comprehend the causal chain by which this leads to particular aircraft flying at particular times. The relevant variables (the entries in the timetable) belong to an irreducible equivalence class of abstract entities coding information that controls what happens in the real world.

Example 4.5.3: *Every physics experiment* is an example of intelligent top-down causation. The experimenter manipulates the world so as to create a repeatable experiment, for example collisions between subatomic particles that are then measured with precision. These collisions would not take place without the active intervention of the experimenter in accordance with his mental plans; hence this is top-down action from the human brain to sub-microscopic scales.

Example 4.5.4: *The value of money.* Physically, money[57] is just coins or pieces of paper with patterned marks on them. This does not explain its causal significance. The effectiveness of money, which can cause physical change in the world such as the construction of buildings, roads, bridges, and so on by top-down action of the mind to material objects, is based in social agreements that lead to the value of money (pricing systems) and exchange rates. These are abstract entities arising from social interaction over an extended period of time, and are neither the same as individual brain states, nor equivalent to an aggregate of current values of any lower level variables (although they may be represented by, and are causally effective through, such states and variables).

Of course we do not fully understand how the mind is able to plan and make choices resulting in top-down action as discussed here (Koch 2004, Murphy and Brown 2007). The fact that we do not know how it works does not affect the fact that we are certain it can happen and does happen (Simon 1992, Chapters 5 and 6). Indeed you could not be reading this article if it were not true: the marks on the paper that constitute the letters you are reading have the form they do because of top-down action from my mind to my hand. This article would not exist were this not possible. Thus this form of top-down causation has been demonstrated many millions of times.

Example 4.5.5: Roles, expectations, and values. Social roles are socially determined abstract entities that are causally effective in structuring society. They are a key aspect of the way individual behaviour links in to the social environment. Roles are developed by an adaptive process which is a combination of bottom up and top down interaction between society and the individuals who make up the society (Berger 1963). They are

---

[57] http://en.wikipedia.org/wiki/Money.

then inculcated into the individual by top-down social processes (Berger and Luckmann 1967; Cacioppo et al, 2002), whereafter they become a core feature of individual psychology in relation to society (Longres 1990), together with expectations guiding the choice of goals and actions and hence being causally effective in a top-down way from the mind to the body.

Roles embody social values, which, together with individual values relating to life purpose, guide the individual and communal choice of goals and the methods used to attain these goals. Thus the highest level adaptive goals are *values*, related to ethics, aesthetics, and meaning, which are all causally effective in a top-down way by determining the set of desirable lower level goals (Murphy and Ellis 1995). The imperative to search for meaning is a key aspect of human nature (Frankl 2004), without which for example the entire edifice of science would not exist. Our understandings of meaning and purpose are abstract entities that form the highest level in the hierarchy of causation in the mind (Murphy and Ellis 1995). The related ethical values are non-reducible higher level variables; by determining the nature of acceptable lower level goals, they are a set of abstract principles that are causally effective in the real physical world, indeed they crucially determine what happens. For example wars will be waged or not depending on ethical stances; large-scale physical devastation of the earth will result if thermonuclear war takes place, so their nature has crucial effects in the way human activity impacts on society and the world.[58] The question of what are the best fitness criteria for values, is one of the key issues for humanity.[59]

Quite different from all these examples is the following:

Example 4.5.6: ***Mathematics Comprehension and Utilisation*** is a case of top-down causation from a Platonic world of mathematical abstractions to the human mind, being realized in details of neuronal connections, and then into the real world where it is causally effective in terms of creating patterns on paper and through underlying physics, engineering, commerce, and planning in general. How this happens is not clear (Penrose 1997, 2005; Changeux and Connes 1998).

The existence of a Platonic world of mathematical objects (Ellis 2004) is strongly argued by Penrose (Penrose 1997, 2005) and Connes (Changeux and Connes 1998), the point being that major parts of mathematics are discovered rather than invented (rational numbers, zero, irrational numbers, and the Mandelbrot set being classic examples). They are not determined by physical experiment, but are rather arrived at by mathematical investigation. They have an abstract rather than embodied character; the same abstract quantity can be represented and embodied in many symbolic and physical ways (Penrose 2005), and these representations form an equivalence class. They are independent of the existence and culture of human beings; it is plausible that the same features will be discovered by intelligent beings in the Andromeda galaxy as here, once their mathematical understanding is advanced enough (which is why they are advocated as the basis for inter-stellar communication).

---

[58] For an in depth examination of a specific example, see Rupert Smith (2006):
[59] For one view on this, see Murphy and Ellis (1995).

This Platonic world is to some degree discovered by humans, and represented by our mathematical theories; that representation is a cultural construct, but the underlying mathematical features they represent are not - indeed like physical laws, they are often unwillingly discovered, for example irrationality of the $\sqrt{2}$ and the number π. This world is causally efficacious in terms of the process of discovery and description: one can for example print graphic versions of the Mandelbrot Set[60] in a book, resulting in a physical embodiment in the ink printed on the page. The causal variables here are not coarse grained lower level variables and exist independent of the mind, even though they are discovered and comprehended by the mind.

Example 4.5.7: ***Physics Theories***: Maxwell's *theory* of electromagnetism (an abstract entity, described by Maxwell's equations[61]) led to the development of radio, and then to existence of cell phones, TV, and so on, based in manipulation of physical materials composed of atoms and electrons. Maxwell's theory is not the same as any single person's brain state. It can be represented in many ways (on blackboards, in print, on computer screens, in spoken words) and many formalisms (via 3-dimensional vectors or 4-dimensional tensors, for example); these various representations together form an equivalence class, as they all lead to the same predicted outcomes.

How do you demonstrate this kind of top-down causation? - design an artefact such as a cell-phone through use of Maxwell's theory, and then construct it and operate it. The abstract theory will have altered physical configurations in the real world, and hence is causally effective. It is an irreducible higher level variable (it cannot be derived by coarse-graining any lower level variables) representing the nature of physical reality.

5 FREEDOM AT THE BOTTOM?

I have claimed here that top-down causation is causally effective, which means that even in principle, micro-level laws fail to fully determine outcomes of complex systems: causal closure is achieved only by appealing to downward causation. But this claim is clearly in trouble if the system is already causally closed at the micro-level, as is supposed by most physicists. For higher levels to be causally efficacious over lower levels, there has to be some causal slack at the lower levels, otherwise the lower levels would be causally over-determined.[62]

5.1 ALLOWING TOP-DOWN CAUSATION
Where does the causal slack lie? Three key features are relevant.

Firstly, in considering specific physical and biological systems, it lies partly in the *structuring of the system* so as to attain higher level functions, for example the specific connections in a computer (which could have been different) act as constraints on lower

---

[60] http://en.wikipedia.org/wiki/Mandelbrot_set.
[61] http://en.wikipedia.org/wiki/Maxwell%27s_equations.
[62] http://www.iscid.org/encyclopedia/Physical_Causal_Closure_Thesis;
    http://www.meta-religion.com/Philosophy/Articles/Other/causal_closure.htm.

level dynamics (Juarrero 2002), thus channeling how they function; and partly in the *openness of the system*: new information can enter across the boundary and affect local outcomes. For example, cosmic rays may enter the solar system and alter the genetic heritage of individual humans; alteration in solar radiation can cause climate change on earth; telephone calls from afar convey vital information that changes how we act. This is top-down causation from the overall context to the system.

Local systems are not isolated in either space or time, and their future evolution cannot be predicted from their internal properties alone. Thus for example weights of network connections shape the outcome of neural net and brain functioning, and these weights are developed through network training using external information. The `neurons' do not function independent of context. Such structural and boundary effects are the core of TDC1 and TDC2 (Figure 4).

However this does not work in the same way on the largest scales: one can always consider a bigger system, including more and more of the universe within its boundaries, until at the cosmological scale we consider all that exists and there is no longer a possibility of such boundary effects occurring. The system considered includes all causally connected events; there is no larger physical context to take into account.

The second key feature (Luisi 2002) is that *top-down causation changes the nature of the lower elements*. There is not just a situation of invariant lower level elements obeying physical laws; rather we have the nature of lower level elements being changed by context.[63] Often the way this occurs ensures that the way the lower level elements obey physical laws fulfils higher level purposes. This is then an aspect of adaptive selection.

Example 5.1: Chemical bonding. A hydrogen atom in isolation is quite different from one covalently bonded with oxygen to form water (Grey 1973). The orbital structure is quite different and the hydrogen atom has therefore different properties than in isolation; indeed it is really to some degree a misuse of language to call it hydrogen when bonded.

Example 5.2: The neutron has a half life of 11 minutes in isolation, decaying to form a proton, an electron, and an anti-neutrino. It is stable with a half life of billions of years when bound in a nucleus. Its properties therefore are dramatically different in these different contexts.[64]

Example 5.3: Cell differentiation. Through the processes of developmental biology (Gilbert 2006), cells get differentiated to perform specific functions;[65] this changes their nature relative to other cells in an adaptive way. Cells differentiate into neurons that get

---

[63] One could argue that it does not change the nature of those elements, but rather some of their properties, through their now different relationship with their surroundings; but in my view it is precisely those properties that characterise their nature (what a thing is, is defined by how it interacts with other things). Strictly speaking, this means that what was a neutron in a free state is really no longer a neutron when it is bound in a nucleus. It has assumed a new identity, or perhaps even lost its identity, as it has become part of a collective structure.

[64] http://en.wikipedia.org/wiki/Neutron.

[65] http://en.wikipedia.org/wiki/Cellular_differentiation.

adapted to their location in the brain, into muscle cells adapted to their role in the heart, and so on. They each develop so as to fit into their allotted role in the body.

Example 5.4: Humans in society. Individual minds develop in the context of their interactions with other minds, and brain development cannot be understood outside this context (Donald 2001). Individuals are shaped by society so that they fit into that society, for example learning a specific language and a variety of societal roles and expectations (Berger 1963). This is top-down causation from the society to the individual, and indeed to their synaptic connections; one can say that their brain is adapted to fit into the society in which they live (Berger and Luckmann 1967; Cacioppo et al, 2002).

Thus *the nature of micro causation is changed by these top-down processes*, profoundly altering the mechanistic view of how things work. This is a key feature of TDC3 and TDC4 (Figure 4).

Thirdly, the required freedom lies in *micro-indeterminism* (random outcomes of microphysical effects), *combined with adaptive selection*: random outcomes at the micro-level allow variation at the macro-level, which then leads to selection at the micro-level but based in macro-level properties and meaning. Statistical variation and/or quantum indeterminacy provides a repertoire of variant systems that are then subject to processes of Darwinian section, based on higher level qualities of the overall system. This is also a key feature of TDC3 and TDC4 (Figure 4).

For this to work, one needs amplifying mechanisms in order to attain macroscopic variation from quantum fluctuations. Some physical systems (such as photomultipliers and the human eye) amplify quantum effects to a macroscopic scale; some classically chaotic systems can amplify fluctuations in initial data that are of quantum origin;[66] some of the effects captured in Thom's catastrophe theory[67] allow large amplification of microscopic changes; and some molecular biology processes (for example involving replication of mutated molecules) act as such amplifiers (Percival 1991). There is considerable evidence that these kinds of effects lead to indeterminacy in brain and behaviour (Glimcher 2005) At a profound level the universe is indeterministic (Feynman, 1992; Polkinghorne, 2002), allowing the needed causal slack. By itself that does not lead to emergence of higher level order; but it does allow this through the process of adaptive selection (Roederer 2005).

Whether these are sufficient to account for TDC5 is not clear – this is essentially the problem of free will (Pink 2004). It has in essence been suggested that they are indeed sufficient (Murphy and Brown 2007); this is however contested.[68] But in any case the evidence for TDC5 is very strong, as we discuss next, so there must be some way it is possible, even if we do not yet know what that way is. Those who claim physics alone

---

[66] When chaotic systems are quantized, their chaotic behaviour normally goes away, but that is not the context envisaged here.

[67] http://en.wikipedia.org/wiki/Catastrophe_theory.

[68] It has been suggested that the solution of the free will problem is related to the as yet unsolved quantum measurement problem (Penrose 1989, 1997; Stapp 1993); this proposal is mainly regarded with scepticism.

underlies consciousness and our actions should take cognisance of the true difficulty of the `hard problem' of consciousness (Chalmers 1997, Clayton 2006); we do not know how to begin to tackle it.

5.2 EVIDENCE OF HIGHER-LEVEL AUTONOMY

There has to be adequate causal slack because it is needed in order to explain the detailed complexity that exists in the universe. The claim made by bottom-up determinism is *physical causal completeness*: *for any specific physical system, including human minds, physical laws alone give a unique outcome for each set of initial data*. To see the improbability of this claim, one can contemplate what is required from this viewpoint when placed in its proper cosmic context. The implication is that the particles that existed at the time of decoupling of the Cosmic Background Radiation in the early universe (Silk 2001, Dodelson 2003) just happened to be placed so precisely as to make it inevitable that fourteen billion years later, human beings would exist and Crick and Watson would discover DNA, Townes would conceive of the laser, Witten would develop M-theory.

In my view, this is absurd. It is inconceivable that truly random quantum fluctuations in the inflationary era – the supposed source of later emergent structure (Dodelson 2003) - can have had implicitly coded in them the future inevitability of the Mona Lisa, Nelson's victory at Trafalgar, Einstein's 1905 theory of relativity. Such later creations of the mind are clearly not random, on the contrary they exhibit high levels of order embodying sophisticated understandings of painting, military tactics, and physics respectively, which cannot possibly have directly arisen from random initial data. This proposal simply does not account for the origin of such higher-level order. In any case it is not possible because of quantum uncertainty, for two reasons (Ellis 2006b).

Firstly, quantum fluctuations can change the genetic inheritance of animals (Percival 1991) and so influence the course of evolutionary history on Earth. Indeed that is in effect what occurred when cosmic rays[69] – whose emission processes are subject to quantum uncertainty - caused genetic damage in the distant past:

"*The near universality of specialized mechanisms for DNA repair, including repair of specifically radiation induced damage, from prokaryotes to humans, suggests that the earth has always been subject to damage/repair events above the rate of intrinsic replication errors ..... radiation may have been the dominant generator of genetic diversity in the terrestrial past*" (Scalo et al 2001).

Consequently the specific evolutionary outcomes of life on Earth (the existence of dinosaurs, giraffes, humans) cannot even in principle be uniquely determined by causal evolution from conditions in the early universe, or from detailed data at the start of life on Earth. Quantum uncertainty prevents this, because it significantly affected the occurrence of radiation-induced mutations in this evolutionary history. The specific outcome that actually occurred was determined as it happened, when quantum emission of the relevant photons took place: the prior uncertainty in their trajectories was resolved by the historical occurrence of the emission event, resulting in a specific photon emission time

---

[69] See http://www.chicos.caltech.edu/cosmic_rays.html for a brief summary of their origin.

and trajectory that was not determined beforehand, with consequent damage to a specific gene in a particular cell at a particular time and place that cannot be predicted even in principle.

Secondly, according to the standard inflationary model of the very early universe, we cannot predict the specific large-scale structure existing in the universe today from data at the start of the inflationary expansion epoch, because density inhomogeneities at later times have grown out of random quantum fluctuations in the effective scalar field that is dominant at very early times:

"*Inflation offers an explanation for the clumpiness of matter in the universe: quantum fluctuations in the mysterious substance that powered the [inflationary] expansion would have been inflated to astrophysical scales and therefore served as the seeds of stars and galaxies*" (Hinshaw 2006).[70]

Thus the existence of our specific Galaxy, let alone the planet Earth, was not uniquely determined by initial data in the very early universe. The quantum fluctuations that are amplified to galactic scale by this process are unpredictable in principle

For these reasons, it is impossible that the initial data in the early universe uniquely specified any particular human action or thought whatever, as neither the existence o0f humans nor even the existence of our Galaxy is uniquely implied by that initial data. So where do they arise from? The obvious explanation is that they arise from the autonomous behavior of the human mind acting in an intelligent way, supervenient on[71] but not causally determined by the underlying physics. Top-down action from the mind to the electrons in the brain allows this to happen; and there thus has to be the causal openness needed to allow this to occur, for otherwise this demonstrable outcome could not arise. If it can happen in this case, it may be possible in other cases too, see the various examples above.

Finally, one can argue there is good evidence that free will plays an autonomous causal role not determined by physics; and is an important part of the causality in operation (Pink 2004). This is clearly controversial territory, and some deny that free will truly exists. Firstly, Libet's experiments and their interpretation (Libet 2004) has generated lots of discussion and speculation, but the experiments and his interpretation of them are very problematic.[72] Secondly, we should recognize that the enterprise of science itself does not

---

[70] See Dodelson (2003) for details.
[71] http://en.wikipedia.org/wiki/Supervenience.
[72] Robert Bishop comments as follows (private communication): First, the motion of raising a finger or a wrist only goes as high as the lower brain stem, so it's not clear there is much, if any, upshot for free will here. Second, the timing of when a "choice" to raise the finger is made and the subject's reporting when they have made this "choice" depends on their concentrating on watching a revolving clock so as to report the approximate time of their "choice." There simply is no way to conclude that there is something significant here for free will in the timing of the rise of the "readiness potential" about 300ms before the subject reports "deciding" to lift their finger, a time interval that is easily accounted for by both the gross accuracy of the clock the subject is watching and the delay in how quickly they can say that they've chosen. Third, there is nothing purposeful or meaningful about the experimental task. Indeed, it has to be roughly

make sense if our minds cannot rationally choose between alternative theories on the basis of the available data, which is indeed the situation if one takes seriously the bottom-up mechanistic view that the mind simply dances to the commands of its constituent electrons and protons, algorithmically following the imperatives of Maxwell's equations and quantum physics. A reasoning mind able to make rational choices is a prerequisite for the academic subject of physics to exist. The proposal that apparent rationality is illusory, being just the inevitable outcomes of micro-physics, cannot account for the existence of physics as a rational enterprise. But this enterprise does indeed make sense; thus one can provisionally recognise the possibility that free will too is an active causal factor, not directly determined by the underlying physics. Indeed I suggest a stronger statement: if your theory does not allow existence of free will in a serious sense, then it is not a good enough theory – for you cannot engage in scientific activity without it!

6 MULTIPLE CATEGORIES OF CAUSATION

Reductionist analysis `explains' the properties of the machine by analysing its behaviour in terms of the functioning of its component parts (the lower levels of structure). Systems thinking tries to understand the properties of the interconnected complex whole (Churchman 1968, Flood and Carson 1990), and `explains' the behaviour or properties of an entity by determining its role or function within the higher levels of structure. For example, the question: `Why is an aircraft flying?' can be answered,[73]

- In *Bottom-up terms*: it flies because air molecules impinge against the wing with slower moving molecules below creating a higher pressure as against that due to faster moving molecules above, leading to a pressure difference described by Bernoulli's law, this counteracts gravity, etc.;

- In terms of *same-level explanation*: it flies because the pilot is flying it, after a major process of training and testing that developed the necessary skills, and she is doing so because the airline's timetable dictates that there will be a flight today at 16h35 from London to Berlin, as worked out by the airline executives on the basis of need and carrying capacity at this time of year;

- In terms of *top-down explanation*: it flies because it is designed to fly! This was done by a team of engineers working in a historical context of the development of metallurgy, combustion, lubrication, aeronautics, machine tools, computer aided design, etc., all needed to make this possible, and in an economic context of a society with a transportation need and complex industrial organisations able to mobilise all the necessary resources for design and manufacture. A brick does not fly because it was not designed to fly. And why was it designed to fly? Because it will make a profit for the manufacturers and the airline company! Without the prospect of that profit, it would not exist

---

automated, a repetitive task, if the electrophysiological recordings are to be averaged; and this is not the kind of task involved in the genuine exercise of free will.

[73] I am indebted to Russell Ackoff for this example.

These are all simultaneously true non-trivial explanations; *the plane would not be flying if they were not all true at the same time.* The higher-level explanations involving goal choices rely on the existence of the lower level explanations involving physical mechanisms in order that they can succeed, but are clearly of a quite different nature than the lower level ones, and are certainly not reducible to them nor dependent on their specific nature. The bottom-up kind of explanation would not apply to a specific context if the higher-level explanations, the result of human intentions, had not created a situation that made it relevant.

This situation was captured by Aristotle through his proposal of four different kinds of causation. According to Falcon (2006),[74] they are**:**

- ***The material cause***: "that out of which", e.g., the bronze of a statue
- ***The formal cause***: "the form", "the account of what-it-is-to-be", e.g., the shape of a statue.
- ***The efficient cause***: "the primary source of the change or rest", e.g., the artisan, the art of bronze-casting the statue, the man who gives advice.
- ***The final cause***: "the end, that for the sake of which a thing is done", e.g., health is the end of walking, losing weight, purging, drugs, and surgical tools

The last is a *teleological explanation* - an explanation that makes a reference to *telos* or purpose. Additionally, circular causation is possible: things can be causes of one another - a relation of reciprocal influence.

These four kinds of causes correspond broadly to those identified above in the case of the aircraft flying. Indeed we can adapt Aristotle's categorisation to the hierarchical context considered here, by seeing the Material Cause as the lower level (Mechanical) cause, the Formal Cause as the same level (Immediate) cause, the Efficient Cause as the immediate higher (Contextual) cause, and the Final Cause as the ultimate higher level cause.[75] The key point is that all four kinds of causation will be at work in all everyday circumstances. These all take place in a context, and one cannot sensibly discuss their nature out of that context. There are four kinds of contexts for physical processes, as mentioned in the introduction: the natural world, life, humanity, and artefacts. How they work out at the macro level is indicated in Figure 7a, and at the micro-level in Figure 7b. Information comes in as a causal factor in the last three contexts, but not the first one (unless quantum theory is based in information, as hypothesized by Wheeler). The key point about causality in real-world contexts, then, is that simultaneous multiple causality (inter-level, as well as within each level) is always in operation in complex systems.

Example 6.1: ***Lock and Key.***[76] A key fits a lock because of simple mechanical principles (the material cause). But these generic principles do not explain why this particular key fits this particular lock. That is a result of human intention and design of this particular

---

[74] See also http://en.wikipedia.org/wiki/Aristotle.
[75] For a more complete characterisation of types of causes, see Auletta (2007).
[76] I thank Guenter Mahler for this example.

lock and this particular key, that fits only this lock (Efficient cause), and that in turn is done to ensure safety of goods and persons (Final cause). And the lock opens the door when it is inserted and turned (Formal cause).

Example 6.2: *Physics Experiments* : Successful completion of a physics experiment, such as observing particle production in a collider,[77] involves all the re-interpreted Aristotelian forms of causation. The Material (Mechanical) cause is the particle interactions that lead to the production of new particles. The Formal (Immediate) cause is that the experimenter turns the accelerator and measuring equipment on at a particular time. The Efficient (Contextual) cause is that the collider was designed and manufactured so that the collisions would take place and outcomes could be observed. The Final Cause might simply be that the experimenter wanted to understand the collision in the context of a theory of AdS/CFT duality, or it might be because she aspired to attaining a Nobel Prize.

Thus one can have top-down system explanations as well as bottom-up and same level explanations, corresponding to the four kinds of Aristotelian explanation, and all being simultaneously applicable. Indeed there will be numerous causal factors in any specific case forming a network of causes, including those identified here but also the overall historical and physical and social environment, without which the identified events would not take place (for example, the laws of physics are as they are, the Earth exists, scientists are able to do experiments, measuring apparatus can be devised reliably, and so on). An explanation usually takes most of this for granted and focuses just on one or two items that are the subject of attention, perhaps because they can be manipulated to alter the result. The point of this paper is to identify top-down or whole-part effects as part of this causal network, in addition to bottom-up and same level effects.

7 CONCLUSION

The fact that physics is not the only form of causation in the real world has been demonstrated above by numerous examples. *Physics provides necessary conditions (but not the sufficient conditions) for what happens; it provides the possibility space for what happens, but does not determine the outcome.* Top-down causation allows higher level causes to be what they appear to be: real effective causes. Context is the key to physical outcomes: multiple causation is always at work. Random fluctuations plus quantum uncertainty provide the freedom at the bottom needed to allow this to happen. It enables the causal power of *abstract entities*: mathematics, theories, ethics, social constructs, and underlies the paradox of the experimenter in physical science: all scientific experiments are based on purposeful activity and free will, enabling decisions based in abstract analysis that lies beyond the explanatory scope of physical science.

The different kinds of top-down causation distinguished here themselves form a hierarchy, with the higher levels largely dependent on the lower levels for their ability to function (Figurer 4). The central feature of all types of downward causation is *functional equivalence classes*, as these are the higher level entities that are causally effective. This

---

[77] http://en.wikipedia.org/wiki/Collider.

applies in each of the cases distinguished. One of the major reasons for use of equivalence classes is that it enables reliable behaviour to emerge from noisy or unreliable components together with statistical fluctuations and irreducible quantum uncertainty at the bottom levels. A key issue is that there are effective higher level variables (equivalence classes of phenomena) that cannot be obtained by aggregation of lower level variables. These will usually in some sense convey information. Examples are the folding patterns of RNA and proteins, sensory data such as smells and images, information coded in languages, and theories of physics such as the theory of the laser. These underlie top-down causation that clearly cannot be reduced to any lower level functions (although they are based in such functions) for the relevant control variables simply are not expressible in lower level terms, or even in aggregations of lower level variables.

Information control occurs when there is top-down causation that fulfils some higher level purpose or goal: that is, the outcome is causally affected by, and indeed predictable in terms, of that higher level goal. Here *goals* are by definition conceptual or potential states that control lower level action through feedback control loops. The higher level goals cannot be described in terms of lower level goals or concepts, and indeed often cannot be obtained by coarse graining of lower level variables. It is in this case that it entails a higher level meaning. Attractors of a dynamical system are not goals. A control system has an aim or purpose, whereas attractors in dynamical system do not: they are just probable outcomes of a large variety of initial conditions. There is no associated goal or purpose. Goals entail information use of some kind (needed for the feedback signal).

Information is causally effective view *feedback control loops*, because without the checks of outcome involved in such loops (taking into account the actual situation) the attainment of desired goals will be unreliable. It has syntactic, semantic and pragmatic aspects (Kuppers 1990, Roederer 2005). A pre-requisite is reliable recognition and classification of information, hence *pattern recognition* is important. *Patterns* are by definition abstract classes, not the same as any physical state. They may be realized mathematically (algebraically or geometrically), physically, temporally, or functionally. They underlie coding and language. They are causally effective via pattern recognition mechanisms, used to interpret the pattern in a context of meaning. This is a form of top-down causation from an abstract space of patterns to the physical world. Through abstract characterizations on the one hand (e.g. symmetry classes) and alternative codings or languages on the other, patterns come in equivalence classes, which then define higher level patterns (e.g. symmetric and asymmetric patterns). Thus there are *Recursive patterns:* one pattern may be an element in a higher order pattern, giving a hierarchical structure. A key ability is to recognize this, treating a whole set of elements as a unit in a higher order pattern. This enables hierarchical complexity to be built up.

Information selection is a key feature of higher level feedback systems: they must ignore what is irrelevant and attend to what is not.[78] How this happens is different at each level and each timescale. In simple control systems, it happens because the system only responds to what is relevant and is not affected by the rest (a thermostat responds only to

---

[78] For a specific example, see Baumohl (2007)

temperature, for example). Lower level processes that select information are acting as *filters* of signals (they select polarisation angles or energy ranges or frequency ranges, for example). They do not have an explicit goal, but do process signals in a well-defined manner. Signals become information when related to purpose in some context (Roederer 2005). In conscious beings, it is a key component of attention: we discard most of the data coming in and pay attention only to that which is relevant to our immediate safety and purposes. Thus we test it against a matrix of relevance and take notice only of the significant, discarding the rest; in particular the ability to recognize novelty is a key feature of animal brains (Greenspan 2007). In evolutionary history, it is in terms of survival value: we develop genetically-based systems that retain information that will help survival, e.g. a genetic code, plus genes for basic physiological and sensory systems and primary emotional systems. In biochemistry, the binding and recognition of specific molecules is the way molecular information is read (Lehn 1995). This identifies chemical signals as they come along as relevant or not, e.g. specific neuro-modulators and proteins are recognised by specific receptors. But overall in living systems it is via Darwinian processes at all timescales and physical scales, whereby in effect different interpretations of data are tried and most discarded, but the relevant ones (in terms of some higher level selection criteria) are retained and stored in some form or other (Roederer 2005).

Adaptive processes are essential in order to allow new classes of information (and associated structures) to come into being. Adaptive processes do not have a goal, they have a *meta-goal* (such as survival, or specified by some value system) that guides the overall direction of the system but in an open manner. They can function on evolutionary, developmental, and functional timescales.They allow suitable responses to be made to environmental changes (whether these are physical or social) because they embody information about the effect of such responses. Purely bottom-up processes cannot have that character: they can create a certain level of complexity, but cannot adapt their outcome to the environment. Adaptive selection creates new classes of information and new instances of those classes, e.g. the genetic code and DNA that uses that code. By contrast hardwired feedback control systems use specific information (e.g. blood pressure) for specific control purposes (maintaining specific levels of blood pressure) via feedback loops. Adaptive evolution can create such control systems.

In the biological context, we will usually be focusing on one or other of the relevant causal timescales, considering evolutionary, (phylogenetic), developmental (ontogenetic), or functional (physiological) situations. Bottom up and top down causation will be occurring in all those explanatory contexts. I have not explicitly considered the functional emergence of complexity in this paper,[79] but this is obviously related to the issues discussed here;[80] indeed contextual emergence (Bishop 2006, Bishop and Atmanspacher 2006) may probably be regarded as a form of top-down causation. However I will not pursue that issue here.

This paper has focused on top-down causation, because it is the mode of causation that is least considered at present. However it should be emphasized that bottom up, same level,

---

[79] See e.g. Clayton (2006), Clayton and Davies (2006).
[80] See e.g. Laughlin (1998), Luisi (2002).

and top down causation all occur at the same time, in concert enabling the emergence of genuine complexity in modular hierarchical systems. The complex whole arises from the interaction of all these modes of causation. Broadly speaking, same level causation is where the action is; bottom up causation enables it to happen; and top down causation decides what happens. Furthermore, this is true for every level of the hierarchy (except the very top-most and very bottom-most; but we do not know what those levels are). Overall, as stated above, physics provides some of the necessary conditions (but not the sufficient conditions) for what happens. Behavior at the higher levels is consistent with the fundamental laws of physics, but not fully determined by them.

REFERENCES

Alon, U (2007): *An Introductin to Systems Biology. Design Principles of Biological Circuits* (Chapman and Hall/CRC)

Anderson, P W (1972): `More is Different'. *Science* **177**, 377. Reprinted in P W Anderson: *A Career in Theoretical Physics.* (World Scientific, Singapore. 1994).

Auletta, G (2007) How Many Causes There Are? (private communication).

Auletta, G, Ellis, G F R, and Jaeger, L (2007) "Top-Down Causation: From a Philosophical Problem to a Scientific Research Program". arXiv:0710.4235.

Barabási, A-L, and Oltvai, Z N (2004). "Network biology: Understanding the cell's Functional organization". *Nature Reviews: Genetics* **5**: 101-114**.**

Baumohl, B (2007). *The Secrets of economic indicators* (Upper SaddleRiver: Wharton).

Beer, M E, Connors, B W, and Paradiso, M A (2007). *Neuroscience: Exploring the Brain* (Philadelphia: Lippincot, Williams and Wilkins).

Beer, S (1966). *Decision and Control* (New York: Wiley).

Berger, P (1963) *Invitation to Sociology: A Humanistic Perspective.* (Doubleday, New York. 1963).

Berger, P and Luckmann T (1967), *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (Anchor, New York).

Birney, E (2007). "Come fly with us" *Nature* **450**: 184-185.

Bishop, C M (1999). *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press).

Bishop, R. (2006), 'Patching physics and chemistry together', *Philosophy of Science* 72: 710-722. http://philsci-archive.pitt.edu/archive/00001880/.

Bishop, R C (2006). **"**Downward Causation in Fluid Convection" *Synthese* (accepted). http://philsci-archive.pitt.edu/archive/00002933/.

Bishop, R and Atmanspacher, H (2006). "Contextual Emergence in the Description of Properties" *Foundations of Physics:* **36,**1753-1777.

Blanchard, Ph., Giulini, D, Joos, E, Kiefer, C, and Stamatescu, I-O (Eds) (2000). *Decoherence: Theoretical, Experimental and Conceptual Problems* (Berlin: Springer).

Booch, G (1994). *Object Oriented Analysis and design with Applications* (New York: Addison Wesley).

Boulding, K E (1961) *The Image: Knowledge in Life and Society* (Ann Arbour: The University of Michigan Press)

Burnet, F. M. (1959) *The clonal selection theory of acquired immunity*. (Cambridge: Cambridge. University Press).

Buyana, T (1997). *Molecular Physics* (Singapore: World Scientific).

Cacioppo, J T, Berntson, G G , Adophs, R, Carter, C S, Davidson, R J, McClintock, M K, Mcewan, B S, Meaney, M J, Schacter, D L, Sternberg, E M, Suomi, S S and Taylor, S E (Eds), (2002): *Foundations in Social Neuroscience* (Cambridge, Mass: MIT Press)

Campbell, D T (1974): "Downward Causation". In *Studies in the Philosophy of Biology: Reduction and Related Problems*, F J Ayala, and T Dobhzansky (eds). (University of California Press, Berkeley. 1974), pp. 179-186.

Campbell, N A and Reece, J B (2005). *Biology* (San Francisco Benjamin Cummings, New York).

Carpenter, G A and Grossberg, S (1991): *Pattern Recognition by Self-organising Neural Networks* (MIT Press).

Chaikin, P M and Lubensky, T C (2000). *Principles of condensed matter physics* (Cambridge: Cambridge University Press).

Chalmers, D (1997). *The Conscious Mind.* (Oxford University Press, Oxford).

Chandrasekhar, B S (1998). Why things are the way they are (Cambridge: Cambridge University Press).

Changeux, J-P and Connes, A (1998). *Conversations on Mind, Matter, and Mathematics* (Princeton: Princeton University Press).

Churchman, C W (1968): *The Systems Approach.* (New York: Delacorte Press).

Clayton, P (2006). *Mind and Emergence* (Oxford: Oxford University Press).

Clayton, P, and Davies, P (2006). *The Re-emergence of Emergence* (Oxford: Oxford University Press).

Conrad, M (1985). "On design principles for a molecular computer". *Communications of the ACM* **28:** 464-480.

Corning, P (2005). *Holistic Darwinism: Synergy, Cybernetics, and the Bioeconomics of Evolution* (Chicago: University of Chicago Press).

Damasio, A R (2000). *Descarte's Error* (New York: Harper Collins).

Deacon, T (1997). *The symbolic species: the co--evolution of language and the human brain* (London: Penguin).

Dodelson, S (2003). *Modern Cosmology* (San Diego: Academic Press).

Donald, M (2001) *A Mind so Rare: The evolution of Human Consciousness* (New York: W W Norton,).

Drosophila 12 Genome Consortium (2007): "Evolution of genes and genomes on the *Drosophila* phylogeny." *Nature* **450**: 293-218.

Edelman, G M. (1989). *Neural Darwinism: The theory of group neuronal selection.* (Oxford: Oxford University Press).

Ellis, G F R (2002) ``Cosmology and Local Physics''. *New Astronomy Reviews* **46**: 645-658 (gr-qc/0102017).

Ellis, G F R (2004). "True Complexity and its Associated Ontology". In *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity*, Ed. J D. Barrow, P C.W. Davies, and C L. Harper. (Cambridge: Cambridge University Press), pp.607-636.

Ellis, G F R (2005) "Physics, Complexity, and Causality" *Nature* **435:** 743.

Ellis, G F R (2006a) "Physics and the real world". *Foundations of Physics*, April 2006, 1–36 http://www.mth.uct.ac.za/~ellis/realworld.pdf .

Ellis, G F. R. (2006b): "Physics in the Real Universe: Time and Spacetime". *GRG* **38**:1797-1824. http://arxiv.org/abs/gr-qc/0605049.

Ellis, G F R (2006c) `On the nature of emergent reality'. In *The Re-emergence of Emergence*, ed P Clayton and P C W Davies (Oxford University Press), pp. 79-110.

Ellis, G F R (2008), "Fundamentalism in Science, Theology, and Academia". In *Understanding Humans in a Scientific Age*, Ed. M Hewlett, C Knight, and W van Huyssteen (Ashgate).

Ellis, G F R and Toronchuk. J A (2005): "Neural Development: Affective and Immune System Influences." In *Consciousness and Emotion*. Ed. R D. Ellis and N Newton (John Benjamins), pp. 81−119 [http://www.mth.uct.ac.za/~ellis/AND.doc].

Elman, J L, Bates, E A, Johnson, M H, Karmiloff-Smith, A, Parisi D, and Plunkett, K (1998). *Rethinking Innateness: A connectionist perspective on development* (Cambridge, Mass: MIT Press).

Emsley, J (2003). *Nature's Building Blocks: An A-Z Guide to the Elements* (Oxford: Oxford University Press).

Falcon, A (2006), "Aristotle on Causality". Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/aristotle-causality/.

Feynman, R (1992). *The Character of Physical Law* (London: Penguin).

Feynman, R (1995) *Six Easy Pieces* (London: Penguin).

Flood, R L and Carson, E R (1990). Dealing with Complexity: An Introduction to the Theory and Application of Systems Science (London: Plenum Press).

Forester, J (1961). *System Dynamics* (Cambridge, Mass: MIT Press).

Fox Keller, E (2000). *The century of the gene* (Cambridge, Mass: Harvard University Press).

Frankl, V (2004) *Man's Search for Meaning. An Introduction to Logotherapy* (Random House / Rider, London).

Gilbert, S F (2006): *Developmental Biology* (Sunderland, MA: Sinauer Associates).

Glendinning, P (1996). *Stability, instability and chaos: an introduction to the theory of non-linear differential equations.* (Cambridge: Cambridge University Press).

Glimcher, P W (2005). "Indeterminacy in brain and behaviour". *Annual Review of Psychology*, **56**: 25.

Goodstein, D L (1985). *States of Matter* (New York: Dover).

Greenland, S, and Pearle, J (2006): "Causal Diagrams". Technical Report R0332, In *Encyclopaedia of Epidemiology*. http://ftp.cs.ucla.edu/pub/stat_ser/r332.pdf.

Greenspan, R J (2007). *An introduction to nervous systems*. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press).

Grey, H B (1973). *Chemical Bonds: An Introduction to Atomic and Molecular Structure* (Menlo Park: Benjamin Cummings).

Guyton, A C (1977). *Basic Human Physiology* (Philadelphia: W B Saunders).

Harrison, H L and Bollinger, J G (1969): *Introduction to automatic controls* (Scranton: International Textbook Company).

Hartmann, S (2001). "Effective Field Theories, Reductionism, and Scientific Explanation". *Stud Hist Phil Mod Phys* **32:** 267**.**

Hewitt, P G (1998). *Conceptual Physics* (Reading, Mass: Addison Wesley).

Hewitt, P G, Lyons, S, Suchocki, J, and Yeh J (2007). *Conceptual Integrated Science* (San Francisco: Addison Wesley).

Higgins, J C (1976). *Information systems for planning and control; concepts and cases* (London: Edward Arnold).

Hinshaw, G (2006) "WMAP data put cosmic inflation to the test". *Physics World* **19**, No 5, 16-19.

Juarrero, A (2002). *Dynamics in Action: Intentional Behaviour as a Complex System* (Cambridge, Mass: MIT Press).

Kandel, E R, Schwartz, J H, and Jessell, T M. (2000). *Principles of Neural Science* (New York: McGraw Hill).

Koch, C (2004). *The Quest for Consciousness: a neurobiological approach* (Englewood, Colorado: Roberts and Company)

Kuppers, B-O (1994). *Information and the origin of life* (Cambridge, Mass: MIT Press).

Laidler, K J, Meiser, J H, and Sanctuary, B C (2003). *Physical Chemistry*. (Boston: Houghton Mifflin).

Laughlin, R B (2000). "Fractional Quantisation ". *Reviews of Modern Physics* **71**, 863.

Lecuit, T, and Le Goff, L (2007). "Orchestrating size and shape during morphogenesis". *Nature* **450**: 189-192.

LeDoux, J (2002): *Synaptic Self* (New York: Viking).

Lehn, J-M (1995). *Supramolecular chemistry* (Weinheim: VCH).

Leopold, P and Perriman, N (2007). "Drosophila and the genetics of the internal milieu". *Nature* **450**: 186-188.

Libet, B (2004*) Mind Time: The Temporal Factor in Consciousness* (Cambridge, Mass: Harvard University Press).

Longres, JF (1990), Human Behaviour in the Social Environment (Itasca, Illinois: F E Peacock).

Luisi, P L (2002). "Emergence in Chemistry: Chemistry as the embodiment of emergence". *Foundations of Chemistry* **4**: 183-200.

Luisi, P L (2006): *The emergence of life: from chemical origins to synthetic biology* (Cambridge: Cambridge University Press).

Maynard Smith, J and Szathmary, E (1997): *The major transitions in evoplution* (Oxford: Oxford University Press).

Mitchell, Melanie (1998) *An Introduction to Genetic Algorithms* (*Complex Adaptive Systems*)*.* (Cambridge, Mass: MIT Press).

Morowitz, H (2002): *The emergence of everything* (Oxford: Oxford University Press)

Murphy, N (2006) "Emergence and Mental Causation". In Clayton, P, and Davies, P (Eds) *The Re-emergence of Emergence* (Oxford: Oxford University Press), 227-243.

Murphy, N. and Brown, W.S (2007): *Did My Neurons Make Me Do It?: Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will* (Oxford: Oxford University Press).

Murphy, N, and Ellis, G F R (1995). *On the moral nature of the universe* (Minneapolis: Fortress Press).

Oerter, Robert (2006). *The theory of almost everything. The Standard Model: the unsung triumph of modern physics* (New York: Plume/Penguin).

Peacocke, A R (1989). *An introduction to the physical chemistry of biological organization* (Oxford: Oxford University Press).

Pearl, J (1998). "Graphs, Causality, and Structural Equation Models". *Sociological Methods & Research,* **27**: 226-284. ftp.cs.ucla.edu/pub/stat_ser/R253.ps.

Pearl, J (2000). *Causality: models, reasoning, and inference* (Cambridge: Cambridge University Press).

Penrose, R (1989). *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics* (New York: Oxford University Press).

Penrose, R (1997). *The Large, the Small and the Human Mind* (Cambridge: Cambridge University Press).

Penrose, R (2005). *The Road to Reality: A Complete Guide to the Laws of the Universe* (New York: Knopf).

Percival, I (1991) "Schrödinger's quantum cat". *Nature* **351**, 357.

Philbin, T (2003). *The 100 Greatest Inventions of All time* (New York: Citadel Press).

Pink, T (2004). *Free Will: A Very Short Introduction* (Oxford: Oxford University Press).

Polkinghorne, J (2002): *Quantum Theory: A Very Short Introduction* (Oxford: Oxford University Press).

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., Baraba´si, A.-L (2002). "Hierarchical Organization of Modularity in Metabolic Networks". *Science* **297**: 1551-1555.

Rhoades, R and Pflanzer, R (1989. *Human Physiology* (Fort Worth: Saunders College Publishing).

Roberts, E B (Ed) (1991). *Managerial Application of System Dynamics* (Cambridge, Mass:MIT Press).

Roederer, J G (2005). *Information and its role in nature* (Berlin: Springer).

Rutherford, F J and Ahlgren, A (1990). *Science for all Americans*(New York: Oxford University Press)..

Sagan, C (1996), *The Demon-Haunted World: Science As a Candle in the Dark* (New York: Random House).

Scalo, J, Wheeler, J Craig and Williams, P (2001). "Intermittent jolts of galactic UV radiation: Mutagenetic effects" In *Frontiers of Life; 12th Rencontres de Blois*, ed. L. M. Celnikier [astro-ph/0104209].

Schlosser, G and Wagner, G P (2004). *Modularity in development and evolution* (Chicago: University of Chicago Press)

Scott, A (1995). *Stairway to the Mind* (New York: Springer-Verlag).

Shearer, J L< Murphy, A T, and Richardson, H H (1971). *System Dynamics* (Reading, Mass: Addison Wesley).

Shepherd, D M, Martin, D P, Varsani, A, Thomson, J A, Rybicki, E P, and Klump, H H (2006). "Restoration of native folding of single stranded DNA sequences through reverse mutations: An indication of a new epigenetic mechanism". *Archives of Biochemistry and Biophysics* **453**: 108-122.

Silk, J (2001) *The Big Bang*. (New York: Freeman).

Simon, H A (1992). *The Sciences of the artificial* (Cambridge, Mass: MIT Press).

Smith, R (2006) *The Utility of Force*: *the art of war in the modern world* (London: Penguin).

Smolin, L (1997). *The Life of the Cosmos* (Oxford University Press US).

Stapp, H (1993): *Mind, Matter and Quantum Mechanics* (Berlin: Springer-Verlag).

Tanenbaum, A S (1990) *Structured Computer Organisation* (Englewood Cliffs: Prentice Hall).

Van Gulick, R (1995). `Who's in charge here? And who's doing all the work?' in Heil, J and Mele, A (eds) *Mental Causation* (Oxford; Clarendon Press), pp, 233-256.

Va´zquez, A., Dobrin, R., Sergi, D., Eckmann, J.-P., Oltvai, Z. N., and Baraba´ si, A.-L. (2004). "The topological relationship between the large-scale attributes and local interaction patterns of complex networks". *Proc Nat Acad Sci* **101**: 17940–17945.

Watson, J D., Baker, T A., Bell, S P, Gann, A, Levine, M, and Losick, R (2003). *The Molecular Biology of the Gene* (New York: Benjamin Cummings).

Wegscheid, B, Condon, C, and Hartmann, R K (2006): "Type A and Type B RNase P RNAs are interchangeable in vivo despite substantial biophysical differences". *EMBO Reports* **7**: 411-417.

Weigel, D and Dean, C (2002): "Development, evolution and adaptation". *Current Opinion in Plant Biology* **5:** 11-13.

West-Eberhard, M J (2003). *Developmental Plasticity and Evolution* (Oxford: Oxford University Press).

Williams, G C (1992): *Adaptation and Natural Selection* (Princeton: Princeton University press).

Zee, H.-D. (1989). *The Physical Basis of the Direction of Time* (Berlin: Springer Verlag)

Zhang, S and Srinivasan, M (2004): "Exploration of cognitive capacity in honeybees: Higher functions emerge from a small brain". In *Complex Worlds from Simpler Nervous Systems*, Ed Frederick R Prete (Cambridge, Mass: MIT Press) pp. 41-74.

Zurek, W H (2003). "Decoherence, Einselection, and the Quantum Origins of the Classical." *Rev. Mod. Phys*. **75**, 715 [http://lanl.arxiv.org/abs/quant-ph/0105127.]

Version 2008-01-31

Level 8	
	Cosmology
Level 7	
	Astrophysics
Level 6	
	Planetary/Earth science
Level 5	
	Geology
Level 4	
	Material science/ mineralogy
Level 3	
	Chemistry
Level 2	
	Atomic Physics
Level 1	
	Particle physics

Figure 1a: **The natural sciences hierarchy**

Level 8	
	Sociology/Economics/Politics
Level 7	
	Psychology
Level 6	
	Physiology
Level 5	
	Cell biology
Level 4	
	Biochemistry
Level 3	
	Chemistry
Level 2	
	Atomic Physics
Level 1	
	Particle physics

Figure 1b: The life sciences hierarchy

Figure 1*: The hierarchy of structure and causation. This figure gives a simplified representation of this hierarchy of levels of reality (as characterised by corresponding academic subjects) in the natural world (1a) and in living beings (1b). .Each lower level underlies what happens at each higher level, in terms of causation. For a more detailed d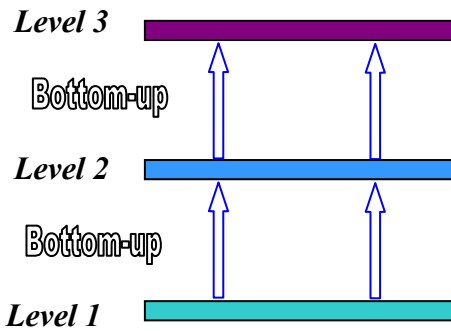escription of this hierarchical structure, see* http://www.mth.uct.ac.za/~ellis/cos0.html*.*

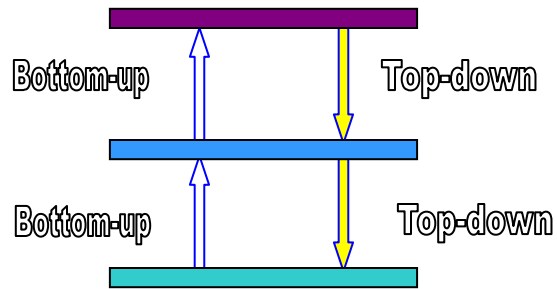**Figure 2a: Bottom-up causation only.**   **Figure 2b: Bottom-up and Top-down causation**

Figure 2: ***Bottom-up and Top-down causation.*** *The fundamental importance of top-down causation is that it changes the causal relation between upper and lower levels in the hierarchy of structure and organisation, cf. the difference between Fig 2a and Fig 2b.*
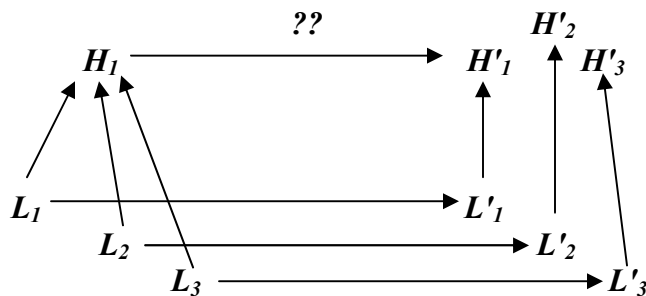

Figure 3a: *The lower level dynamics does not lead to coherent higher level dynamics when the lower level dynamics acting on different lower level states corresponding to a single higher level state, give new lower level states corresponding to different higher level states.*
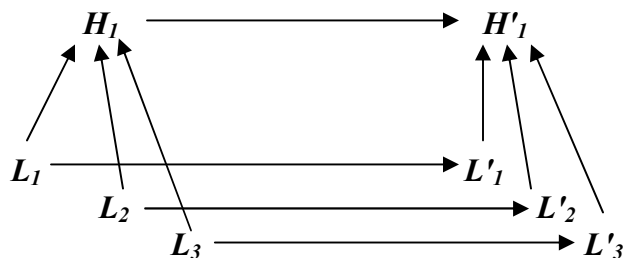

Figure 3b: *The lower level dynamics leads to coherent higher level dynamics when the lower level dynamics acting on different lower level states corresponding to a single higher level state, give new lower level states corresponding to the same higher level state.*

.

TDC5	*Intelligent top-down causation*
TDC4	*Top-down causation via adaptive information control.*
TDC3	*Top-down causation via adaptive selection*
TDC2	*Top-down causation via non-adaptive information control.*
TDC1	*Mechanical Top-down causation.*

Figure 4: *The five kinds of top-down causation identified in this paper. The higher ones depend on the lower ones for their functioning.*
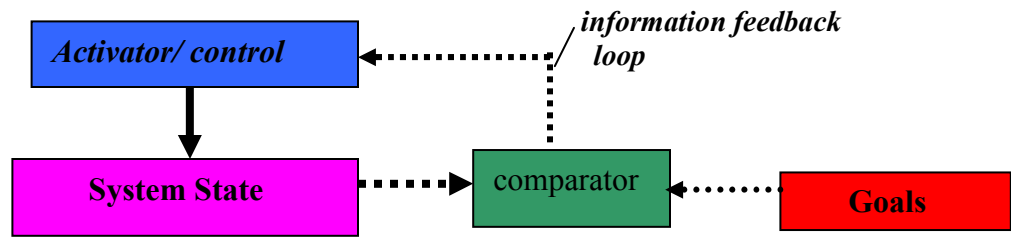
Figure 5: *The basic feedback control process.* *The goals tend to lead to a specific final state via a specific mode of physical action. The initial state of the system is then irrelevant to its final outcome, provided the system parameters are not exceeded.*
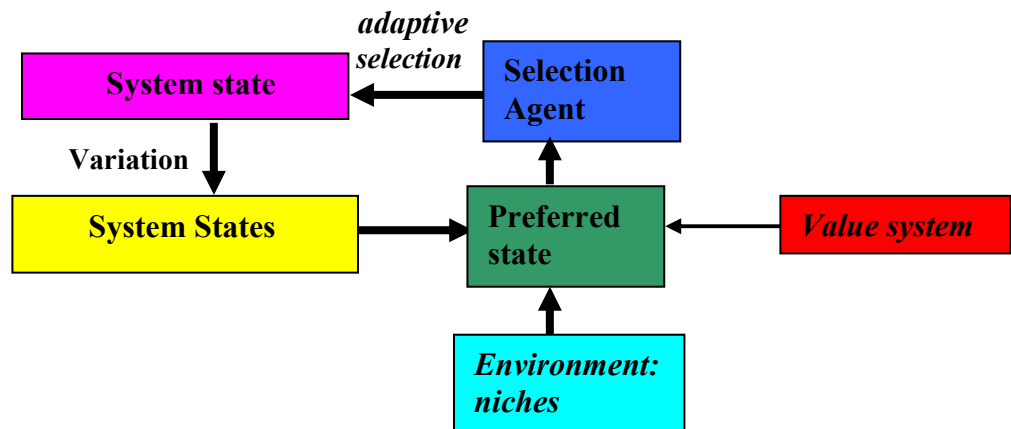

Figure 6: *Adaptive selection*. *The meta-goals embodied in the value system do not lead to a specific final state: rather they lead to any one of a class of states that tends to promote the meta-goals. Thus the final state is not uniquely determined by the meta-goals; random variation influences the outcome by leading to a suite of states from which an adaptive selection is made in the context of both the value system and the environment.*

Causes	1: Natural world	2: Life	3:Humanity	4:Artifacts
Ultimate	???	Adaptive selection	Telos	Purpose
Contextual	Cosmos	Functioning in Ecosystem	Purposes in social setting	Planning and fabrication
Immediate	Astronomical, geological interactions	Birth, life, death	Thoughts, actions	Utilisation
Mechanical	Physical interactions	Bodily functions	Bodily functions	Atoms and forces

Figure 7a: ***The four kinds of causes in relation to context****. Aristotle's four kinds of causation, reinterpreted in the context of the hierarchy of causation, and their nature in the four different physical contexts.*

Causes	Physics	Chemistry	Biochemistry
Ultimate	Variational principles	Energy minimisation	Adaptive selection
Contextual	Specific Laws of physics	Reaction context	Living system
Immediate	Random chance or necessity	Self assembly or chemical transformation	Folding, binding
Mechanical	Particles and forces	Atoms and forces	Molecules and bonds

Figure 7b: ***The four kinds of causes in relation to micro levels****. Aristotle's four kinds of causation, reinterpreted in the context of the basic levels of the hierarchy.*