# Multiscale Analysis of Biological Systems

**Article** *in* Acta Biotheoretica · January 2013

DOI: 10.1007/s10441-013-9170-z · Source: PubMed



**1 author:**

Annick Lesne
French National Centre for Scientific Research
**210** PUBLICATIONS **2,667** CITATIONS

Acta Biotheor (2013) 61:3–19
DOI 10.1007/s10441-013-9170-z

REGULAR ARTICLE

# Multiscale Analysis of Biological Systems

**Annick Lesne**

Received: 12 October 2012 / Accepted: 7 January 2013 / Published online: 18 January 2013


**Abstract** It is argued that multiscale approaches are necessary for an explanatory modeling of biological systems. A first step, besides common to the multiscale modeling of physical and living systems, is a bottom-up integration based on the notions of effective parameters and minimal models. Top-down effects can be accounted for in terms of effective constraints and inputs. Biological systems are essentially characterized by an entanglement of bottom-up and top-down influences following from their evolutionary history. A self-consistent multiscale scheme is proposed to capture the ensuing circular causality. Its differences with standard mean-field self-consistent equations and slow-fast decompositions are discussed. As such, this scheme offers a way to unravel the multilevel architecture of living systems and their regulation. Two examples, genome functions and biofilms, are detailed.

**Keywords** Multiscale approaches · Effective models · Regulation · Circular causality · Self-consistent equations · Integrative biology · Systems biology

## 1 Introduction

Biological functions involve processes at various scales. This statement is obviously true for organismic processes like development, or for a bacterial colony. It is already relevant for intracellular functions like gene expression and signaling pathways. Understanding biological functions requires to *integrate knowledge or*

A. Lesne (✉)
CNRS UMR 7600, Université Pierre et Marie Curie-Paris 6,
4 place Jussieu, 75252 Paris Cedex 05, France
e-mail: lesne@lptmc.jussieu.fr

A. Lesne
Institut des Hautes Études Scientifiques, 35 route de Chartres, 91440 Bures-sur-Yvette, France

*data of various natures*, available at various levels, and described *within various frameworks*, from quantum mechanics (for elementary processes like light transduction) to stochastic kinetics to deterministic rate equations and continuous medium theory (e.g. elasticity theory). Beyond the epistemic issue of capturing a real process in descriptions and measurements prescribed by our own abilities and scale limitations, biological systems present a greater challenge: they are *intrinsically and irreducibly multiscale processes*. Regulation of a biological function has to bridge the state of the cells and some surrounding features with basic mechanisms at the atomic or molecular scale, in an adaptive way. A cell itself has to perform a multiscale integration. For instance, transcription in eukaryotes relies at the same time on information about DNA sequence and bound proteins, histone chemical status (e.g. acetylated or not), chromatin conformation (e.g. condensed and topologically constrained or not), nuclear localization (e.g. near a nuclear pore), cell state and its surroundings; it involves influences from each of these various levels, either directly controlling the polymerase binding and activity, or being mediated by signaling pathways or mechanical constraints. Our analysis and modeling should follow the same line (Lesne et al. 2012; Lesne and Victor 2006). A similar challenge is met in Karsenti et al. (2006) in the context of microtubules assemblies and in Muskhelishvili et al. (2010) in the context of bacterial transcriptional regulation. I will argue that the multiscale organization of biological systems has been established by natural selection and ensuing coadaptation of their various parts and levels. Specific multiscale approaches are thus to be devised for their integrated understanding. A possible one, based on a self-consistent procedure, is proposed in this paper, Sect. 9.

## 2 The Example of Genome Functions

In biological systems, one aims at understanding *functions*, whose fulfillment in general involve specific and adapted features at several scales. Let us consider the example of a genomic process, e.g. replication, transcription or repair. Its initiation and regulation typically involve specific DNA sequences (e.g. TATA box or some other specific binding site), nucleotide relative positioning (e.g. in DNA restriction by a ribozyme), specific nucleosome conformations with spatial alignment of histone residues [e.g. nucleosome gaping (Mozziconacci and Victor 2003) or reversome conformation (Bécavin et al. 2010)], nucleosome positioning, post-translational modifications of histone tails (Jenuwein and Allis 2001), proper conformation of the chromatin fiber and even fiber positioning within the nucleus (Spector 2003). A relevant explanatory model has to somehow keep track at the same time of the chromatin loop location and conformation, nucleosome shape, histone tails chemical status and certain atomic details. This requires to articulate investigations at several scales (Lesne et al. 2012; Lesne and Victor 2006) for instance:

1. all-atoms simulation to determine base-pair or residue precise positioning,

2. molecular mechanics, Brownian dynamics or inverse kinematics to determine the position of linkers or histone tails and account for the presence of the linker histone H1 and histone-tails post-translational modifications,
3. mechanical modeling of the 30-nm fiber and topological features of the chromatin loop (i.e. topologically closed stretch of chromatin fiber delineated by boundaries or insulators) at the scale of about 100 kbps.

The challenge is here to take into account the *reciprocal influences* between the levels, e.g. to describe residue positioning *within the chromatin fiber*.

## 3 The Example of Biofilms

Biofilms are complex structures formed by a (often multispecies) bacterial colony and the matrix it secretes. One of their major functional features is their organization at several scales, from that of a single cell up to macroscopic scales. To understand the growth of a biofilm, its structural properties (density, porosity, thickness), its mechanical properties (attachment/detachment under the action of a flow), or its activity in consuming substrate (e.g. in applications to waste treatments), we have to describe jointly the individual and the global level. Microscopic simulation can be developed to give an explicit basis to effective macroscopic models of biofilms and provide a framework to integrate experimental data (Deygout et al. 2013). Two main types of microscopic simulation are usually considered: *cellular automata*, describing the evolution of the particle contents of microscopic spatial cells according to simple rules involving only the states of the neighboring cells, and *individual-based models*, also called *agent-based models*, prescribing the behavioral rules of each individual particle. Both can be used to integrate microscopic knowledge and e.g. determine the explicit expression of local rate constants, to be plugged in macroscopic models in terms of ordinary or partial differential equations. Such a bottom-up approach is well-suited to take into account refined microscopic mechanisms or complicated geometry at the bacteria level.

The challenge is here to simultaneously take into account top-down influences, for instance to describe the physiological behavior of a bacterial cell *within the biofilm*, with e.g. possible functional changes according to its local surroundings and spatial location.

## 4 Bottom-up Approaches Using Effective Parameters

A first way for implementing a multiscale analysis is to travel across the scales in the bottom-up direction. This way is the standard one in physics and it has produced both technical contributions [for instance the whole domain of statistical physics, see (Castiglione et al. 2008)] and epistemological insights [see for instance (Anderson 1972; Simon 1962)]. A basic notion is that of *effective parameter*, encapsulating the net result of several (complicated or not fully known) mechanisms into a single quantity parameterizing a structural feature, an interaction, or a

contribution to an evolution law at a higher description level. Bottom-up approaches aim at determining the collective behavior of an assembly of elements. Effective parameters play an essential role in this integration, in reducing a wealth of complicated and possibly not fully known microscopic ingredients to a single effective one, *having the same impact at higher scales*. Henceforth, a tractable bottom-up integration will typically be hierarchical, with the first steps devoted to the design of a *minimal model* for the elements, and the next steps to the determination of the emergent features of the assembly. Let us cite a few examples of effective parameters and associated dimensional reduction.

- *Rate constants* involved in chemical kinetics are derived under simplifying assumptions from the stochastic (possibly quantum) analysis of the reaction process. Within Kramers theory, they are expressed as a function of the temperature and the limiting-step free-energy barrier (Hänggi et al. 1990). The concept of reaction rates depending on the temperature has been introduced on empirical grounds by Arrhenius in 1889, about 50 years earlier than Kramers.
- Embedding microscopic fluctuations into a macroscopic dynamic model can be done by adding an *effective noise term* to deterministic ordinary or partial differential equations describing the dynamics at large scale. These stochastic differential equations are called *Langevin equations*, by extension of the equation introduced by Langevin (1908) to account for Brownian motion. This noise term appears as the net result of microscopic degrees of freedom that we do not intend to take into account (Lemarchand et al. 1995). This amounts to cutting off high frequencies and short wavelengths and replacing the detailed description of the most microscopic modes by an effective noise term. Stochastic calculus has been developed to handle such equations (Gardiner 1983), however its efficiency is limited to special kinds of noise terms (white or colored noises). Another way is provided by *stochastic processes*, whose evolution rules are essentially random, for instance a Markov chain whose dynamics is fully prescribed by the *probabilities of transition* between the instantaneous state and the following one. These probabilities appear as the effective parameters of the model. In contrast to Langevin equations, Markov processes can cope with discontinuous forms of noise (Van Kampen 1981).
- An *effective diffusion coefficient* $D_{eff}$ can account in an average way of microscopic heterogeneities that are present within a porous substrate provided they have a finite characteristic size $a$. At mesoscopic scales $dx \gg a$, diffusion is described using a a plain diffusion equation, with a spatially constant diffusion coefficient $D_{eff}$ and simple boundary conditions at the border of the sample (Lesne 2006; Nicholson 2001). Computing an effective diffusion coefficient is an instance of a general method called *homogenization* (Hornung 1997; Torquato 2002), intensively developed in the context of composite materials (Mathias et al. 2006). The intuitive justification has been supplemented with a mathematical analysis to determine what is the appropriate "representative elementary volume", that is, the size $dx$ of the regions considered as the elementary volumes of the homogenized system (Ben Arous and Owhadi 2003) and in which all the microscopic structures and processes will be averaged:

$dx$ has to be large compared to the characteristic lengths of these structures and processes.

- The *relative dielectric constant* $\epsilon_r \approx 80$ accounts for the electrostatic influence of water by replacing $\epsilon_0$ by $\epsilon_0\epsilon_r$ in electrostatic interactions and Maxwell equations. This effective description, termed "implicit solvent" is a good approximation at supra-molecular scales. It fails when only a few hydration shells are involved, as in the computation of RNA tertiary structure or protein-protein interactions. An explicit description of water molecules is then required.
- In some instances a nucleosome can be described as a *solid body* (Ben Haïm et al. 2001). Two parameters are sufficient to reproduce the main structural and elastic features of the 30-nm chromatin fiber: the angle $\alpha$ between the DNA left-handed helical path onto the histone core and the core axis, and the *angle* $\Phi$ *between the incoming and outgoing linkers*. This angle $\Phi$ is a typical example of an effective parameter: it is considered as a tunable quantity in the study of the fiber structural and elastic properties (Ben Haïm et al. 2001; Woodcock et al. 1993) without entering the details of the molecular determinants responsible of its value and variations. It would be another issue to relate the precise description of a nucleosome and its surroundings (ionic strength, presence of histone H1, presence of polyamines, post-translational modifications of histones, non-histone binding proteins) with the value of $\Phi$.
- In more refined functional studies (Bancaud et al. 2007; Mozziconacci et al. 2006; Sivolob et al. 2003; Wong et al. 2007), nucleosome conformational transitions have to be taken into account. It is often enough to consider possible switches between a finite number of conformations, i.e. discrete states (Bécavin et al. 2010) and describing the nucleosome as an effective shape characterized by a few effective parameters remain valid.
- DNA entropic elasticity can be described using an effective continuous model, the *worm-like-chain* model (Kratky and Porod 1949). DNA molecule, although composed of discrete atoms, is identified above nanometer scale with an homogeneous one-dimensional semi-flexible filament and entirely described by its local curvature $\rho(x)$ where $x$ is the arc length. This model involves a single effective parameter, the *persistence length* $\mathcal{L}_p$, defined by the expression of the energy density $\mathcal{E}(x) = kT\mathcal{L}_p\rho(x)^2/2$ at the temperature $T$. This description does not intend to take into account sequence effects, that require a description at a finer scale or an effective description in terms of disorder.

As shown by the above examples, an effective parameter can relate a digital (i.e. discrete) description with an analog (i.e. continuous) one, for instance encapsulate the entropic elasticity of a discrete sequence of DNA base pairs in the persistence length of a continuous elastic filament, or conversely describe the wells of a continuous energy landscape as discrete states (Lesne 2007). Most often an effective parameter should have a specific use: it makes sense only in the dimensionally-reduced description in which it is involved, and with regards to the integrative study in which it is involved. Considering e.g. a solid shape for the nucleosome is relevant on geometric and topological grounds but not necessarily for dynamic studies. An effective parameter does not necessarily share all the interpretations and properties

satisfied by the corresponding bare parameter, if any, despite sharing the same name: An effective diffusion coefficient does not necessarily satisfy the Einstein relation linking diffusion and viscosity. An experimental force-extension curve of a chromatin fiber can be fitted within a worm-like-chain model, considering the fiber as a continuous filament with a single elastic degree of freedom, bending. The fit yields an effective persistence length $\mathcal{L}_{p,eff}$ (Bystricky et al. 2004). But this length $\mathcal{L}_{p,eff}$ should not be confused with the actual bending persistence length $\mathcal{L}_p$ of the fiber when it is described within an elastic-worm-like-rod model (Ben Haïm et al. 2001), that is, as an elastic rod with three elastic degrees of freedom (bending, twisting, stretching). Effective parameters will play a central role in the self-consistent integrative approach proposed in this paper (Sect. 9), as the loci where some top-down control can be specified.

## 5 Top-down Approaches Using Effective Constraints and Inputs

Another way of analyzing a multiscale system is to investigate *top-down* relationships, namely how macroscopic inputs, structures and constraints might affect the features of the constitutive elements and the elementary processes. Such influences are often termed *top-down causation* (Ellis 2005). Again effective quantities are useful to encapsulate in a low-dimensional expression (e.g. a field, a force, a geometrical constraint, an energy landscape, a source term or a boundary condition) involving only a few parameters, a wealth of top-down influences. When external constraints apply via boundary conditions, it may be useful to process these conditions into local prescriptions at work inside the system. For instance, an external concentration (e.g. of oxygen) may be replaced after suitable computations involving assumptions on the diffusivity and decoupling between diffusion and consumption, by an effective distribution inside the system. Effective noise terms can be introduced to account for a variety of ill-identified external influences, for instance an high-dimensional input that we do not want to describe in detail. What matters is only the resulting influence on the system at the chosen level of description.

A benefit of effective descriptions, either reducing the description of the elements (effective parameters) or the description of the surroundings (effective inputs and constraints), is their *parsimony*. Because neither the microscopic details nor the macroscopic surroundings are fully known, often not even all identified nor fixed, the models should involve only a coarse description, so as to avoid over-interpretation or spurious sensitivity of the results to the precise knowledge of what is taken for granted in devising the model. This requirement of parsimony does not intend to mean that the biological reality is bound to be that simple, but that our description has to be *unbiased* and *robust* with respect to an additional detail. This is all the more demanded since the intrinsic variability observed between identical biological systems, e.g. cells within a clonal population, is superimposed to the variability of our observations and fuzziness of our knowledge. Another option would be to explicitly take into account this intrinsic variability in the modeling, which would require a (more complicated) probabilistic setting.

## 6 A Basic Classification: Plain, Critical and Complex Systems

At this point, an essential distinction in the multiscale logic of physical and biological systems has to be underlined.

- In plain (mostly physical) systems, microscopic fluctuations average out. Macroscopic observables can be identified with average quantities and described by deterministic continuous fields, that obey ordinary, partial or integro-differential equations with no mention of an underlying microscopic level. Standard examples are classical mechanics, hydrodynamics (Navier-Stokes equations) or chemical kinetics (mass action law).
- In striking but less frequently encountered *critical phenomena*, fluctuations are enhanced by long-range correlations and persist at all scales up to macroscopic ones. Microscopic fluctuations are thus able to qualitatively modify macroscopic behaviors, typically leading to anomalous laws.
- In biological systems, situation is yet different. Microscopic fluctuations are either *buffered by regulatory circuits*, either exploited and possibly amplified as a *source of variability* feeding selection-driven adaptation mechanisms. In the latter case, microscopic fluctuations potentially have repercussions at all scales. In the former case, they are controlled from above by means of feedback circuits that monitor the microscopic ingredients so as to get the proper macroscopic regime and ensure its maintenance.

This tripartite categorization has a parallel formulation in terms of the relevant number of degrees of freedom. In plain physical systems, *a few collective variables*, defined as averages over the microscopic degrees of freedom, are sufficient to describe the macroscopic behavior. Both critical systems and biological systems depart from this case by exhibiting *very many coupled degrees of freedom*. Averaging is no longer efficient to bring out the essential behavior. In case of critical systems, the efficient strategy is a recursive integration and associated renormalization-group methods (Castiglione et al. 2008; Lesne 1998). In biological systems, I claim that dimensional reduction should be achieved jointly at all scales in a *self-consistent way*, replacing at each level several degrees of freedom by a few effective terms reciprocally coupled (Sect. 9).

Biological systems are a specially wide and important instance of *complex systems*, which can be generally defined as assemblies of interacting elements where emergent features directly or indirectly *modify* the elements. Such a logical scheme is often termed *circular causality* (Muskhelishvili et al. 2010). This term underlines the coupling of bottom-up and top-down relationships, leading to self-organized and possibly adaptive behaviors. Typically, elements collectively modify their surroundings, in a way sufficient to influence back the elementary interactions, which in turn may change the collective behavior of the assembly. Only a few purely physical systems display such features, for instance sand dunes (Hersen et al. 2004) or coast reliefs (Werner 1999). In the case of sand dunes, the complexity originates in a change in the interaction between the wind and the sand heap when the heap has passed some critical size; the heap then behaves as a whole, a "dune", able to modify the wind in its vicinity. It is then qualitatively different from a small heap of

grains which only experience the influence of the wind without exerting any feedback on it.

## 7 Mean-Field Self-Consistent Approaches

My proposal is that a good strategy to solve the chick-and-egg problem raised by the circular causality of living systems is to bridge in a self-consistent way the bottom-up and top-down effects. This does not sounds new: in statistical physics, self-consistent equations are encountered for a long time in *mean-field approaches*, in which the influence onto a given microscopic elements of all the other ones is expressed as a function of some average quantity. However, the circular causality apparent in their formulation is far different from the stronger instance observed in living systems. Let us substantiate this claim by discussing the historical example of the mean-field study of ferromagnetism.

A ferromagnetic system is described as an assembly of $N$ spins located at the nodes $i$ of a spatial lattice, each experiencing pairwise alignment interaction with its nearest neighbors and possibly the influence of an external magnetic field $h_0$. Our macroscopic perception of this system of interacting spins reduces to the overall magnetization $M$. In the standard mean-field approximation, correlations between the spatial heterogeneities of the spin orientations are ignored. Accordingly, macroscopic observables describing the collective behavior can be identified with statistical averages of the microscopic features according to the law of large numbers, that is, $M \approx N\langle s \rangle$. The trick is to replace the resultant influence on a given spin $s_i$ of all its neighbors by that of a *mean field* $h_{eff}(M) = aM$ where $a$ is some multiplicative factor depending on the interaction strength. Henceforth spins can be formally considered as independent entities, whose average magnetization $M$ is given by a well-known formula of the form $M = B(h)$, see e.g. the textbook (Laguës and Lesne 2011). Plugging in the field expression $h = h_0 + h_{eff}(M)$ yields a self-consistent equation $M = B(h_0 + aM)$. In considering that the same uniform and deterministic field $h_{eff}(M)$ applies identically onto each element, local fluctuations and correlations between the elements are neglected. Accordingly, the validity of mean-field approaches fails in case of long-range correlations generating fluctuations at all scales. Such critical situations require renormalization-group approaches (Castiglione et al. 2008; Lesne 1998).

This example introduced a general method to determine the macroscopic behavior of an assembly of interacting elements, termed a *mean-field approach*. Formally, the behavior of the system can be properly formulated via *self-consistent equations*, expressing the identity of the macroscopically observed quantity $M$, and the collective quantity $\mathcal{M}_N[s_1(M), \ldots, s_N(M)]$ resulting from the interaction of $N$ elements with state $s_i$. The dependence $s_i(M)$ of the individual states $s_i$ on $M$ is an *approximation* of the resultant influence on a given element $i$ of all the other ones. I underline that *only interactions between the elements are at work* in such systems. The expression $s_i(M)$ provides a closure relation yielding a self-consistent equation involving only the macroscopic variable $M$ : $M = \mathcal{M}_N[s_1(M), \ldots, s_N(M)]$. It has the mathematical expression of a fixed-point equation: $M = \Phi(M)$. In practice,

when a direct analytical resolution is too tricky, the solution can be obtained recursively, given an initial value $M_0$ and computing $M_{n+1} = \Phi(M_n)$, provided the recursive scheme converges; the limit is the desired value $M$.

## 8 Slow-fast Decomposition

The term "multiscale analysis" is sometimes restricted to a more specific method encountered for long in dynamical systems theory. This method is devised to solve a special kind of coupled evolution equations where a small parameter $\epsilon \ll 1$ enforces a time-scale separation between slow and fast components in the dynamics. The common physical wisdom identifies slow variables to macroscopic observables and fast variables to microscopic degrees of freedom, however this intuitive interpretation is not necessarily true for living systems.

The typical instance is an ordinary differential equation where a multiplicative small parameter appears in front of a highest-order time derivative:

$$\begin{cases} \epsilon dx/dt = f(x, Y) \\ dY/dt = g(x, Y) \end{cases} \tag{1}$$

The dependence on $\epsilon$ is termed *singular*, insofar as the behavior of the solution for $\epsilon \to 0$ qualitatively differs from the behavior of the solution for $\epsilon = 0$. Equations of this kind are known for long in control theory (the slow variable $Y$ is controlled by the experimenter and the slaving of the fast variables $x$ to $Y$ is designed by the engineer) and in physics (where the slow variable $Y$ is some macroscopic observable and the dependence of the fast variables $x$ onto $Y$ is usually derived using mean-field-like arguments). The idea is to exploit time-scale separation to decouple slow and fast variables and get formally independent time evolutions. The first step is to solve the equation for the fast variable $x$ at fixed value of $Y$. This step, called a *quasi-static* or *parametric* approximation since $Y$ is considered as a constant parameter, yields the value $X(Y)$ such that $f[X(Y),Y] \equiv 0$. Then, focusing on the slow behavior, we plug the stationary value $X(Y)$ into the evolution of $Y$ to obtain a closed equation:

$$dY/dt = g[X(Y), Y] \equiv G(Y) \tag{2}$$

This second step is called a *quasi-stationary* approximation since it amounts to identify the fast variables with their stationary value (i.e. such that the right hand-side $f$ of $dx/dt$ vanishes) in the slow evolution. The small parameter $\epsilon$ re-enters when, in a possible third step, one looks for the detailed evolution of the fast variable $x$ (Lesne 2006; Nayfeh 1973). Fast variables elimination when the fast dynamics is non-autonomous can be found in Artstein (1999).

Based on scale separation, slow-fast decoupling methods may be seen as a temporal analog of a mean-field approach (Lesne 2006). Plugging the closure relation $x = X(Y)$ into the dynamics of $Y$ yields a dimensionally reduced self-consistent equation for the macroscopic variable $Y$. These methods illustrate the general philosophy according which reduction of the dimension of the complete description and extraction of a relevant macroscopic description may be achieved in

describing only quantities varying at macroscopic scales (Castiglione et al. 2008; Givon et al. 2004). In a biological context, this approach has been implemented e.g. in the study of biochemical networks (Radulescu et al. 2008). Often a clue is to consider the evolution of quantities averaged in phase space, for instance the moments of the microscopic variables (Dieckmann and Law 2000) or aggregated variables associated with a coarse-grained version of the initial model (Auger et al. 2008; Gaveau et al. 1999). Time-scale separation is also a strong argument to eliminate fast oscillating variables $X$ by averaging their oscillations at fixed $Y$ and plugging their average value $\langle X \rangle_Y$ in the equation describing the evolution of the slow variable $Y$ (Bogoliubov and Mitropolskii 1961; Nayfeh 1973; Sanders et al. 2007). Derivation of the dimensionally-reduced dynamics for time-averaged variables is described e.g. in Acharya and Sawant (2006).

In the above analysis, self-consistent equations are given. In practice, it is another matter to get there and to delineate slow and fast variables. I claim that a radically new approach is necessary to fully capture the multiscale logic of living systems. Indeed, due to its evolutionary imprint, the articulation between the levels of organization of a biological *system* cannot in general be expressed as a system of coupled equations for the local and the global variables. Each level often requires to be described in its own formalism, thus reducing the applicability of standard slow-fast decomposition methods. Singular perturbations and decoupling of slow and fast variables remain relevant in a biological context in cases where a local mechanism or functional process can actually be described using a coupled set of *differential equations* involving a small parameter in the above singular way. A well-known example is the derivation of Michaelis-Menten kinetics for enzymatic catalysis; the small parameter $\epsilon$ is there related to the ratio of enzyme concentration to substrate concentration (the enzyme is present in minute amount), and the fast variable is the concentration of the intermediary complex made of the enzyme and the substrate (Lesne 2006; Murray 2003). Other applications can be found e.g. in the context of population dynamics (El Hajji and Rapaport 2009).

## 9 Unravelling the Coevolved Consistency of Living Systems

Considering now living *systems*, the relationship between the state or evolution of the elements and the higher-level processes or superstructures cannot, most often, be expressed as a system of coupled equations for the individual and collective variables using mean-field-like arguments. Natural selection has been at work, at the organism and even at the species level, to produce the observed functioning of living systems. The ensuing coevolution of the various components of organisms led to the appearance of tinkered entities (like allosteric enzymes) and coordinated processes, which cannot be understood using only physical arguments. I mean here that their appearance do not correspond to any reasonably probable physical event. This evolutionary design of living systems biased the generic physical behavior and harnessed the physical laws to produce adapted and regulated organisms. Although never being in contradiction with the laws of physics (this question was warmly

debated in the past), living systems differ from inanimate systems in an irreducible way (Polanyi 1968; Schrödinger 1944).

To substantiate these claims, let us go back to the example of ferromagnetism (see Sect. 7). A spin does not "know" that it participates to the mean-field $h_{eff}$; it only experiences interactions with its nearest neighbors. In contrast, although the overall conformation of a chromatin loop is determined by the assembly of nucleosomes and DNA, this loop actually behaves as an autonomous entity, endowed with a topological invariant, its linking number. Due to this conservation law, any local DNA or nucleosome modification will be perceived at any other place. Another example is transcriptional regulation in a bacterial cell (Muskhelishvili et al. 2010). The cell is sensitive at the level of its metabolism to a change in the environment. It experiences a metabolic switch which in turn controls an adapted switch in gene expression via a coordinated interplay between the pattern of DNA-bound proteins, the molecular composition of the polymerase and the spatial distribution of DNA supercoiling. The association of these factors is an instance of coadaptation achieved by means of natural selection, with no physical necessity In living systems, the evolutionary imprint reflects in a stronger downward causation than in physical systems. As if, in ferromagnetism, some source of variability would have produced different values for the spins and natural selection would have selected the distribution of values yielding the best magnet (best in a sense to be specified according to the biological context, either the most efficient, or the most robust, or the most adapted, or the most tunable).

My point is that self-consistent approaches can be extended far beyond mean-field models and tailored to capture the above specificity of living systems. This point has been addressed several times on epistemological grounds, e.g. (Ellis 2005; Polanyi 1968) but more rarely as a practical methodology. Let us explain the scheme in general terms before giving some examples. Effective parameters involved in the minimal modeling of elementary actors (Sect. 4) are seen as the loci where some functional knowledge about top-down regulation can be injected: dependence of effective structural features (e.g. the gaping angle) of the nucleosome onto the overall conformation of the chromatin fiber and higher levels of organization, dependence of the effective parameters describing the nucleosome structure on the level of supercoiling of the chromatin fiber, dependence of the effective bacterial growth rate onto some features (assumed or measured) of the biofilm. In a complementary step, some prior analysis or experimental knowledge is used to delineate how effective constraints or inputs encapsulating top-down influences in a compact way depend on the elementary features and mechanisms. In short, each level embeds entries that are fed by the other levels, and produces outputs feeding the other levels, as represented in Fig. 1. Running the different parts of the model in following this loop should yield a stable picture, in which inputs and outputs of each part are no longer updated. This achieves the integration of bottom-up and top-down influences and yields an "operationally closed" description, according to the terminology used in (Muskhelishvili et al. 2010).

In the case of biological systems, in contrast to self-consistent approaches encountered in physics, the relations $h = H(r)$, Fig. 1, expressing how the features $h$ of emergent processes and superstructures depend on the effective parameters

$r$ describing the elements, do not simply follow from an averaging approach. Neither do the relations $r = R(h)$, capturing how the elements $r$ are affected by the higher-level features $h$, follow from a mean-field approximation. They involve a wealth of intermediary mechanisms and ingredients, all devised and tuned by numerous runs of natural selection. In particular, these two relations are not simply the inverse one of the other (in which case the self-consistent scheme would be a mere tautology). Part of the articulation may be done "by hand" using some biological knowledge, mainly the identification of the coordinating go-between (Bécavin et al. 2010; Lesne et al. 2012; Malo et al. 2010, 2012) allowing us to bridge in a mechanistic way what is otherwise a statistical relationship (i.e. a correlation). This bridge, termed *structural coupling* in Muskhelishvili et al. (2010), amounts to a short-cut of the (often contingent) evolutionary history. I underline that the aim is to devise a self-consistent *scheme*, not necessarily within the restricted setting of self-consistent equations. Each level may demand to be described in its own relevant formalism or theoretical framework. Accordingly, the microscopic variables $r$ and the macroscopic variables $h$ may well belong to different settings, e.g. the variables $r$ to a molecular dynamics or multi-agent simulation, and the variables $h$ to a partial differential equation. The proposed approach may not be tractable in any situation, since it requires as an input in the modeling some prior biological functional knowledge and understanding of the elementary biological mechanisms. It rather arrives in a second step, when several pieces of understanding at different levels are gathered and need to be articulated in an integrated scenario.

- A first and already mentioned example is provided by the topological constraints experienced by a DNA molecule when its ends are anchored. Its linking number $L_k$ is conserved and provides a global feature of the molecular assembly. The overall conformation of the DNA molecule is characterized by its writhe $W_r$, and

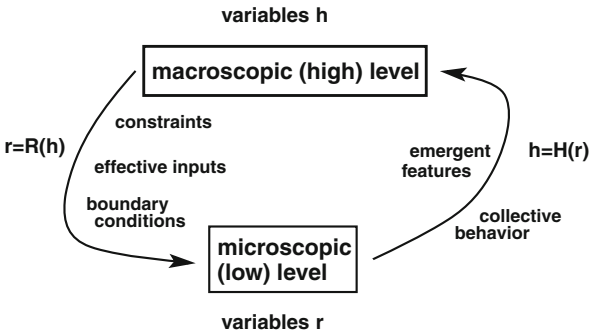


**Fig. 1** Proposed self-consistent scheme for modeling biological systems. The expression $h = H(r)$ indicates in a compact way how the features $h$ of emergent processes and superstructures depend on the effective parameters $r$ describing the elements in a minimal way. Similarly $r = R(h)$ summarizes how the elements are affected by the higher-level entities. They are not the inverse one of the other. The terms "microscopic" and "macroscopic" denote two different levels of organization, e.g. individual and population level (e.g. bacteria and biolfilms), or molecular and organite or cell level (e.g. DNA and chromatin fiber)

$L_k - W_r$ corresponds to its total twist $T_w$. Any local modification of DNA molecule, e.g. a local structural change induced by some protein binding, should be compensated so as to ensure the conservation of $L_k$. Such coordination of local events by global topological constraints is essential in procaryote transcriptional regulation (Travers and Muskhelishvili 2005). The same statement applies to the chromatin fiber in eukaryotes (Lesne and Victor 2006). The elastic coefficients of the chromatin fiber depend on the atomic details of its local architecture. Conversely, fiber supercoiling and balance between toroidal and plectonemic conformations induce mechanical constraints and strains down to the base-pair level.

- We have shown in Bécavin et al. (2010) that RNA-polymerase activity within a condensed chromatin loop is possibly coordinated with a conformational change of nucleosomes downwards. RNA-polymerase activity generates supercoiling $h$ in the chromatin fiber, described as a continuous elastic filament. In turn, supercoiling $h$ induces a conformational change of some nucleosomes (individual state $r$), described as a transition between discrete states (nucleosome and reversome) and measured by the local density of reversomes. These transitions partly relax torsional constraints (i.e. induce a decrease of $h$) while allowing polymerase to process by providing a permissive substrate (polymerase can pass through a reversome), which in turn increases $h$ again. Quantitatively, the self-consistent scheme bridges on the one hand the results of in vitro micromanipulations (Bancaud et al. 2007) and modeling of the nucleosome structure (Zlatanova et al. 2009) with, on the other hand, the activity of RNA-polymerase, generating torsional constraints and requiring a permissive state of the nucleosome. The mediation is achieved by the coordinating influence of fiber supercoiling, and described using a partial differential equation for the propagation of torsional constraints in the chromatin fiber (Bécavin et al. 2010; Lesne et al. 2012).
- The proposed scheme has been implemented in Malo et al. (2010, 2012) to construct an integrated scenario of metastatic escape based on partial in vitro experiments. The principle is to bridge a system of ordinary differential equations describing in a minimal way the physiological switch of the cell between a proliferating and a migrating state, with a multi-agent simulation for the growing tumor, i.e. a population of cells. The coordinating intermediary is here the state-dependent secretion by the cells of proteins able to collectively modify the surroundings and induce a change of the cell physiological state.
- A last example, quite similar to the previous one, is provided by bacterial biofilms (Deygout et al. 2013). Emergent properties of the bacterial population and, above all, the biofilm it secretes may exert feedbacks onto the very behavior of individuals, which conversely control the biofilm composition and geometry. The modeling challenge is to bridge a bottom-up approach, in which some emergent features are derived from a microscopic model, for instance an IBM simulation, with a top-down approach in which some global features control the IBM evolution. This scheme seems similar in spirit to the multiscale simulation proposed by E and Engquist (2003; E et al. 2009) in hydrodynamics. However, the key difference is that top down constraints cannot be computed using some

mean-field argument or more refined closure relations. They originate from a new entity, the biofilm, which does not reduce to the average or collective behavior of the bacterial population, but involves their secretion of a highly structured extracellular matrix. In principle, it could be possible to include the secretion of the bacteria and its regulation (and several other relevant physiological processes) in the IBM. In practice, a realistic simulation of a biofilm involving all the relevant physiological details for the bacterial cells is simply out of reach computationally, and would anyway raise huge difficulties regarding the robustness of the results. We need to recourse to a short-cut, namely devising a minimal model for a bacterial cell, involving only a few effective parameters that at the same time summarize the individual features controlling the biofilm properties and reflect the response of a bacterial cell to the constraints exerted by the biofilm. The difficulty is the identification and computation of these parameters, which essentially involves experimental knowledge and possibly an auxiliary microscopic simulation.The challenge is to quantitatively account for remarkable behaviors such as species coexistence (whereas the macroscopic mean-field equations predict the persistence of a single species and exclusion of the other ones), functional differentiation, spatial segregation, and other complex features of the bacterial population within a biofilm.

Self-consistency implicitly evokes self-consistent equations of physics involving continuous variables or fields. However, in principle, the proposed self-consistent scheme could fit in a discrete setting, as shown also by the above examples involving transitions between discrete states. Part of the difficulty in bridging discrete and continuous views might well originates in our descriptions (Lesne 2007). The Sorites (heap) paradox, like the paradox of Achilles and the tortoise, can be nowadays solved by the mathematical notion of limit, which was unknown to the ancient Greeks. Passing to the limit allows one to reach at infinity an entity of possibly different nature than the steps leading to it, e.g. continuous whereas the steps are discrete (continuous limit, involved e.g. in describing the DNA molecule, made of discrete base pairs, as a continuous elastic filament, Sect. 4), or conversely discrete whereas the steps are continuous (step-wise limit of a sequence of increasingly steeper sigmoidal curves). The notion of emergence can be formulated is such terms: as summarized by Anderson (1972), "more is different". Another clue is to place the description at the unifying level of probability distributions, either for discrete-valued or continuous-valued variables [see (Lesne and Benecke 2008a, b) in which this framework has been developed for the features involved in eukaryote transcriptional regulation]. However, another facet of the difficulty is more intrinsic, and refers to the articulation between digital and analog *informations*. It relates to the question of biological codes, namely information storing, transfer and processing specific to living systems, much discussed since Schrödinger (1944). Inspiring insights about this delicate and still largely open issue as well as a way to solve it in the context of prokaryote transcriptional regulation can be found in Marr et al. (2008; Muskhelichivili et al. 2010).

## 10 Conclusion

The first challenge of a multiscale analysis is to articulate experimental data, knowledge and models available at various levels. The aims are to bridge observations at different scales, to confront observations with mechanisms envisioned at another scale, and to integrate partial models into a consistent and explanatory account of the biological function.

The second and far more difficult challenge is to unravel the multiscale logic of living systems. Strikingly, collective effects may as a whole exert a feedback on the very properties of elementary ingredients and endow them with new functionalities. A specific multiscale approach is thus required to capture the architecture of biological systems and their regulation. A main guideline is their inter-level consistency, reflecting that evolution occurred *jointly* at all the levels of their organization. I propose an approach that makes use not only of effective parameters encapsulating at a given scale lower-level details, but also of effective inputs encapsulating the influences and constraints coming from superstructures and higher-level processes. It is then completed by writing the self-consistency of these reciprocal couplings and influences. This amounts to considering elements and elementary mechanisms within their higher-level context, instead of in isolation. In this way, reductionist and holist views of biological systems are reconciled, which paves the way for systems biology.

I finally underline that the very notion of what is a model prevents from devising an "all-purposes" model and replacing the study of the system by questions addressed on this model. To be fruitful, a model has to be specific to the investigated issue. It should ignore degrees of freedom irrelevant *to this issue* and its characteristic scales. Similarly, a multiscale model should not intend to keep track of all details at all scales but only of the relevant details, whatever their scales, essential to the biological function and its regulation.

**Acknowledgments** This work has been funded by the Agence Nationale de la Recherche, SYSCOMM program, grant DISCO 09-SYSC-003 and by the Institut National de la Santé et de la Recherche Médicale, grant MICROMEGAS PC201104. It also greatly benefited from discussions during the XXXI seminar of the Société Francophone de Biologie Théorique, 15–18 May 2011, Autrans, France, and from the comments of two anonymous referees. I thank all the members of my team "Multiscale modeling of living matter" and especially Jean-Marc Victor for numerous fruitful discussions.

## References

Acharya A, Sawant A (2006) On a computational approach for the approximate dynamics of averaged variables in nonlinear ODE systems: toward the derivation of constitutive laws of the rate type. J Mech Phys Solids 54:2183–2213

Anderson PW (1972) More is different. Sci Agric 177:393–396

Artstein Z (1999) Singularly perturbed ordinary differential equations with nonautonomous fast dynamics. J Dyn Differ Equ 11:297–318

Auger P, de la Parra RB, Poggiale JC, Sánchez E, Sanz L (2008) Aggregation methods in dynamical systems and applications in population and community dynamics. Phys Life Rev 5:79–105

Bancaud A, Wagner G, Conde e Silva N, Lavelle C, Wong H, Mozziconacci J, Barbi M, Sivolob A, Le Cam E, Mouawad L, Viovy JL, Victor JM, Prunell A (2007) Nucleosome chiral transition under positive torsional stress in single chromatin fibers. Mol Cell 27:135–147

Bécavin C, Barbi M, Victor JM, Lesne A (2010) Transcription within condensed chromatin: steric hindrance facilitates elongation. Biophys J 98:824–833
Ben Arous G, Owhadi H (2003) Multiscale homogenization with bounded ratios and anomalous slow diffusion. Comm Pure Appl Math 56:80–113
Ben Haïm E, Lesne A, Victor JM (2001) Chromatin: a tunable spring at work inside chromosomes. Phys Rev E 64:051921
Bogoliubov NN, Mitropolskii YA (1961) Asymptotic methods in the theory of nonlinear oscillations. Gordon and Breach, New York
Bystricky K, Heun P, Gehlen L, Langowski J, Gasser SM (2004) Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. Proc Natl Acad Sci USA 101:16495–16500
Castiglione P, Falcioni M, Lesne A, Vulpiani A (2008) Chaos and coarse-graining in statistical mechanics. Cambridge University Press, Cambridge
Deygout C, Lesne A, Campillo F, Rapaport A (2013) Homogenised model linking microscopic and macroscopic dynamics of a biofilm: application to growth in a plug flow reactor. Ecol Model 250:15–24
Dieckmann U, Law R (2000) Relaxation projections and the method of moments. In: Dieckmann U, Law R, Metz JAJ (eds) The geometry of ecological interactions: simplifying spatial complexity. Cambridge University Press, Cambridge, pp 412–455
E W, Engquist B (2003) The heterogeneous multiscale methods. Comm Math Sci 1:87–132
E W, Ren W, Vanden-Eijnden E (2009) A general strategy for designing seamless multiscale methods. J Comput Phys 228:5437–5453
El Hajji M, Rapaport A (2009) Practical coexistence of two species in the chemostat—a slow-fast characterization. Math Biosci 218:33–39
Ellis GFR (2005) Physics, complexity, and causality. Nat Biotechnol 435:743
Hänggi P, Talkner P, Borkovec M (1990) Reaction rate theory: fifty years after Kramers. Rev Mod Phys 62:251–342
Gardiner CW (1983) Handbook of stochastic methods. Springer, Berlin
Gaveau B, Lesne A, Schulman LS (1999) Spectral signatures of hierarchical relaxation. Phys Lett A 258:222–228
Givon D, Kupferman R, Stuart A (2004) Extracting macroscopic dynamics: model problems and algorithms. Nonlinearity 17:R55–R127
Hersen P, Andersen K, Elbelrhiti H, Andreotti B, Claudin P, Douady S (2004) Corridors of barchan dunes: stability and size selection. Phys Rev E 69:011304
Hornung U (1997) Homogenization and porous media. Springer, Berlin
Jenuwein T, Allis CD (2001) Translating the histone code. Sci Agric 293:1074–1080
Karsenti E, Nedelec F, Surrey T (2006) Modelling microtubule patterns. Nat Cell Biol 8:1204–1211
Kratky O, Porod G (1949) Röntgenuntersuchung gelöster Fadenmoleküle. Rec Trav Chim Pays-Bas 68:1106–1123
Laguës M, Lesne A (2011) Scale invariance. Springer, Berlin
Langevin P (1908) On the theory of Brownian motion. C.R. Acad Sci (Paris) 146:530–533. Reprinted in Am J Phys 65:1079–1081 (1997)
Lemarchand A, Lesne A, Mareschal M (1995) Langevin approach to a chemical wave-front: selection of the propagation velocity by internal noise. Phys Rev E 51:4457–4465
Lesne A (1998) Renormalization methods. Wiley, New-York
Lesne A (2006) Multiscale approaches. In: Françoise JP, Naber G, Tsun TS (eds) Encyclopedia of mathematical physics. Elsevier, Amsterdam, pp 465–482
Lesne A (2007) Discrete vs continuous controversy in physics. Math Struct Comput Sci 17:185–223
Lesne A, Bécavin C, Victor JM (2012) The condensed chromatin fiber: an allosteric chemo-mechanical machine for signal transduction and genome processing. Phys Biol 9:013001
Lesne A, Benecke A (2008a) Probability landscapes for integrative genomics. Theor Biol Med Model 5:9
Lesne A, Benecke A (2008b) Feature context-dependency and complexity reduction in probability landscapes for integrative genomics. Theor Biol Med Model 5:21
Lesne A, Victor JM (2006) Chromatin fiber functional organization: some plausible models. Eur Phys J E 19:279–290
Malo M, Cartier-Michaud A, Fabre-Guillevin E, Hutzler G, Delaplace F, Barlovatz-Meimon G, Lesne A (2010) When a collective outcome triggers a rare individual event: a mode of metastatic process in a cell population. Math Pop Stud 17:136–165

Malo M, Cartier-Michaud A, Charrière-Bertrand C, Gadea G, Anguille C, Supiramaniam A, Lesne A, Delaplace F, Hutzler G, Roux P, Lawrence DA, Barlovatz-Meimon G (2012) Matrix-bound PAI-1 supports membrane blebbing via RhoA-Rock1 signaling. PLoS One 7:e32204

Marr C, Geertz M, Hütt MT, Muskhelishvili G (2008) Two distinct logical types of network control in gene expression profiles. BMC Syst Biol 2:18

Mathias JD, Grediac M, Balandraud X (2006) On the bidirectional stress distribution in rectangular bonded composite patches. Int J Solids Struct 43:6921–6947

Mozziconacci J, Lavelle C, Barbi M, Lesne A, Victor JM (2006) A physical model for the condensation and decondensation of eukaryotic chromosomes. FEBS Lett 580:368–372

Mozziconacci J, Victor JM (2003) Nucleosome gaping supports a functional structure for the 30 nm chromatin fiber. J Struct Biol 143:72–76

Murray JD (2003) Mathematical biology (3rd edition). Springer, New York

Muskhelishvili G, Sobetzko P, Geertz M, Berger M (2010) General organisational principles of the transcriptional regulation system: a tree or a circle?. Mol BioSyst 6:662–676

Nayfeh AH (1973) Perturbation methods. Wiley, New York

Nicholson C (2001) Diffusion and related transport mechanisms in brain tissue. Rep Prog Phys 64:815–884

Polanyi M (1968) Life's irreducible structure. Sci Agric 160:1308–1312

Radulescu O, Gorban AN, Zinovyev A, Lilienbaum A (2008) Robust simplifications of multiscale biochemical networks. BMC Syst Biol 2:86

Sanders J, Verhulst F, Murdock J (2007) Averaging methods in nonlinear dynamical systems. Springer, New York

Schrödinger E (1944) What is life. Cambridge University Press, Cambridge

Simon HA (1962) The architecture of complexity. Proc Am Phil Soc 106:467–482

Sivolob A, Lavelle C, Prunell A (2003) Sequence-dependent nucleosome structural and dynamic polymorphism. Potential involvement of histone H2B N-terminal tail proximal domain. J Mol Biol 326: 49–63

Spector DL (2003) The dynamics of chromosome; organization and gene regulation. Annu Rev Biochem 72:573–608

Torquato S (2002) Random heterogeneous materials: microstructure and macroscopic properties. Springer, Berlin

Travers A, Muskhelishvili G (2005) DNA supercoiling—a global transcriptional regulator for enterobacterial growth. Nat Rev Microbiol 3:157–169

Van Kampen N (1981) Stochastic processes in physics and chemistry. North-Holland, Amsterdam

Werner BT (1999) Complexity in natural landform patterns. Sci Agric 284:102–104

Wong H, Victor JM, Mozziconacci J (2007) An all-atom model of the chromatin fiber containing linker histones reveals a versatile structure tuned by nucleosomal repeat length. PLoS One 2:e877

Woodcock CL, Grigoryev SA, Horowitz RA, Whitaker N (1993) A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. Proc Natl Acad Sci USA 90:9021–9025

Zlatanova J, Bishop TC, Victor JM, Jackson V, van Holde K (2009) The nucleosome family: dynamic and growing. Struct Bond 17:160–171

Interface Focus

*Interface Focus* (2012) **2**, 55–64
doi:10.1098/rsfs.2011.0067
*Published online* 9 November 2011

Review

# A theory of biological relativity: no privileged level of causation

**Denis Noble***

*Department of Physiology, Anatomy and Genetics, University of Oxford, Parks Road, Oxford OX1 3PT, UK*

Must higher level biological processes always be derivable from lower level data and mechanisms, as assumed by the idea that an organism is completely defined by its genome? Or are higher level properties necessarily also causes of lower level behaviour, involving actions and interactions both ways? This article uses modelling of the heart, and its experimental basis, to show that downward causation is necessary and that this form of causation can be represented as the influences of initial and boundary conditions on the solutions of the differential equations used to represent the lower level processes. These insights are then generalized. *A priori*, there is no privileged level of causation. The relations between this form of 'biological relativity' and forms of relativity in physics are discussed. Biological relativity can be seen as an extension of the relativity principle by avoiding the assumption that there is a privileged scale at which biological functions are determined.

**Keywords: downward causation; biological relativity; cardiac cell model; scale relativity**

## 1. INTRODUCTION

Have we reached the limits of applicability of the relativity principle? And could it have relevance to biology?

By 'relativity principle' in this context, I mean distancing ourselves in our theories from specific absolute standpoints for which there can be no *a priori* justification. From Copernicus and Galileo through to Poincaré and Einstein, the reach of this general principle of relativity has been progressively extended by removing various absolute standpoints in turn. People realized that those standpoints represent privileging certain measurements as absolute, for which there is and could be no basis. First, we removed the idea of privileged location (so the Earth is not the centre of the Universe), then that of absolute velocity (since only relative velocities can be observed), then that of acceleration (an accelerating body experiences a force indistinguishable from that of gravity, leading to the idea of curved space–time). Could biology be the next domain for application of the relativity principle? This article will propose that there is, *a priori*, no privileged level of causality in biological systems. I will present evidence, experimental and theoretical, for the existence of downward causation from larger to smaller scales by showing how mathematical modelling has enabled us to visualize exactly how multi-level 'both-way' causation occurs. I will discuss the consequences for attempts to understand organisms as multi-scale systems. Finally, I will assess where some of the extensions of the relativity principle now stand in relation to these goals.

*denis.noble@dpag.ox.ac.uk

One contribution of 15 to a Theme Issue 'Top-down causation'.

## 2. THE HIERARCHY OF LEVELS: 'UP' AND 'DOWN' ARE METAPHORS

In biological science, we are used to thinking in terms of a hierarchy of levels, with genes occupying the lowest level and the organism as a whole occupying the highest level of an individual. Protein and metabolic networks, intracellular organelles, cells, tissues, organs and systems are all represented as occupying various intermediate levels. The reductionist causal chain is then represented by upward-pointing arrows (figure 1). In this figure, I have also represented the causation between genes and proteins with a different kind of arrow (dotted) from the rest of the upward causation since it involves a step that is usually described in terms of coding, in which particular triplets of nucleic acids code for specified amino acids so that a complete protein has a complete DNA template (or, more correctly, a complete mRNA template that may be formed from various DNA exons). The standard story is that genes code for proteins, which then go on to form the networks. Coding of this kind does not occur in any of the other parts of the causal chain, although signalling mechanisms at these levels could also be described in terms of coding (a signal can always be described as using a code in this general sense).

The concepts of level, and of 'up' and 'down', 'higher' and 'lower', however, are all metaphors. There

*Received* 1 July 2011
*Accepted* 17 October 2011

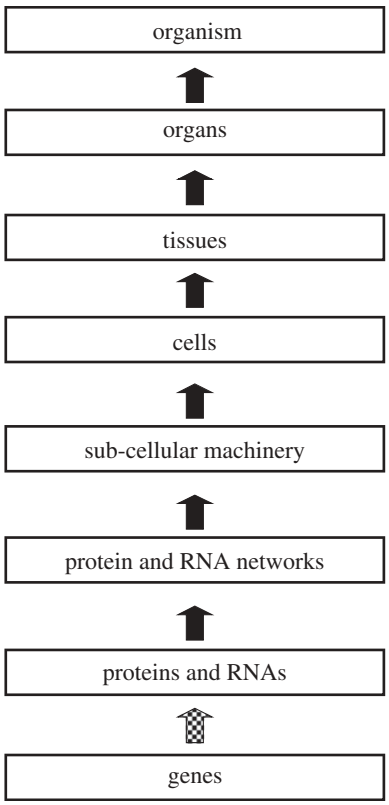


Figure 1. Upward causation: the reductionist causal chain in biology. This is a gross simplification, of course. No one today seriously believes that this diagram represents all causation in biology. Reductive biological discourse, however, privileges this form of causation and regards it as the most important. In particular, the nature and the direction of the lowest arrow (dotted) are fixed and represent the impact of the central dogma of molecular biology. Adapted from Noble [1, fig. 1].

is no literal sense in which genes lie 'below' cells, for example. Genes are all over the body, so also are cells, and the organism itself, well, that is very much everywhere. This is why I prefer 'scale' to 'level'. The real reason for putting genes, as DNA sequences, at the bottom of the hierarchy is that they exist at the smallest (i.e. molecular) scale in biological systems. The formation of networks, cells, tissues and organs can be seen as the creation of processes at larger and larger scales.

Does the metaphorical nature of the way we represent upward and downward causation matter? The bias introduced by the metaphor is that there is a strong tendency to represent the lower levels as somehow more concrete. Many areas of science have proceeded by unravelling the small elements underlying the larger ones. But notice the bias already creeping in through the word 'underlying' in the sentence I have just written. We do not use the word 'overlying' with anything like the same causal force. That bias is reinforced by the undeniable fact that, in biology, many of the great advances have been made by inventing more and more powerful microscopical and other techniques that allow us to visualize and measure ever smaller components. I was a graduate student when the first electron microscopes were introduced and I recall the excitement over the ability to visualize individual molecules of, for example, the contractile proteins in muscle cells. This enabled the contractile protein machinery to be understood: and so the sliding filament model of muscle contraction was born [2,3]. Taking a system apart to reveal its bits and then working out how the bits work together to form the machinery is a standard paradigm in science.

That paradigm has been remarkably successful. Breaking the human organism down into 25 000 or so genes and 100 000 or so proteins must be one of the greatest intellectual endeavours of the twentieth century, with completion of the first draft sequencing of the entire human genome occurring appropriately at the turn of the millennium [4,5].

As a scientific approach, therefore, the reductionist agenda has been impressively productive. The question remains though. If 'up' and 'down' are metaphorical, how can causation in one direction be privileged over that in the reverse direction? Are molecular events somehow causally more important than events that occur at the scales of cells, organs or systems? And are there causally efficacious processes that can only be characterized at higher scales?

## 3. THE CENTRAL DOGMA OF MOLECULAR BIOLOGY: WHAT DOES IT SHOW?

It is hard to think of an *a priori* reason why one level in a biological system should be privileged over other levels when it comes to causation. That would run counter to the relativity principle. Moreover, I will outline later in this article how mathematical modelling has enabled us to visualize exactly how multi-level 'both-way' causation occurs. If the reductionist view is to be justified, therefore, it must be done *a posteriori*: we need empirical evidence that information that could be regarded as 'controlling' or 'causing' the system only passes in one direction, i.e. upwards. In biology, we do not have to look very far for that empirical evidence. The central dogma of molecular biology [6,7] is precisely that. Or is it?

Let us pass over the strange fact that it was called a 'dogma', first by Crick and then by very many who followed him. Nothing in science should be a dogma of course. Everything is open to question and to testing by the twin criteria of logic (for mathematical ideas) and experimental findings (for theories with empirical consequences). So, let us look more closely at what is involved. The essence of the central dogma is that 'coding' between genes and proteins is one-way. I prefer the word 'template' to 'coding' since 'coding' already implies a program. Another way to express the central point of this article is to say that the concept of a genetic program is part of the problem [1]. I will briefly explain why.

The sequences of DNA triplets form templates for the production of different amino acid sequences in proteins. Amino acid sequences do not form templates for the production of DNA sequences. That, in essence, is what was shown. The template works in only one direction, which makes the gene appear primary. So what does the genome cause? The coding sequences form a list of proteins and RNAs that might be made in

a given organism. These parts of the genome form a database of templates. To be sure, as a database, the genome is also extensively formatted, with many regulatory elements, operons, embedded within it. These regulatory elements enable groups of genes to be coordinated [8] in their expression levels. And we now know that the non-coding parts of the genome also play important regulatory functions. But the genome is not a fixed program in the sense in which such a computer program was defined when Jacob and Monod introduced their idea of 'le programme génétique' [9–11]. It is rather a 'read–write' memory that can be organized in response to cellular and environmental signals [12]. Which proteins and RNAs are made when and where is not fully specified. This is why it is possible for the 200 or so different cell types in an organism such as the human to make those cell types using exactly the same genome. A heart cell is made using precisely the same genome in its nucleus as a bone cell, a liver cell, pancreatic cell, etc. Impressive regulatory circuits have been constructed by those who favour a genetic program view of development [13,14], but these are not independent of the 'programming' that the cells, tissues and organs themselves use to epigenetically control the genome and the patterns of gene expression appropriate to each cell and tissue type in multi-cellular organisms. As I will show later, the circuits for major biological functions necessarily include non-genome elements.

That fact already tells us that the genome alone is far from sufficient. It was Barbara McClintock, who received the Nobel Prize for her work on jumping genes, who first described the genome as 'an organ of the cell' [15]. And so it is. DNA sequences do absolutely nothing until they are triggered to do so by a variety of transcription factors, which turn genes on and off by binding to their regulatory sites, and various other forms of epigenetic control, including methylation of certain cytosines and interactions with the tails of the histones that form the protein backbone of the chromosomes. All of these, and the cellular, tissue and organ processes that determine when they are produced and used, 'control' the genome. For further detail on this issue, the reader is referred to Shapiro's article on re-assessing the central dogma [16] and to his book *Evolution: the view from the 21st century* [12]. A good example in practice is the way in which neuroscientists are investigating what they call electro-transcription coupling [17], a clear example of downward causation since it involves the transmission of information from the neural synapses to the nuclear DNA.

To think that the genome completely determines the organism is almost as absurd as thinking that the pipes in a large cathedral organ determine what the organist plays. Of course, it was the composer who did that in writing the score, and the organist himself who interprets it. The pipes are his passive instruments until he brings them to life in a pattern that he imposes on them, just as multi-cellular organisms use the same genome to generate all the 200 or so different types of cell in their bodies by activating different expression patterns. This metaphor has its limitations. There is no 'organist'. The 'music of life' plays itself [1], rather as some musical ensembles perform without a conductor. And, of course, the 'organ' varies between individuals in a species. But it is quite a good metaphor. The pipes of an organ are also 'formatted' to enable subsets to be activated together by the various stops, manuals and couplers. Like the regulatory parts of the genome, these parts of the organ make it easier to control, but both, genome and organ, still do nothing without being activated. The patterns of activation are just as much part of the 'program' as the genome itself [18].

So, even at the very lowest level of the reductionist causal chain, we discover a conceptual error. The protein-coding sequences are templates. They determine which set of proteins the organism has to play with, just as a child knows which pieces of Lego or Meccano she has available for construction. Those parts of the genome are best regarded as a database. Even when we add in the regulatory and non-coding regions, there is no program in the genome in the sense that the sequences could be parsed in the way in which we would analyse a computer program to work out what it is specifying. The reason is that crucial parts of the program are missing. To illustrate this, I will use the example of cardiac rhythm to show that the non-genomic parts are essential.

## 4. INSIGHTS FROM EXPERIMENTAL AND MODELLING WORK ON HEART CELLS

Over many years, my research has involved experimental and computational work on heart cells. I was the first to analyse the potassium ion channels in heart muscle [19,20] and to construct a computer model based on the experimental findings [21,22]. Since that time, a whole field of heart modelling has developed [23,24].

How do we construct such models? The trail was blazed by Hodgkin & Huxley [25] in their Nobel prize-winning work on the nerve impulse. The ion channel proteins that sit across the cell membrane control its electrical potential by determining the quantity of charge that flows across the cell membrane to make the cell potential become negative or positive. The gating of these channels is itself in turn controlled by the cell potential. This is a multi-level loop. The potential is a cell-level parameter; the ion channel openings and closings are protein-level parameters. The loop, originally called the Hodgkin cycle, is absolutely essential to the rhythm of the heart. Breaking the feedback (downward causation) between the cell potential and the gating of the ion channels and cellular rhythm are abolished. A simple experiment on one of the cardiac cell models will demonstrate this computationally.

In figure 2 [26], a model of the sinus node (the pacemaker region of the heart) was run for 1300 ms, during which time six oscillations were generated. These correspond to six heartbeats at a frequency similar to that of the heart of a rabbit, the species on which the experimental data were obtained to construct the model. During each beat, all the currents flowing through the protein channels also oscillate in a specific sequence. To simplify the diagram, only three of those protein channels are represented here. At 1300 ms, an experiment was

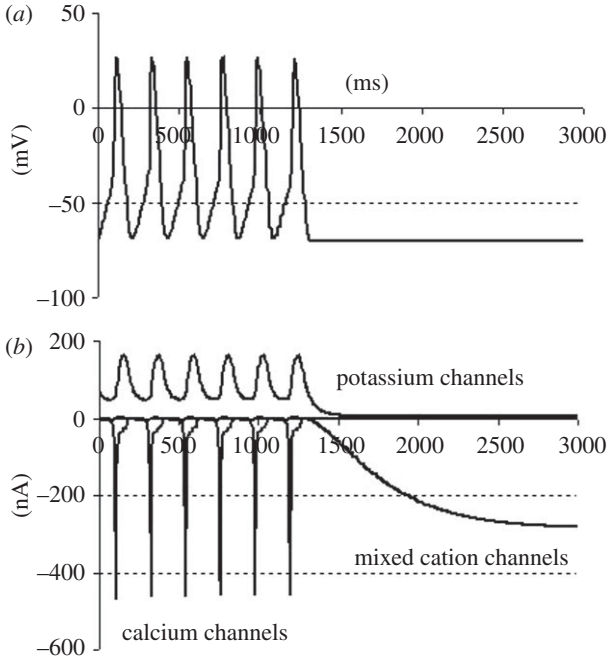


Figure 2. Computer model of pacemaker rhythm in the heart [27]. For the first six beats, the model is allowed to run normally and generates rhythm closely similar to a real cell. Then the feedback from cell voltage (*a*) to protein channels ((*b*) currents in nanoamps) is interrupted by keeping the voltage constant (voltage clamp). All the protein channel oscillations then cease. They slowly change to steady constant values. Without the downward causation from the cell potential, there is no rhythm. Adapted from Noble [1, fig. 3].

performed on the model. The 'downward causation' between the global cell property, the membrane potential and the voltage-dependent gating of the ion channels was interrupted. If there were a sub-cellular 'program' forcing the proteins to oscillate, the oscillations would continue. In fact, however, all oscillations cease and the activity of each protein relaxes to a steady value, as also happens experimentally. In this case, therefore, the 'program' includes the cell itself and its membrane system. In fact, we do not need the concept of a separate program here. The sequence of events, including the feedback between the cell potential and the activity of the proteins, simply *is* cardiac rhythm. It is a property of the interactions between all the components of the system. It does not even make sense to talk of cardiac rhythm at the level of proteins and DNA, and it does not make sense to suppose that there is a *separate* program that 'runs' the rhythm.

Of course, all the proteins involved in cardiac rhythm are encoded by the genome, but these alone would not generate rhythm. This is the sense (see above) in which I maintain that there is not a program for cardiac rhythm in the genome. The non-genomic structural elements are also essential. Similar arguments apply, for example, to circadian rhythm [1,28] and, indeed, to all functions that require cellular structural inheritance as well as genome inheritance. Indeed, I find it hard to identify functions that do not involve what Cavalier-Smith [29,30] has characterized as the membranome. Much of the logic of life lies in its delicate oily membranes.

## 5. GENERALIZATION OF THE ARGUMENT IN MATHEMATICAL TERMS

We can generalize what is happening here in mathematical terms. The activity of the ion channels is represented by differential equations describing the speed and the direction of the gating processes on each protein. The coefficients in those differential equations are based on experimental data. One might think that, provided all the relevant protein mechanisms have been included in the model and if the experimental data are reliable, and the equations fit the data well, cardiac rhythm would automatically 'emerge' from those characteristics. It does not. The reason is very simple and fundamental to any differential equation model. In addition to the differential equations you need the initial and boundary conditions. Those values are just as much a 'cause' of the solution (cardiac rhythm) as are the differential equations. In this case, the boundary conditions include the cell structure, particularly those of its membranes and compartments. Without the constraints imposed by the higher level structures, and by other processes that maintain ionic concentrations, the rhythm would not occur. If we were to put all the components in a Petri dish mixed up in a nutrient solution, the interactions essential to the function would not exist. They would lack the spatial organization necessary to do so.

This fact tells us therefore how higher levels in biological systems exert their influence over the lower levels. Each level provides the boundary conditions under which the processes at lower levels operate. Without boundary conditions, biological functions would not exist.

The relationships in such models are illustrated in figure 3. The core of the model is the set of differential equations describing the kinetics of the components of the system (e.g. the channel proteins in figure 2). The initial conditions are represented as being on the same level since they are the state of the system at the time at which the simulation begins. The boundary conditions are represented as being at a higher level since they represent the influence of their environment on the components of the system. So far as the proteins are concerned, the rest of the cell is part of their environment.

The diagram of figure 1 therefore should look more like figure 4. There are multiple feedbacks from higher levels to lower levels in addition to those from lower to higher levels. In any model of lower level systems, these form the constraints that would need to be incorporated into the boundary and initial conditions. As figure 4 indicates, these include triggers of cell signalling (via hormones and transmitters), control of gene expression (via transcription factors), epigenetic control (via methylation and histone marking), and note also that it is the protein machinery that reads genes—and continually repairs copying errors and so makes the genome reliable. To reverse a popular metaphor, that of the selfish gene [31], it is the 'lumbering robot' that is responsible for any 'immortality' genes may possess!

## 6. DIFFERENTIAL AND INTEGRAL VIEWS OF THE RELATIONS BETWEEN GENOTYPES AND PHENOTYPES

All of this is fundamental and, even, fairly obvious to integrative physiologists. Physiologists have been

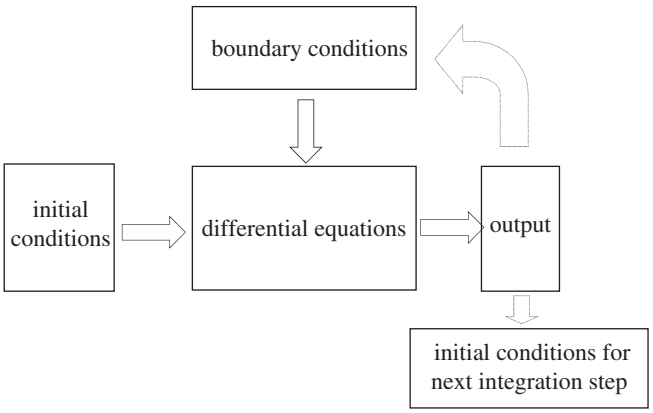


Figure 3. Many models of biological systems consist of differential equations for the kinetics of each component. These equations cannot give a solution (the output) without setting the initial conditions (the state of the components at the time at which the simulation begins) and the boundary conditions. The boundary conditions define what constraints are imposed on the system by its environment and can therefore be considered as a form of downward causation. This diagram is highly simplified to represent what we actually solve mathematically. In reality, boundary conditions are also involved in determining initial conditions and the output parameters can also influence the boundary conditions, while they in turn are also the initial conditions for a further period of integration of the equations. As with the diagrams (see §§2 and 5) of levels in biological systems, the arrows are not really unidirectional. The dotted arrows complete the diagram to show that the output contributes to the boundary conditions (although not uniquely), and determines the initial conditions for the next integration step.

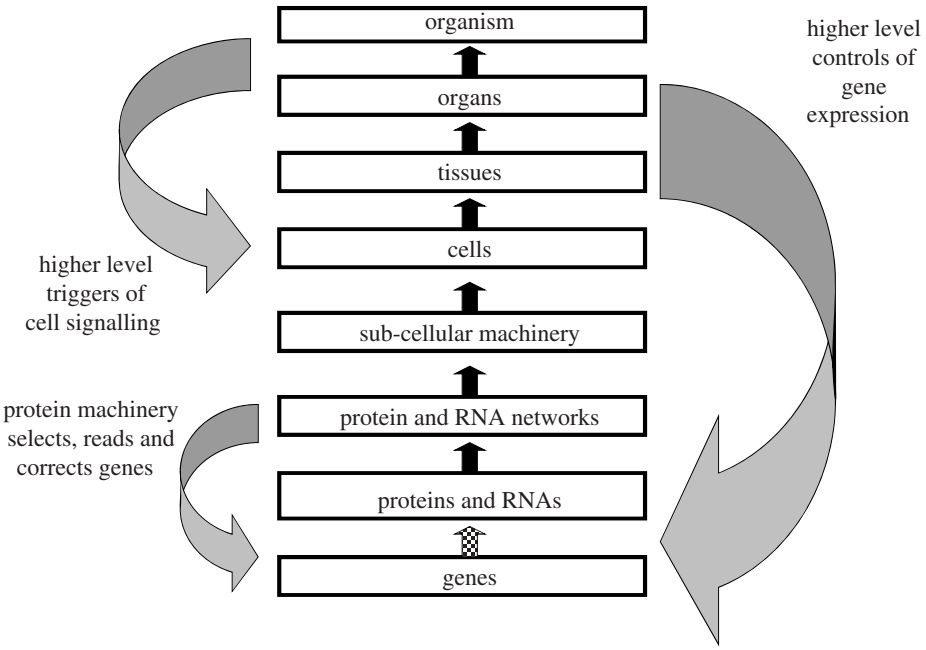


Figure 4. The completion of figure 1 with various forms of downward causation that regulates lower level components in biological systems. In addition to the controls internal to the organism, we also have to take account of the influence of the environment on all the levels (not shown in this diagram). Adapted from Noble [1, fig. 2]. Causation is, therefore, two-way, although this is not best represented by making each arrow two-way. A downward form of causation is not a simple reverse form of upward causation. It is better seen as completing a feedback circuit, as the examples discussed in the text show.

familiar with the basic ideas on multi-level control ever since Claude Bernard formulated the concept of control of the internal environment in his book *Introduction à l'étude de la médecine expérimentale* in 1865 [32] and Walter B. Cannon developed the idea of homeostasis in *The wisdom of the Body* in 1932 [33]. So, how has mainstream biology tended to ignore it, as has physiology also with some exceptions, for example Guyton's modelling of the circulation [34]? I think the main culprit here has been neo-Darwinism and particularly the popularizations of this theory as a purely gene-centric view [31]. The essential idea of gene-centric theories is what I have called the differential view of the relationships between genes and phenotypes [35–38]. The idea is essential in the sense that it excludes alternative theories by arguing that what matters in evolutionary terms are *changes* in the genotype that are reflected in *changes* in phenotype. Selection of the phenotype is therefore, according to this logic, fundamentally equivalent to selection of particular genes (or, more strictly, gene alleles). This view might have been appropriate for a time when genes were regarded as hypothetical entities defined as

the cause of each phenotype. It is not appropriate for the current molecular and systems biology-inspired definition of a gene as a particular DNA sequence, replicating and being expressed within cellular and multi-cellular systems. In principle, we can now investigate all the functions that DNA sequence is involved in, though that goal still remains very ambitious in practice. We do not have to be restricted to investigating differences. Anyway, that would be to focus on the tip of the iceberg. Considering just differences at the genetic level is as limiting as it would be for mathematics to limit itself to differential equations without integrating them, as though the integral sign and what it stands for had never been invented [37].

The analogy with the mathematics of differential calculus is strongly revealing. Integration requires knowledge of the initial and boundary conditions in addition to the differential equations themselves (figure 3). One can only ignore those by restricting oneself to the differential equation 'level'. In a similar way, the neo-Darwinist synthesis tends to ignore downward causation precisely because such causation requires an integral rather than a differential view of genetics for its analysis.

Specifically, when neo-Darwinists refer to the 'genes' for any particular phenotype on which selection may act, they are not referring to complete protein-coding sequences of DNA, they are really referring to *differences* between alleles. The 'gene' is, therefore, defined as this inheritable difference in phenotype. It would not even matter whether this difference is a difference in DNA or in some other inheritable factor, such as inherited cytoplasmic changes in *Paramecium* [39], or the cytoplasmic influences on development observed in cross-species cloning of fish [40].

By contrast, the integral view for which I am arguing does not focus on differences. Instead it asks: what are all the functions to which the particular DNA sequence contributes? Indeed, it would not matter whether those functions are ones that result in a different phenotype. Through the existence of multiple back-up mechanisms, many DNA changes, such as knockouts, do not have a phenotypic effect on their own. As many as 80 per cent of the knockouts in yeast are normally 'silent' in this way [41]. Their functionality can be revealed only when the boundary conditions, such as the nutrient environment, are changed. The analogy that I am drawing with differential and integral calculus draws its strength precisely through this dependence on the boundary conditions. A differential equation, on its own, has an infinite set of solutions until those are narrowed down by the boundary conditions. Similarly, a difference in DNA sequence may have a wide variety of possible phenotypic effects, including no effect at all, until the boundary conditions are set, including the actions of many other genes, the metabolic and other states of the cell or organism, and the environment in which the organism exists.

## 7. A (BIOLOGICAL) THEORY OF RELATIVITY

I and my colleagues have expressed many of the ideas briefly outlined here in the form of some principles of systems biology [1,42–44]. One of those principles is that, *a priori*, there is no privileged level of causation in biological systems. Determining the level at which a function is integrated is an empirical question. Cardiac rhythm is clearly integrated at the level of the pacemaker sinus node cell, and does not even exist below that level. The principle can be restated in a more precise way by saying that the level at which each function is integrated is at least partly a matter of experimental discovery. There should be no dogmas when it comes to causation in biological systems.

## 8. CONNECTING LEVELS

One way to connect levels in biological simulation can be derived immediately from figure 3. Since the boundary conditions for integration are set by the higher level, determining those conditions at that level either by measurement or by computation can enable them to be inserted into the equations at the lower level. This is the way, for example, in which the structural organization of the whole heart is used to constrain the ordinary and partial differential equations describing the protein channels and the flow of ionic current through the structure—conduction is faster along a fibre axis, for example, than across and between fibres. These kinds of constraints turn out to be very important in studying cardiac arrhythmias, where the sequence of events from ordered rhythm to tachycardia and then to fibrillation is dependent on the high-level structure [45–52].

A similar approach could be used to simulate other biological processes such as development. If we had a sufficiently detailed knowledge of the fertilized egg cell structure and networks, including particularly the concentrations and locations of transcription factors and the relevant epigenetic influences, we could imagine solving equations for development involving gene expression patterns determined by both the genome and its non-DNA regulators. In this case, the various levels 'above' the cell (better viewed as 'around' the cell) would actually develop with the process itself, as it moves through the various stages, so creating the more global constraints in interaction with the environment of the organism. We cannot do that kind of ambitious computation at the present time, and the reason is not that we do not know the genome that has been sequenced. The problem lies at a higher level. We cannot yet characterize all the relevant concentrations of transcription factors and epigenetic influences. It is ignorance of all those forms of downward causation that is impeding progress. Even defining which parts of the DNA sequence are transcribed (and so to identify 'genes' at the DNA level—and here I would include sequences that form templates for RNAs as 'genes') requires higher level knowledge. This approach would naturally take into account the role of cell and tissue signalling in the generation of organizing principles involved in embryonic induction, originally identified in the pioneering work of Spemann & Mangold [53–55]. The existence of such induction is itself an example of dependence on boundary conditions. The induction mechanisms emerge as the embryo interacts with its

environment. Morphogenesis is not entirely hard-wired into the genome.

## 9. EMERGENCE AND BOUNDARY CONDITIONS

Reference to emergence leads me to a fundamental point about the limits of reductionism. An important motivation towards reductionism is that of reducing complexity. The idea is that if a phenomenon is too complex to understand at level X then go down to level Y and see, first, whether the interactions at level Y are easier to understand and theorize about, then, second, see whether from that understanding one can automatically understand level X. If indeed all that is important at level X were to be entirely derivable from a theory at level Y, then we would have a case of what I would call 'weak emergence', meaning that descriptions at level X can then be seen to be a kind of shorthand for a more detailed explanatory analysis at level Y. 'Strong emergence' could then be defined as cases where this does not work, as we found with the heart rhythm model described above. They would be precisely those cases where what would be merely contingent at level Y is systematic at level X. I am arguing that, if level Y is the genome, then we already know that 'weak emergence' does not work. There is 'strong emergence' because contingency beyond what is in the genome, i.e. in its environment, also determines what happens.

This kind of limit to reductionism is not restricted to biology. Spontaneous symmetry breaking in particle physics is a comparable case. An infinitesimal change can determine which way symmetry is broken [56]. How that happens in particular cases is not derivable from particle theory itself. Biological reductionists whose motivation is that of reducing biology to physics need to be aware that physics itself also displays the kind of limits I am describing here. Nor are these limits restricted to particle theory.

Connecting levels through setting initial and boundary conditions derived from multi-level work has served biological computation very well so far. The successes of the Physiome Project attest the same [23,57]. But there are two reasons why I think it may not be enough.

## 10. COMPUTABILITY

The first is the problem of computability.

Consider the heart again. Since the very first supercomputer simulations [58,59] in which cell models were incorporated into anatomical structures representing heart tissue and the whole organ [23,60,61], we have continually pushed up against the limits of computer speed and memory. Even today, we are only beginning to be within reach of whole organ simulations of electrical activity running in real time, i.e. that it should take only 1 s of computer time to calculate a second of heart time. Yet, such models represent only a few per cent of the total number of proteins involved in cardiac function, although, of course, we hope we have included the most important ones for the functions we are representing. And the equations for each component are the simplest that can capture the relevant kinetics of ion channel function. Expanding the models to include most, rather than a very few, gene products, extending the modelling of each protein to greater detail, and extending the time scale beyond a few heartbeats would require orders of magnitude increases in computing power.

In fact, it is relatively easy to show that complete bottom-up reconstructions from the level of molecules to the level of whole organs would require much more computing power than we are ever likely to have available, as I have argued in a previous article [37]. In that article, I began by asking two questions. First, 'are organisms encoded as molecular descriptions in their genes?' And, second, 'by analysing the genome, could we solve the forward problem of computing the behaviour of the system from this information, as was implied by the original idea of the "genetic program" and the more modern representation of the genome as the "book of life"?' (for a recent statement of these ideas see [62]). The answer to both questions was 'no'. The first would have required that the central dogma of molecular biology should be correct in excluding control of the genome by its environment, while the second runs into the problem of combinatorial explosion. The number of possible interactions between 25 000 genes exceeds the total number of elementary particles in the whole-known Universe [63], even when we severely restrict the numbers of gene products that can interact with each other (see also [64]). Conceivably, we might gain some speed-up from incorporating analogue computation to go beyond the Turing limits [65], but it is still implausible to expect that increased computer power will provide all we need or that it is the best way forward [66].

## 11. SCALE RELATIVITY

The second reason why connecting levels via boundary conditions may not be enough is that it assumes that the differential equations themselves remain unchanged when they form part of a hierarchy of levels. This is what we would expect in a classical analysis. But is this necessarily correct?

One of the reasons I introduced this article with some remarks on the general principle of relativity and its history of distancing us from unwarranted assumptions concerning privileged standpoints is that we can ask the same question about levels and scales. If there is no privileged level of causation, then why should there be a privileged scale? This is the question raised by Laurent Nottale's theory of scale relativity [67,68]. As Nottale *et al.* [69] shows in his recent book, the consequences of applying the relativity principle to scales are widespread and profound, ranging from understanding the quantum–classical transition in physics to potential applications in systems biology [70,71].

I will conclude this article, therefore, by describing what that theory entails, how it relates to the general theory of biological relativity I have outlined here and what is the status of such theories now?

The central feature from the viewpoint of biological modelling can be appreciated by noting that the equations for structure and for the way in which elements move and interact in that structure in biology

necessarily depend on the resolution at which it is represented. Unless we represent everything at the molecular level which, as argued above, is impossible (and fortunately unnecessary as well), the differential equations should be scale-dependent. As an example, at the level of cells, the equations may represent detailed compartmentalization and non-uniformity of concentrations, and hence include intracellular diffusion equations, or other ways of representing non-uniformity [72–74]. At the level of tissues and organs, we often assume complete mixing (i.e. uniformity) of cellular concentrations. At that level, we also usually lump whole groups of cells into grid points where the equations represent the lumped behaviour at that point.

These are *practical* reasons why the equations we use are scale-dependent. The formal theory of scale relativity goes much further since it proposes that it is theoretically *necessary* that the differential equations should be scale-dependent. It does this by assuming that space–time itself is continuous but generally non-differentiable, therefore fractal, not uniform. The distance between two points, therefore, depends on the scale at which one is operating and that, in the limit, as $dx$ or $dt$ tend to zero, the differential is most often not defined. This does not mean that differential equations cannot be used, simply that terms corresponding to scale should be included as an extension of the usual differential equations as explicit influences of scale on the system. The derivation of these extension terms can be found in Auffray & Nottale [70, pp. 93–97] and in Nottale [69, pp. 73–141].

The idea of fractal space–time may seem strange. I see it as an extension of the general relativity principle that space–time is not independent of the objects themselves found within it, i.e. space–time is not uniform. We are now used to this idea in relation to the structure of the Universe and the way in which, according to Einstein's general relativity, space–time is distorted by mass and energy to create phenomena such as gravitational lensing [75,76]. But, it is usually assumed that, on smaller scales, the classical representations of space–time are sufficient. It is an open question whether that is so and whether scale should be incorporated in explicit terms in the equations we use in multi-scale models. Remember also that the utility of a mathematical concept does not depend on how easily we can visualize the entities involved. We find it difficult to imagine a number like $\sqrt{-1}$, but it has great utility in mathematical analysis of the real world. We may need to think the unimaginable in order fully to understand the multi-scale nature of biology. The concept of scale is, after all, deeply connected to our conception of space–time.

## 12. CONCLUSIONS

While I think we can be certain that multi-level causation with feedbacks between all the levels is an important feature of biological organisms, the tools we have to deal with such causation need further development. The question is not whether downward causation of the kind discussed in this article exists, it is rather how best to incorporate it into biological theory and experimentation, and what kind of mathematics needs to be developed for this work.

This article is based on a presentation of a meeting on Downward Causation held at the Royal Society in September 2010. I should like to acknowledge valuable discussion with many of the participants of that meeting. I also thank Charles Auffray, Jonathan Bard, Peter Kohl and Laurent Nottale for suggesting improvements to the manuscript, and the journal referees for valuable criticism. I acknowledge support from an EU FP7 grant for the VPH-PreDiCT project. Following acceptance of this article, my attention was drawn to the article on downward causation by Michel Bitbol [77]. He approaches the issue of downward causation from Kantian and quantum mechanical viewpoints, but I would like to acknowledge that many of his insights are similar to and compatible with the views expressed here, particularly on the role of boundary conditions and the relativistic stance.

## REFERENCES

1 Noble, D. 2006 *The music of life.* Oxford, UK: Oxford University Press.
2 Huxley, A. F. 1957 Muscle structure and theories of contraction. *Prog. Biophys. Mol. Biol.* **7**, 255–318.
3 Huxley, H. 2004 Fifty years of muscle and the sliding filament hypothesis. *Eur. J. Biochem.* **271**, 1403–1415. (doi:10.1111/j.1432-1033.2004.04044.x)
4 International Human Genome Mapping Consortium. 2001 A physical map of the human genome. *Nature* **409**, 934–941. (doi:10.1038/35057157)
5 Venter, C. *et al.* 2001 The sequence of the human genome. *Science* **291**, 1304–1351. (doi:10.1126/science.1058040)
6 Crick, F. H. C. 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163.
7 Crick, F. H. C. 1970 Central dogma of molecular biology. *Nature* **227**, 561–563. (doi:10.1038/227561a0)
8 Jacob, F., Perrin, D., Sanchez, C., Monod, J. & Edelstein, S. 1960 The operon: a group of genes with expression coordinated by an operator. *C. R. Acad. Sci. Paris* **250**, 1727–1729.
9 Jacob, F. 1970 *La Logique du vivant, une histoire de l'hérédité.* Paris, France: Gallimard.
10 Jacob, F. 1982 *The possible and the actual.* New York, NY: Pantheon Books.
11 Monod, J. & Jacob, F. 1961 Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harbor Symp. Quant. Biol.* **26**, 389–401.
12 Shapiro, J. A. 2011 *Evolution: a view from the 21st century.* Upper Saddle River, NJ: Pearson Education Inc.
13 Davidson, E. H. 2006 *The regulatory genome: gene regulatory networks in development and evolution.* New York, NY: Academic Press.
14 Davidson, E. H. *et al.* 2002 A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Dev. Biol.* **246**, 2–13. (doi:10.1006/dbio.2002.0635)
15 McClintock, B. 1984 The significance of responses of the genome to challenge. *Science* **226**, 792–801. (doi:10.1126/science.15739260)
16 Shapiro, J. A. 2009 Revisiting the central dogma in the 21st century. *Ann. N. Y. Acad. Sci.* **1178**, 6–28. (doi:10.1111/j.1749-6632.2009.04990.x)
17 Deisseroth, K., Mermelstein, P. G., Xia, H. & Tsien, R. W. 2003 Signaling from synapse to nucleus: the logic behind the mechanisms. *Curr. Opin. Neurobiol.* **13**, 354–365. (doi:10.1016/S0959-4388(03)00076-X)

18 Coen, E. 1999 *The art of genes.* Oxford, UK: Oxford University Press.
19 Hutter, O. F. & Noble, D. 1960 Rectifying properties of heart muscle. *Nature* **188**, 495. (doi:10.1038/188495a0)
20 Noble, D. 1965 Electrical properties of cardiac muscle attributable to inward-going (anomalous) rectification. *J. Cell. Comp. Physiol.* **66**(Suppl. 2), 127–136. (doi:10.1002/jcp.1030660520)
21 Noble, D. 1960 Cardiac action and pacemaker potentials based on the Hodgkin–Huxley equations. *Nature* **188**, 495–497. (doi:10.1038/188495b0)
22 Noble, D. 1962 A modification of the Hodgkin–Huxley equations applicable to Purkinje fibre action and pacemaker potentials. *J. Physiol.* **160**, 317–352.
23 Bassingthwaighte, J. B., Hunter, P. J. & Noble, D. 2009 The cardiac physiome: perspectives for the future. *Exp. Physiol.* **94**, 597–605. (doi:10.1113/expphysiol.2008.044099)
24 Noble, D. 2007 From the Hodgkin–Huxley axon to the virtual heart. *J. Physiol.* **580**, 15–22. (doi:10.1113/jphysiol.2006.119370)
25 Hodgkin, A. L. & Huxley, A. F. 1952 A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544.
26 Noble, D., Denyer, J. C., Brown, H. F. & DiFrancesco, D. 1992 Reciprocal role of the inward currents $i_{b,Na}$ and $i_f$ in controlling and stabilizing pacemaker frequency of rabbit sino-atrial node cells. *Proc. R. Soc. Lond. B* **250**, 199–207. (doi:10.1098/rspb.1992.0150)
27 Noble, D. & Noble, S. J. 1984 A model of sino-atrial node electrical activity based on a modification of the DiFrancesco–Noble (1984) equations. *Proc. R. Soc. Lond. B* **222**, 295–304. (doi:10.1098/rspb.1984.0065)
28 Foster, R. & Kreitzman, L. 2004 *Rhythms of life.* London, UK: Profile Books.
29 Cavalier-Smith, T. 2000 Membrane heredity and early choroplast evolution. *Trends Plant Sci.* **5**, 174–182. (doi:10.1016/S1360-1385(00)01598-3)
30 Cavalier-Smith, T. 2004 The membranome and membrane heredity in development and evolution. In *Organelles, genomes and eukaryite phylogeny: an evolutionary synthesis in the age of genomics* (eds R. P. Hirt & D. S. Horner), pp. 335–351. Boca Baton, FL: CRC Press.
31 Dawkins, R. 1976, 2006 *The selfish gene.* Oxford, UK: Oxford University Press.
32 Bernard, C. 1865 *Introduction à l'étude de la médecine expérimentale.* Paris, France: Bailliere. (Reprinted by Flammarion 1984).
33 Cannon, W. B. 1932 *The wisdom of the body.* Norton, MA: Boston.
34 Guyton, A. C., Coleman, T. G. & Granger, H. J. 1972 Circulation: overall regulation. *Annu. Rev. Physiol.* **34**, 13–46. (doi:10.1146/annurev.ph.34.030172.000305)
35 Noble, D. 2008 Genes and causation. *Phil. Trans. R. Soc. A* **366**, 3001–3015. (doi:10.1098/rsta.2008.0086)
36 Noble, D. 2010 Biophysics and systems biology. *Phil. Trans. R. Soc. A* **368**, 1125–1139. (doi:10.1098/rsta.2009.0245)
37 Noble, D. 2011 Differential and integral views of genetics in computational systems biology. *J. R. Soc. Interface Focus* **1**, 7–15. (doi:10.1098/rsfs.2010.0444)
38 Noble, D. 2011 Neo-Darwinism, the modern synthesis, and selfish genes: are they of use in physiology? *J. Physiol.* **589**, 1007–1015. (doi:10.1113/jphysiol.2010.201384)
39 Sonneborn, T. M. 1970 Gene action on development. *Proc. R. Soc. Lond. B* **176**, 347–366. (doi:10.1098/rspb.1970.0054)
40 Sun, Y. H., Chen, S. P., Wang, Y. P., Hu, W. & Zhu, Z. Y. 2005 Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. *Biol. Reprod.* **72**, 510–515. (doi:10.1095/biolreprod.104.031302)
41 Hillenmeyer, M. E. *et al.* 2008 The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365. (doi:10.1126/science.1150021)
42 Kohl, P., Crampin, E., Quinn, T. A. & Noble, D. 2010 Systems biology: an approach. *Clin. Pharmacol. Ther.* **88**, 25–33. (doi:10.1038/clpt.2010.92)
43 Kohl, P. & Noble, D. 2009 Systems biology and the virtual physiological human. *Mol. Syst. Biol.* **5**, 291–296.
44 Noble, D. 2008 Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* **93**, 16–26. (doi:10.1113/expphysiol.2007.038695)
45 Niederer, S. A., Ter Keurs, H. E. & Smith, N. P. 2009 Modelling and measuring electromechanical coupling in the rat heart. *Exp. Physiol.* **94**, 529–540. (doi:10.1113/expphysiol.2008.045880)
46 Panfilov, A. & Holden, A. V. 1993 Computer simulation of re-entry sources in myocardium in two and three dimensions. *J. Theor. Biol.* **161**, 271–285. (doi:10.1006/jtbi.1993.1055)
47 Panfilov, A. & Keener, J. 1993 Re-entry generation in anisotropic twisted myocardium. *J. Cardiovasc. Electrophysiol.* **4**, 412–421. (doi:10.1111/j.1540-8167.1993.tb01280.x)
48 Panfilov, A. & Kerkhof, P. 2004 Quantifying ventricular fibrillation: *in silico* research and clinical implications. *IEEE Trans. Biomed. Eng.* **51**, 195–196. (doi:10.1109/TBME.2003.820608)
49 Plank, G. *et al.* 2009 Generation of histo-anatomically representative models of the individual heart: tools and application. *Phil. Trans. R. Soc. A* **367**, 2257–2292. (doi:10.1098/rsta.2009.0056)
50 Trayanova, N. & Eason, J. 2002 Shock-induced arrhythmogenesis in the myocardium. *Chaos* **12**, 962–972. (doi:10.1063/1.1483955)
51 Trayanova, N., Eason, J. & Aguel, F. 2002 Computer simulations of cardiac defibrillation: a look inside the heart. *Comput. Vis. Sci.* **4**, 259–270. (doi:10.1007/s00791-002-0082-8)
52 Whiteley, J. P., Bishop, M. J. & Gavaghan, D. J. 2007 Soft tissue modelling of cardiac fibres for use in coupled mechano-electric simulations. *Bull. Math. Biol.* **69**, 2199–2225. (doi:10.1007/s11538-007-9213-1)
53 De Robertis, E. M. 2006 Spemann's organizer and self-regulation in amphibian embryos. *Nat. Rev. Mol. Cell Biol.* **7**, 296–302. (doi:10.1038/nrm1855)
54 Sander, K. & Faessler, P. E. 2001 Introducing the Spemann-Mangold organizer: experiments and insights that generated a key concept in developmental biology. *Int. J. Dev. Biol.* **45**, 1–11.
55 Spemann, H. & Mangold, H. 1924 Über induktion von Embryonalagen durch Implantation Artfremder Organisatoren. *Wilhelm Roux's Arch. Dev. Biol.* **100**, 599–638.
56 Anderson, P. W. 1972 More is different. *Science* **177**, 393–396. (doi:10.1126/science.177.4047.393)
57 Hunter, P., Smaill, B. H., Smith, N. P., Young, A., Nash, M., Nielsen, P. F., Vaughan-Jones, R. D., Omholt, S. & Paterson, D. J. In press. The Heart physiome project. *WIREs Syst. Biol. Med.*
58 Winslow, R., Kimball, A., Varghese, A. & Noble, D. 1993 Simulating cardiac sinus and atrial network dynamics on the connection machine. *Physica D Non-linear Phenom.* **64**, 281–298. (doi:10.1016/0167-2789(93)90260-8)

59 Winslow, R., Varghese, A., Noble, D., Adlakha, C. & Hoythya, A. 1993 Generation and propagation of triggered activity induced by spatially localised Na-K pump inhibition in atrial network models. *Proc. R. Soc. Lond. B* **254**, 55–61. (doi:10.1098/rspb.1993.0126)
60 Nash, M. P. & Hunter, P. J. 2001 Computational mechanics of the heart. *J. Elast.* **61**, 113–141. (doi:10.1023/A:1011084330767)
61 Smith, N. P., Pullan, A. J. & Hunter, P. J. 2001 An anatomically based model of transient coronary blood flow in the heart. *SIAM J. Appl. Math.* **62**, 990–1018. (doi:10.1137/S0036139999359860)
62 Brenner, S. 2010 Sequences and consequences. *Phil. Trans. R. Soc. B* **365**, 207–212. (doi:10.1098/rstb.2009.0221)
63 Feytmans, E., Noble, D. & Peitsch, M. 2005 Genome size and numbers of biological functions. *Trans. Comput. Syst. Biol.* **1**, 44–49. (doi:10.1007/978-3-540-32126-2_4)
64 Lewontin, R. C. 1974 *The genetic basis of evolutionary change.* New York, NY: Columbia University Press.
65 Siegelmann, H. T. 1995 Computation beyond the Turing limit. *Science* **268**, 545–548. (doi:10.1126/science.268.5210.545)
66 Garny, A., Noble, D. & Kohl, P. 2005 Dimensionality in cardiac modelling. *Progr. Biophys. Mol. Biol.* **87**, 47–66. (doi:10.1016/j.pbiomolbio.2004.06.006)
67 Nottale, L. 1993 *Fractal space-time and microphysics: towards a theory of scale relativity.* Singapore: World Scientific.
68 Nottale, L. 2000 *La relativité dans tous ses états. Du mouvements aux changements d'échelle.* Paris, France: Hachette.
69 Nottale, L. 2011 *Scale relativity and fractal space-time: a new approach to unifying relativity and quantum mechanics.* London, UK: Imperial College Press.
70 Auffray, C. & Nottale, L. 2008 Scale relativity theory and integrative systems biology. I. Founding principles and scale laws. *Progr. Biophys. Mol. Biol.* **97**, 79–114. (doi:10.1016/j.pbiomolbio.2007.09.002)
71 Nottale, L. & Auffray, C. 2008 Scale relativity and integrative systems biology. II. Macroscopic quantum-type mechanics. *Progr. Biophys. Mol. Biol.* **97**, 115–157. (doi:10.1016/j.pbiomolbio.2007.09.001)
72 Hinch, R., Greenstein, J. L., Tanskanen, A. J. & Xu, L. 2004 A simplified local control model of calcium-induced calcium release in cardiac ventricular myocytes. *Biophys. J.* **87**, 3723–3736. (doi:10.1529/biophysj.104.049973)
73 Hinch, R., Greenstein, J. L. & Winslow, R. L. 2006 Multi-scale modelling of local control of calcium induced calcium release. *Progr. Biophys. Mol. Biol.* **90**, 136–150. (doi:10.1016/j.pbiomolbio.2005.05.014)
74 Tanskanen, A. J., Greenstein, J. L., Chen, A., Sun, X. & Winslow, R. L. 2007 Protein geometry and placement in the cardiac dyad influence macroscopic properties of calcium-induced calcium release. *Biophys. J.* **92**, 3379–3396. (doi:10.1529/biophysj.106.089425)
75 Einstein, A. 1936 Lens-like action of a star by the deviation of light in the gravitational field. *Science* **84**, 506–507. (doi:10.1126/science.84.2188.506)
76 Petters, A. O., Levine, H. & Wambsganss, J. 2001 *Singularity theory and gravitational lensing.* Boston, MA: Birkhäuser.
77 Bitbol, M. In press. Downward causation without foundations. *Synthese.* (doi:10.1007/s11229-010-9723-5)

*Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht*

*Scientometrics, Vol. 69, No. 3 (2006) 615–637*

# Is chemistry 'The Central Science'? How are different sciences related? Co-citations, reductionism, emergence, and posets

ALEXANDRU T. BALABAN, DOUGLAS J. KLEIN

*Texas A&M University at Galveston, Galveston, TX (USA)*

According to a widely used introductory chemistry text by T. E. Brown et al.,[1] chemistry is 'The Central Science'. But scientometric co-citation analyses indicate that biochemistry seems presently to be more interconnected to other sciences. On the other hand, mathematics is considered by many to permeate all sciences and hence might compete as the choice for centrality. A critical commentary and argument leads to a proposal for an alternative partially ordered hierarchical "framework" map of sciences. This argument is supplemented by a scientometric approach based on university course requirements for different curricula, so as to support our partially ordered map. This alternative "framework" mapping then is seen to indicate a special position for chemistry, as where significant branching begins.

## Chemistry and scientometric inter-relations

In order to find an answer to the lead title question, the second question inextricably arises. Thence we here address these two questions, through an examination of several related topics: different proposed inter-relationships amongst the sciences, their partially ordered "framework" hierarchy, the related complementarity of reductionism and emergence, the development of interdisciplinary sciences, trends in the publication of new scientific journals, and points relating to the centrality of chemistry. It is emphasized that the title questions may be answered from two viewpoints – firstly, on

Received April 28, 2006

*Address for correspondence:*
ALEXANDRU T. BALABAN
Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 77551, USA
E-mail: balabana@tamug.edu

*0138–9130/US $ 20.00*

the basis of a historically manifested structural framework of dependences, here presented as a novel partially-ordered map of the sciences. Secondly, "scientiometric" evidence is critically considered and developed in a manner differently than in some previous work.

A widely read chemistry textbook by T. E. Brown and co-authors is entitled *Chemistry – The Central Science* and it is now at its 10th edition.[1] Indeed, this catchy phrase has been frequently used, with the rationale that chemistry is everything accessible to our senses of touch, taste, and smell. But also centrality may be argued from the intuitive idea that chemistry makes the connection between the "hard sciences" such as physics and the "soft sciences" associated with biology and medicine. Perfect or almost perfect reproducibility of experiments is expected in the hard sciences, whereas multiple experiments are the standard mode in the soft sciences in order to ensure a reliable statistical basis for coping with the poorer reproducibility.

As a parenthetical remark, an alternative textbook title *Chemistry – The Molecular Science* (by Moore, Stanitski, and Jurs, 2005[2]) appears too restrictive within a common understanding of "molecular". That is, not all materials are well characterized by a particular reasonably finite interconnection pattern of atoms – there being many (even common) essentially infinitely extended, or amorphous, or non-stoichiometric materials which are ill-characterized as molecules but which nevertheless are chemical. Note that among the common types of stronger intermolecular interactions it is primarily covalent bonding which leads to the formation of (non-extended) molecules. Most of the millions of substances listed in the Chemical Abstracts Database are organic molecules, but organic compounds also include many ionic salts. Thus, considering the chemical constitution of various forms of matter in our planet, the earth's core (metallic iron-nickel) and many rocks in the earth's mantle such as the ionic limestone ($CaCO_3$) are not molecular compounds, but water, air, petroleum, and the bulk of the biomass are (although supramolecular interactions also intervene in biomolecules). What Moore et al. evidently take as a definition of "molecular" is simply a collection of atoms – though again this meaning is not typically always so understood.

Another abbreviated definition of chemistry by van Brakel[3] is as the "science of the transformations of substances". But in fact often much effort (both in teaching and in research) is just spent on describing and characterizing what structures occur, aside from their mode of formation and reaction. But also nuclear matter, neutron stars, black holes, and the hypothesized "dark" matter are all often viewed as substances, albeit exotic non-chemical substances – and all of these (presumably even dark matter) undergo transformations.

To address "centrality", the inter-relationships of one sort or another amongst the different scientific fields is relevant. One type of inter-relationship yielding an objective mapping of scientific fields was initiated in 1973 by Small and coworkers[4-6] and independently by Marshakova[7] – both using co-citations which were revealed by

Science Citation Index (SCI). The results were then published in book form some years later as the encyclopedic *Atlas of Science* with several volumes covering various scientific areas.[8] The methodology was discussed in detail by Small[9] and Garfield.[10] In 1974 Aaronson presented a similar map for biomedical papers that had been published during the two years preceding the citations.[11]

Recently, Moya-Anegón et al.[12] and Boyack et al.[13,14] published maps argued to reveal a "backbone of science". The latter authors, after analyzing co-citations and inter-citations from over one million documents, concluded that it was not chemistry, but biochemistry which appears as the most interdisciplinary discipline in science. However, from Figure 2 of the former authors (on the basis of an analysis of co-citations from over 26,000 documents published by Spanish authors starting with the year 2000)[12] one can see that there are two sciences that serve as bridges (cut-points) between the hard and the soft sciences, namely chemistry and mathematics.

From these objective scientometric analyses there are three candidates for centrality among the sciences: biochemistry or mathematics or chemistry. On the other hand, the argument of Brown based on casual observations "bridging" and widespread usage of concepts and techniques indicates that chemistry is the central science. Thence the questions arise as to whether Brown's arguments might be further validated, perhaps through an alternative mapping interrelating the sciences. In fact there are other differently based (and even older) schemes for inter-relating the sciences, as we next describe.

## From Comte's classification of intellectual pursuits to partial orderings

Auguste Comte, in his *Course of Positive Philosophy* of 1830, argued that sciences may be ordered in a linear manner such that earlier sciences provide more general frameworks within which later sciences are to be placed.[15] This then seemingly suggests an order as in Figure 1, with mathematics in the lead, followed by astronomy, physics, chemistry, physiology and medicine, and ending with the social sciences.

The ordering from mathematics to physics is made on the basis that the fundamental laws of physics are to be made in a mathematical framework. Presumably there is little argument otherwise. This ordering is clearly suggested by P. A. M. Dirac, who developed the fundamental transformational view of quantum mechanics, as well as a (special) relativistic quantum formulation, encompassing anti-matter, and who won the Nobel Prize for Physics in 1933. In particular Dirac wrote:[16]

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

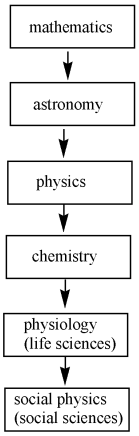


Figure 1. Comte's classification of the Sciences

The ordering from chemistry to biology is similarly nicely illustrated by a statement from Arthur Kornberg (who shared the 1959 Nobel Prize for Physiology or Medicine with Severo Ochoa) for the discovery of the mechanism of the biological synthesis of ribonucleic acid and deoxyribonucleic acid):[17-19] He repeatedly argued that chemistry is the *lingua franca* of medicine and biology, and deplored the gap between the two cultures – chemistry and biology (paraphrasing C. P. Snow's dichotomy between scientists and humanists). Thence Kornberg wrote:

"Life, after all, is only chemistry, in fact, a small example of chemistry on a single, mundane planet…. It is my long and deep conviction that life, including human cognition and behavior, can ultimately be expressed in the language of chemistry. By this reduction we can achieve a more rational and richer appreciation of the living world and our place in it…."

Kornberg's view seems stronger than Dirac's in that there seems less qualification as to any short-comings of the higher level view – in Kornberg's case the chemical view of biology. And this is further emphasized in Kornberg's follow-up comments:

[Today] "Research and teaching in all the basic [biomedical] science departments are interdependent and really indistinguishable. The current unity of the basic medical sciences has come about because these previously diverse disciplines are now expressed in one common language, the language of chemistry….The major obstacle, even for physicians, is to accept that mind is matter and only matter…"

And qualification in Kornberg's opinion finally emerges slowly, when further he says:

"Chemical language has great esthetical beauty and links the physical sciences to the biological sciences. It is an international language, a language for all the time, a language without dialects. It is language that explains where we came from, what we were, and where the physical world will allow us to go."

That is, it seems Kornberg sees chemistry related to biology just as he sees physics related to chemistry. This then also supports Comte's classification of Figure 1.

But besides mathematics, physics, chemistry, biology, and the social sciences, there arguably are other sciences. Astronomy often has been viewed as separate from physics, though (especially on a grand scale) it might be viewed as part of physics. The same seems less clearly to be so for several Earth sciences (geology, oceanography, and meteorology) which are dependent on our particular Earthly environment. Of course the great bulk of our biology is referenced to Earthly circumstances, though in a rather different way – and somewhat the same might be said of ecology. And yet further there are different further sciences related to the human condition: engineering and computer science which together are insightfully viewed by H. Simon (1978 Nobel Prize winner for Economics) as "the science of the artificial".[20] But further there are the agro-sciences, general health science, economics, and perhaps even law, which also could be placed under the science of the artificial.

It seems quite difficult to place all these diverse sciences into a linear order, though it is a little easier to imagine a *partial ordering* – with the ordering still being that earlier sciences are viewed to form the framework in which the later ones are to be placed – just not demanding that every pair be so ordered. Such a partial ordering then appears as indicated in the diagram in Figure 2, where the higher sciences connected *via* a line (or sequence of lines) are to form the framework for those lower in this hierarchical picture. [For a general partially ordered set (poset), such an arrangement is termed a Hasse diagram. See, e.g., refs. 21-23]

Following ratiocination mathematics is presumed to be a parent overarching structure, with every other field of knowledge becoming science in so far as the ideas can be formulated mathematically (and logically). It is emphasized that the precursor ratiocination here encompasses much beyond mathematical logic to include "inference", as involved in scientific induction, and also to include analogy (and even metaphor), whereby different things can be seen as "like" in some sense, with the consequence that different objects or actions or modifiers can be represented by different language, codes, or symbols. Further mathematics here is not restricted to be that of numerical manipulation, but is also to deal with classification (i.e., equivalence classes), partial orderings, graph theory, group theory (and symmetry), statistics, and (much) more.

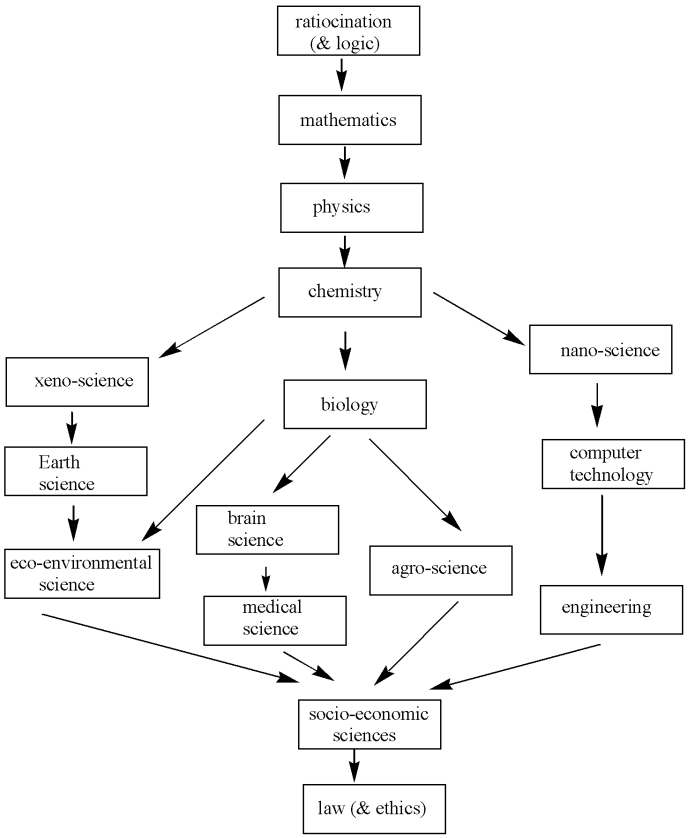


Figure 2. The "framework" partial ordering of the sciences

Cosmology, astronomy, and planetary science have been subsumed under xeno-science, and on our diagram this does not appear directly under physics, but rather under chemistry, as the molecules of interstellar space and stars are studied, as well as the chemical compositions of various planets, moons, asteroids, comets, and meteors in our solar system. Xeno-science then formally serves as an ancestor to the Earth sciences (encompassing geology, oceanography, and meteorology).

Computer science and computer technology are rapidly expanding areas, with computer science a more theoretical mathematical area, concerned not only with different useful algorithms but also with artificial intelligence, and perhaps also with (the theory of) quantum computation. Many (or arguably all) algorithms may be classed as mathematics, but some (e.g., as to do with sorting and listing and parallel processing) seem so very intimately related to the usage of computers and are so exceptionally

important that they might conceivably be separately identified to computer science, though we do not do so in our figure. Computer technology has to do with the devices making the computations, and might be viewed as part of engineering, though again so important now (and more so in the future) that this may be viewed separately. Knuth[24] pointed out that traditional areas of engineering have led to factors of ~$10^3$ or so of increase in speeds of travel or in weights of loads so movable as compared to what a single human can manage, whereas computational speeds under the guidance of computer technology have increased by a factor of ~$10^{12}$. Moreover, Knuth suggested that the societal changes so wrought by computer technology are likely to ultimately be "proportionately" greater than by traditional engineering.[24] Thence with the impact of the "computer revolution" anticipated to so exceed that of the historical "industrial revolution", it seems appropriate to distinguish computer technology. Thence in our figure computer technology has been separately listed (and sometimes one might wish to distinguish computer science from mathematics).

Nano-science might be construed as part of chemistry though we anticipate that ultimately it will manifest great importance, ultimately with diverse applications. It may be viewed as the lead (most fundamental) of the "sciences of the artificial". This, along with the great overlap with physics and engineering, suggests that it might be separately identified.

Ecology and environmental sciences are distinguished from the traditional Earth sciences, because of the strong biologic dependence and singular problems of complex interactions. In some broad sense eco-environmental science might be viewed as the "science of structure in complex systems" – such systems being characterized by a diversity of distinctive parts, which complexly interact with one another, and also have internal structure. This is somewhat more than ecology in that here it has been taken to encompass environmental science which has to do with such complex systems with less of a "biologic" tone, though in fact the "biology" seems to us to often be too much suppressed in much of "environmental science". The science of the brain, encompassing neuroscience & psychology (and perhaps even biologic systems in general) would seem to qualify as involving complex systems, though for such systems there is a great deal of special character and there is a singular amount of special (separate) interest.

One may perceive many of the same aspects of this overall partial ordering of the sciences reflected in E. O. Wilson's *Consilience* which argues[25] for such a unifying interconnected framework, also focusing much on ecology and relations to it including not only law and ethics but also both art and religion. [For his overall theme of a grandiose unification, Wilson does not say much about mathematics or the physical sciences, including engineering, and he is rather abbreviated in indicating how art fits in to his view, though he asserts that it does.] Still, though Wilson does not present explicit diagrams as in figures 1 or 2, his discussion is supportive of the general overall ordering.

A more explicit partial ordering somewhat like an abbreviated form of Figure 2 is found in Ellis' review.[26] In particular, Ellis' diagram quite clearly shows chemistry as central in that it is where he identifies a single bifurcative branching to occur, to lower level sciences, one branch (biochemistry, physiology, psychology, sociology, and ethics) which is human oriented while the other branch (materials, geology, earth science, astronomy and cosmology) is not so limited in orientation. Though this bifurcation is the sole one occurring in Ellis' diagram, it is manifested to some extent in our partial ordering, although our diagram emphasizes that the inter-relations are somewhat more complicated with the bifurcated branches eventually leading back together in the "softer" societally directed sciences. And also the "sciences of the artificial" occur as the right-hand side of our diagram. Indeed "our" diagram may be perceived in some form in many other (earlier) contexts – and again there is a degree of centrality for chemistry, where branching begins.

Our diagram, with higher linked sciences providing a framework for those lower in the diagram, thence indicates a fundamental inter-relation of the different sciences. Chemistry is seen to be "central" in being at a location where significant branching begins. Biology appears nearer the center of gravity of this diagram. But there may be more to the idea of "centrality", as well as to the idea of "inter-relationships" between sciences.

## Qualifications to the hierarchy of sciences

There are further relevant qualifications as to the ordering of sciences indicated in Figure 2. In particular it is to be understood that there is feedback of some sort in the reverse directions in this diagram. Many of the problems, techniques, and foundational theories may be motivated by or even "derived from" other sciences lower in the diagram. This is clearly illustrated in the relation between physics and mathematics, where calculus was developed by Isaac Newton, who is typically considered first as a physicist. (An independent developer of calculus, W. Leibniz seems to be generally recognized more inseparably as a scientist, philosopher, and mathematician.) Following a tradition dating back to Pythagoras, renaissance scientists Galileo, Kepler, Descartes, Leibniz, and Newton all believed in an utterly fundamental (perhaps even divine) role of mathematics in the design of the physical universe. Indeed, it is clear from the name that geometry has its historical foundations in the study of the "science of space" (as is a part of physics). And this view is not confined to antiquity, *e.g.*, as is seen in János Bolyai's foundational commentary[27] of non-Euclidean geometry, where he announces that he is addressing a scientific (physical) question, even in his commentary's title: *The Science of Absolute Space – independent of the truth or falsity of Euclid's axiom XI* (*which can never be decided a priori*). And following Bolyai, one finds Bernhard Riemann introducing (in 1854) his even more general view[28] of geometry in terms of

what are now termed "Riemann spaces", and particularly utilizing a physical motivation for his general metric structure, with the comment: "The basis of metrical determination must be sought outside the manifold in the binding forces which act on it". And thereafter one finds A. Einstein's special theory of relativity of 1905 enlarging geometry to a theory of space-time, which then was reworked into a fundamental mathematical framework by H. Minkowski.[29] And yet again following Einstein's general theory of relativity in 1916, there is mathematical development surveyed by H. Weyl.[30] This interaction forth and back does not seem to be slowing, with e.g., quantum mechanics engendering the mathematical study of von Neumann algebras, C*-algebras, Gelfand triples, rigged Hilbert spaces, gauge fields, and more. Recently a nice small example comes from the idea[31,32] of "quantum computation" arising from physics within the last decade or two, with most of the theoretical development in this field to date having been done by physicists, though one can expect that ultimately the general theory will be incorporated into mathematical computer science. Indeed it may be argued that mathematics actually arose from physics – via a process of abstraction. Evidently in many cases the ideas in physics provide a motivation for the mathematical ideas. Even the development of something as fundamentally mathematical as the abstract idea of "function" has very much to do with physics, as is seen from Luzin's insightful historical review.[33] Thus Vizgin[34] describes the views of several famous mathematicians and physicists as to the intimacy of connection between mathematics and physics, e.g., quoting David Hilbert[35] as to the " 'preestablished harmony' of which 'general relativity and quantum mechanics' are the most magnificent and remarkable examples". And yet further there is Dirac's declaration[36] of enthrallment:

"It seems to be one of the fundamental features of nature that fundamental physical laws are described in terms of a mathematical theory of great beauty and power, needing quite a high standard of mathematics for one to understand it. … Why is nature constructed along these lines? One can only answer that our present knowledge seems to show that nature is so constructed. … One could perhaps describe the situation saying that God is a mathematician of a very high order, and He used very advanced mathematics in constructing the universe."

Even more recently this intimate relationship is seen to be on-going, e.g., as is witnessed by Roger Penrose's recent impressive book *The Road to Reality*, where perhaps half the chapter titles might be identified to concern new foundational mathematics, and the other half to concern new foundational physics, though each chapter has much about both subjects, inextricably mixed.[37] Penrose follows Platonic ideas in viewing mathematics simply as the root influences governing "Nature". As such mathematics is deeply entangled with a systematic understanding of the physical world, and the derivation (or deduction) of mathematics from the physical world is not unsurprising. A perhaps less personal view is found in Monastyrsky's survey of *Mathematics at the Turn of the Millennium*,[38] where he emphasizes that before the

twentieth century mathematics and physics were so connected that "it was difficult to separate one field from the other", though in his analysis he goes on to claim that "the connection was broken in the 1930s" but that subsequent developments "restored the relationship". Both Penrose and Monastyrsky build from the millennia-old physico-geometric interconnections to focus on the last century up through the current time, with opinions as to future developments. Generally the interconnection between physics and mathematics is very deep and very strong. It is so deep and so strong that one must surely question just what the earlier mentioned scientometric articles fundamentally might be measuring if they miss this prominent connection – arguably the most prominent connection between the different scientific fields. In fact, it arguably is precisely this math/physics interconnection which historically gave birth to modern science.

But back-and-forth interconnections also arise between other neighbor pairs of sciences, though typically not so manifest as between physics and mathematics. Perhaps agro-science and medicine are so intimately connected to biology, with these two applied sciences arguably being the historical motivational source for much of biology (much as physics has been argued to be the historical motivational source for mathematics). Nevertheless we see biology as forming the fundamental framework for the agro-sciences and medicine.

There also are interconnections at a finer (less grand) scale. Thus there are intermediate fields of physical chemistry (going back somewhat over a century) and of chemical physics (going back in a recognized form for about 70 years). And areas exploding dramatically in the recent past are molecular biology (really starting about 60 years ago) and biochemistry (with an older history, but with recent great growth). Another example of a somewhat controversial intermediate field (advocated around 30 years ago) is sociobiology. And several fields given separate recognition in the diagram of Figure 2 may be viewed to have appeared as a result of such an interdisciplinary development, e.g., as computer technology or nano-science. These provide two examples of interdisciplinary fields which have arisen primarily from the bringing together of ideas from other separate sciences (physics, chemistry, and engineering). But also interdisciplinary fields can arise via another mechanism (either in whole or in part) – via specialization. One such is mathematical physics where before a century ago, it was perhaps viewed that a mathematician rather often could also do fundamental research in physics, and that a (theoretical) physicist could also do fundamental research in mathematics – while more recently specialization occurred, resulting in a "crystallization" of this field from the ambient mathematico-physical traditions.

Sometimes the developed special fields are intermediate between pairs of fields not very closely connected in the hierarchical diagram indicated. Such is notably the case with computer technology, so that there are numerous computer fields arising as an intermediate derivative of almost every one of the other sciences in our diagram. But

each one of these deriving fields already has its own works phrased in terms of the nomenclature and ideas, while comparatively little attention is directed to the fundamental detail of the hardware of the computer technology, or of any wider applicability of any methods outside of the context of the deriving field. E.g., in computational quantum chemistry the great bulk of the work is framed purely in theoretical chemical terms (SCF theory, many-body perturbation theory, density-functional theory, natural orbitals, electronic densities, etc.) which are entirely unfamiliar to computer technologists. That is, such "computer fields" are here identified as part of the deriving application field (i.e. the field other than computer science) without drawing in an interconnection in our diagram. There does not seem to be much "scientific" feedback to computer technology – though there is another type of feedback: there is an ever increasing demand for more powerful computers and computational algorithms. And also at least in the dramatic founding stage of computer science and technology a few decades ago, most computer scientists emigrated from other scientific areas (quite clearly because initially there were so few trained as computer scientists). One may see this last effect going on at a lesser rate of evolution in the field of oceanography, where a few decades ago most oceanographers were trained in another field, while now almost all new oceanographers have a PhD in oceanography, and it is even beginning that there are undergraduate degrees in the area.

It is seen that our diagram of Figure 2 does not manifest all the types of relations between the different sciences. That is, the space of inter-relationships is multi-dimensional and likely complicated. In particular, we can see that the citation and co-citation maps while having some degree of correspondence with our hierarchical diagram, evidently at least in some cases identify something rather different. Still, Figure 2 does identify the general directed relationship that earlier sciences form frameworks for later ones, and it clearly identifies this as a partial ordering. Moreover, it is emphasized that the mode of presentation of the partial ordering does not preclude the framework of a given science to be solely that to which it is directly connected, the framework including all those higher up on the diagram. Thus beyond chemistry, physics is viewed (along with mathematics) as part of the framework for biology, e.g., as even may be emphasized in some presentations, such as in Max Delbrück's overarching survey *Mind from Matter?*.[39]

Further there are qualifications as to "reductionistic" character of the relations indicated in the hierarchical framework diagram. In particular this diagram does *not* presume that the lower fields can be fully reduced to those foundational ones at a higher level. That is, the results in these various lower fields are each presumed to manifest *emergent* properties and related ideas and theories. Thence Darwinistic evolution in biology is an emergent property, not clearly present in "higher-level" sciences. In chemistry the Mendeleevian periodic table (and consequent periodic relations) are effectively argued to be emergent.[40,41] And much of thermodynamics can be seen as

emergent, particularly temperature, entropy, and relations involving these quantities. Again the higher-level sciences in this hierarchical diagram are to be such that the lower level ones are to be framed in terms of them, using them, and building from them – without the requirement that all results neatly follow in a purely logical way from the earlier. In an artistic metaphor, the frame, canvas, and paint are to the picture much as the earlier sciences are to the later sciences. Dirac's earlier quote[16] framing chemistry in terms of physics leaves an opening for this when he says that what arises are "…equations much too complicated to be soluble", though in reality it is more than a vast set of individual solutions which is desired – the desired thing being the emergent principles – as the ideas of "chemical bonds", of "periodicity", etc. which makes chemistry "more" than physics. Thus C. A. Coulson argued[42] that extensive quantum-chemical computational output appears much like naught but further experimental data, which then is to be processed by theory, which then presumably is "emergent". And Kornberg's earlier quoted ultimate qualification[18] that "Chemical language … links the physical sciences to the biological sciences" is even more appropriate in formulating chemistry as the background in terms of which biology is framed. E.g., evolution[43] ends up being framed in terms of mutations at the chemical level of modifications in the genetic DNA – though there is much more to evolution, in terms of competition at the macro scale and the "survival of the fittest" with much elaborate mathematical modeling possible – and on-going. See, e.g., Van Regenmortel[44] for a discussion emphasizing (and illustrating) that biology is not fully reducible to chemistry. The general idea of emergence (or of "emergentism") has been widely discussed in the philosophy of science literature,[45,46] with there being several different stages identifiable between full reductionism and overbearing emergence, which is more often described as "holism". There are now extended more popular discussions of emergence,[47,48] with one iconoclastic discussion being R. B. Laughlin's *A Different Universe*[49] – in particular this book has nice discussions of emergent behavior involving superconductivity and the quantum Hall effect. Reductionism and emergence may be viewed as complementary aspects of science. Further elaboration and examples we do not pursue here.

In conjunction with these ideas of emergence a further sort of qualification is noted in that the lead subject in our diagram was "ratiocination", rather than simply mathematics. This provides a type of qualification in the idea that there is much deeper background behind mathematics, where analogy and representation and expression are the fertile ground on which the later subjects are built. This block is unlike the others in the diagram, in that it does not represent what is a well recognized science – though beyond "mathematical logic" there are aspects of this block formally covered in topics of "inferential logic", Chomsky's foundational "linguistics", and especially "artificial intelligence".

## Interdisciplinary developments yielding border-line fields

Progress in scientific research is especially likely to appear at contact points between scientific fields, giving rise to "hyphenated sciences". Pairwise combination between the main sciences (mathematics, physics, chemistry, biology, and medicine) has given rise to well-defined interdisciplinary areas, with their own journals. Such a development between physics and mathematics we have already described in the preceding section, but we elaborate on a few more, especially those involving a relation to chemistry, our candidate "central" science.

Historically physical chemistry developed to involve electrochemistry, then thermodynamics and kinetics, and now has been well recognized for over a century, with its own journals (*Zeitschrift für physikalische Chemie*, *Journal of Physical Chemistry*, *Journal de chimie physique et du radium*, *Transactions of the Faraday Society*, *Zhurnal Fizicheskoi Khimii*, and more). But in the 1930s an interesting development of a seemingly further field of chemical physics occurred. At that time, a powerful clique of (American) physical chemists in control of the *Journal of Physical Chemistry*, insisted that physical chemistry consisted of thermodynamics, kinetics, and electrochemistry, while excluding spectroscopy, quantum theory, and even statistical mechanics. Another group of (primarily) chemists managed to found the *Journal of Chemical Physics*. Then around this journal a field of chemical physics congealed, with several further journals in later decades, and after a couple of decades physical chemistry began more to accept the new topics and methods. As a consequence, journals like *J. Chem. Phys*. and *J. Phys. Chem.* came (in the last few decades) to be much less distinguishable in the type of articles published, and recently there have appeared journals like *Phys. Chem. – Chem. Phys.*, and *ChemPhysChem*. A much more complete discussion of the development of physical chemistry and then chemical physics is found in the histories of Servos[50] and of Nye.[51] A further yet more specialized area that has developed over the last half century is that of quantum chemistry, even with its own journals, as *International Journal of Quantum Chemistry*. It is often viewed that chemical physics as a whole has two foundational aspects: quantum chemistry and statistical mechanics. The statistical mechanical wing seems to have a somewhat stronger foot-hold within physics proper.

Even intermediates between sub-disciplines may emerge, as with the case of physical organic chemistry, emerging a little over a half century ago. This deals much with property/structure correlation for organic compounds and with organic reaction mechanisms. It has much overlap to theoretical organic chemistry and to photochemistry. Now there is a series *Advances in Physical Organic Chemistry* (with over three dozen volumes), another series *Progress in Physical Organic Chemistry* (with about two dozen volumes), and some journals, as well as dozens of texts (with that of L. P. Hammett in 1940 perhaps marking the field's inception).[52]

The founding of mathematical chemistry is also quite interesting, it arguably being "submerged" in physical chemistry and chemical physics for perhaps almost a century. But then following several preliminary intimations[53] over several decades, the field was finally formally recognized.[54] First a journal of mathematical chemistry, *Communications in Mathematical Chemistry* (*MATCH*), was founded in 1975. The *Journal of Mathematical Chemistry* was founded in 1987, and also (in 1987) the *International Society of Mathematical Chemistry* was inaugurated. In 2005 the *International Academy of Mathematical Chemistry* was founded, so far with about 50 elected members. The general focus in these various developments was on graph-theoretic aspects, such as are so peculiar to chemistry, starting with chemical-structure ideas in the 19th century. Yet further a *Mathematical Chemistry Series* on edited monographs was initiated, with over a half-dozen volumes up to the present, covering a wide range of mathematical topics in chemistry.

Medicinal chemistry and pharmaceutical chemistry have developed immensely, and each have their specialized journals, illustrated with *J. Med. Chem.*, *J. Pharm. Chem., J. Pharm. Sci.,* and many other journals. The recent appearance of *ChemMedChem*. provides a further example (paralleling in some manner *ChemPhysChem* and *Phys.Chem.-Chem.Phys.* mentioned earlier) of how interdisciplinary fields tend to blend and intermingle. And related to these two fields is molecular biology (which seemingly may develop to encompass both medicinal and pharmaceutical chemistry) with too many journals to be mentioned here.

Further on the applied side, one finds agro-chemistry. With an impressive history starting a century or so ago, with the development of crops to re-nourish the soil, and the finding of commercial uses for such re-nourishing crops (such as peanuts, sweet potatoes, and soybeans – as so seminally researched and advocated by George Washington Carver). Again there are numerous research journals in this area, too many to list.

With the development of electronics and materials science (both disciplines benefiting enormously from physics and chemistry) computers now allow a wide range of approximate solutions for the Schrödinger equation, so that the "computational chemistry" sub-discipline of mathematics, physics and chemistry has emerged, with several of its own journals: *Journal of Computational Chemistry*, *Computers in Chemistry*, and *Journal of Chemical Information and Computer Science* (which continues under the name *Journal of Chemical Information and Modeling,* along with a further derivative journal named *Journal of Chemical Theory and Computation*). But also one finds much similar computational work in *Theoretica Chimica Acta, Theochem*, *International Journal of Quantum Chemistry,* in addition to many journals in physical chemistry and chemical physics. Many of these have also been receptive to mathematical chemistry.

Another interdisciplenary area developing explosively within the last two decades is that of nano-science, including nano-chemistry, nano-physics, and nano-technology, even with much overlap to molecular biology. This field involves specifically designed nano-scale structures, especially to be utilized in nano-devices in much the same way that macro-structures have been utilized in traditional engineering. The novel materials and their preparation are naturally a focus of nano-chemistry, while their novel physical properties in addition to being addressed by physicists and engineers are also addressed from a physical chemistry perspective. Interestingly the preparation of such materials *via* vapor phase deposition and especially via molecular-beam epitaxy are at least as much addressed by scientists with a physics background as by scientists with a chemistry background. Thence a new separate subject seems to be emerging with its own areas of expertise different somewhat from its parents. Already there have developed several new journals – and numerous books.

There are also recognizable emerging fields of chemical oceanography, marine chemistry, mineral chemistry, and immunochemistry. Thence there is considerable evidence for the contention that chemistry is "the central science". That is, a considerable degree of branching occurs at chemistry.

As a final bit of evidence we might quote from the web page of the University of York: "Chemistry is everywhere in the world around us – in what we eat, in what we breathe, in how we live, in what we are. Chemists study not only what things are, but also what they do and how they do it, from sub-atomic particles to large arrays of molecules. Chemistry is also very creative: new substances are made every day in a constant search for new materials, new drugs, and new ways of doing things better and new understandings of how Nature works. *Chemistry is central to all sciences* and is used to study biological, physical, medical and environmental phenomena as well."

The above argument that chemistry is central (because everything we touch, taste, and smell is chemical) is a little different than the argument as centrality being where bifurcation occurs – though in fact these two features are not unconnected. That is, the anthropocentric sciences involve intimately (and extensively) what we are able to "touch" and are so are founded on the science of what we touch. A significant qualification however applies to the above underlined assertion which would imply that chemistry is centrally pre-eminent with respect to physics (as well as mathematics and ratiocination if these be admitted as sciences). But this is contrary to our hierarchy of sciences, unless the above statement is intended to mean that chemistry is connected in a non-directional manner, perhaps not unlike the mode indicated in the citation and co-citation diagrams – though then as already noted, there are other candidates for centrality.

## Scientometric evidence for a partially ordered hierarchy of sciences

Evidently our argument for the centrality of chemistry rests on the observation that chemistry concerns everything we touch and further on its location as the point of initial branching in our partially ordered hierarchy of sciences. Thence it would be nice to have a more objective scientometric evidence of this latter claim – in-as-much-as the scientometric citation-based diagrams of Small and Garfield do not seem to reflect this hierarchy so clearly.

A comparatively "objective" manner of constructing a "framework" interconnection pattern would be to look at university course requirements for students in different fields. That is, a dependence of science X on science Y may be perceived when there is a requirement for students majoring in X to take a course in Y. Thence as a preliminary illustration of this idea we have done this for one university, namely our own parent university: Texas A&M University.[55] In Table 1 we thence list row by row different majors, and the number of semesters of course requirements in different other scientific fields, which presumably should turn out to be the "framework" fields lying at a higher level in our diagram. These requirements for a Bachelor of Science are subdivided in the table to list the number of semesters of lower and upper division required courses separately. We report the requirements for degrees in several of our sciences in an averaged manner when more than 1 department seems to correspond to a science as identified in Figure 2. Thus requirements for chemical, civil, electrical, mechanical, and industrial engineering are reported as an average, with the electrical engineering requiring 1 more semester of physics than the others, and chemical engineering requiring 4 more semesters of chemistry than the others. This and further averages are indicated in the footnotes to Table 1. In fact in all our cases of averaging one might surmise from an examination of their other science requirements that the so averaged results must be very closely related because of the great similarity in requirements – e.g., "applied mathematics" differs from pure mathematics, in replacing 4 of the mathematics courses with 3 statistics courses, and one economics course. In biology, zoology and genetics differ from the other biology departments in not requiring a course in statistics (which we have categorized as mathematics). A further point in this table is that we have not reported broad science requirements which have not been specified to a particular department – e.g., many of the natural science degrees require a behavioral or social science elective, without specifying which department it is to come from – and such requirements we have not counted in the Table 1. Besides not having a particular department to identify this to, these requirements presumably are made for the purpose of making a well rounded individual, much as the university wide requirements of "kinesiology" or of political science. Evidently these broad requirements do not identify particular material needed for the parent science.

From this table our earlier argued deep connection between physics and mathematics, is evidenced on the extensive mathematics requirements of physics majors – more so even than the biology requirements concerning chemistry. And yet further this physics/math connection is somewhat manifested with a requirement of 2 semesters of physics for mathematics majors, this being one of the few examples of a lower-level requirement in a science in our hierarchy of a science at a lower level. [In fact there is another exception, with mathematics requiring 1 semester of a modeling course in oceanography – though we suspect that this is an "aberration" peculiar to Texas A&M University and would almost never show up in other University's mathematics requirements – if for no other reason than that most Universities do not have an Oceanography department.] The requirement of 2 semesters of biology in chemistry comes from treating chemistry requirements as an average of "chemistry" and "bio-chemistry" curricula – it not being at all surprising that biochemistry has some biology requirement – indeed it might be deemed surprising that it has so little biology requirement as compared to chemistry (and biochemistry) courses. Another sort of exception, this time across a parallel branch in Figure 2, occurs for the Earth Sciences, where one of the constituent departments (Marine Sciences) requires two semesters of biology – there being a rather different tradition in oceanography (as compared to Meteorology or Geology) of considering biological influences in the ocean, where in fact there is more manifest influences from the biosphere – some of the problem here has presumably arisen from a blurring between Earth sciences and eco-environmental sciences.

Table 1. Summary of requirements by subject area (Texas A&M University)

| | Math | | Physics | | Chemistry | | Biology | |
|---|---|---|---|---|---|---|---|---|
| | lower | upper | lower | upper | lower | upper | lower | upper |
| Mathematics[a] | 6 | 11 | 2 | | | | | |
| Physics | 4 | 4 | 4 | 10 | | | | |
| Chemistry[b] | $3^+$ | | 2 | | $6^-$ | 8 | $(1^+)$ | |
| Biology[c] | 2 | 1 | 2 | | 4 | 2 | 4 | 5 |
| Earth Sciences[d] | $3^+$ | | 2 | | 2 | | $(1^-)$ | |
| Eco-environmental Sci.[e] | 2 | 1 | 2 | | 3 | | 2 | 1 |
| Computer Technology[f] | 2 | 7 | 2 | | 1 | | | |
| Engineering[g] | $4^-$ | $1^+$ | 2 | | $1^+$ | $1^-$ | | |
| Brain Sciences[h] | 2 | 1 | | | | | 2 | |
| Socio-economic Sci.[i] | 2 | 1 | | | | | | |

[a] average: mathematics, applied mathematics, computer science
[b] average: chemistry & biochemistry
[c] average: biology, microbiology, molecular/cell biology, genetics, botany , zoology
[d] average: geology, metereology, marine sciences, earth sciences
[e] average: environmental geosci., plant/environmental soil sci., renewable natural science
[f] average of computer engineering/CS & computer engineering/EE
[g] average of chemical , civil, electrical industrial, & mechanical engineering
[h] psychology
[i] average of sociology, political science, anthropology, & economics

In Table 1 there are some scientific areas only poorly represented, or even absent. Medical sciences do not appear in our current presentation in that this is virtually always pursued solely as a graduate degree, with our present tabulation restricted to undergraduate (Bachelor of Science) degrees. And a similar result occurs for law, and for what we have termed "Xeno-science" and "Nano-science". For many universities oceanography is similarly treated purely as a graduate program. Indeed some brief examination of requirements for graduate degrees seems to indicate that our "framework" relations would not be so clearly manifest there – with there typically being much more specialization within the department of the major, and further any extra-departmental requirements not being required of all graduate students, but being just for those within a specialized subfield in a department. E.g., though a physics course or two might be required of a physical chemistry graduate student, such would not be required of other chemistry graduate students. A final example of a field perhaps poorly represented is eco-environmental science, with there being no ecology department at Texas A&M University and the environmental science departments focused more on *using* the environment and its natural resources than on *studying* the environment.

Overall Table 1 seems (at least roughly) to parallel our hierarchical diagram of Figure 2, and thereby it tends to support earlier arguments presented in connection with this figure.

More explicitly one may use the data of Table 1 to empirically draw out a "framework" diagram. That is, between nodes representing the different scientific fields, one would place directed interconnections between the different fields – the direction being from field X to field Y if a degree in Y demands a course in X, a result which (for $X^1Y$) we then denote as X>Y. We could even weight the interconnections in accordance with the number of semesters of requirements, and perhaps also the level of the requirements. We choose to weight the upper division requirements at twice the weight of the lower division ones. The result of a drawing instituting just these features looks very "busy" in that many interconnections often occur from several of the higher-level sciences to a lower-level one. But the directed feature of our Figure 2 is fully manifested, with the exception of the mathematics requirement of 2 semesters of physics (and the 1 semester of oceanographic modeling, such as we wish to ignore as an "aberration"). The "busyness" of non-neighbor interconnections was avoided in our Figure 2, with the understanding that the only explicitly drawn such connections were between neighbors. Paralleling this, a step towards an avoidance of "busyness" in our empirical diagram may be made by first identifying the empirical relation which with but a few exceptions (involving mathematics requiring physics and oceanography, and marine sciences requiring biology) is in fact a partial ordering. That is, with this exception, one never finds (for $X^1Y$) both Y required by X and X required by Y. As a second step to "busyness" avoidance, we then allow subtraction of interconnections

from a level X to a level Z two steps down by the number of interconnections from X to the intermediate level Y (where X>Y>Z). This subtraction is rationalized in that Y in providing a framework for Z has its own framework in X which then is to some extent inherently required for the framework for Z – and will then be implicitly so understood much as in our Figure 2. The subtraction is taken with upper division courses counting double the lower division ones. The result is shown in Figure 3.

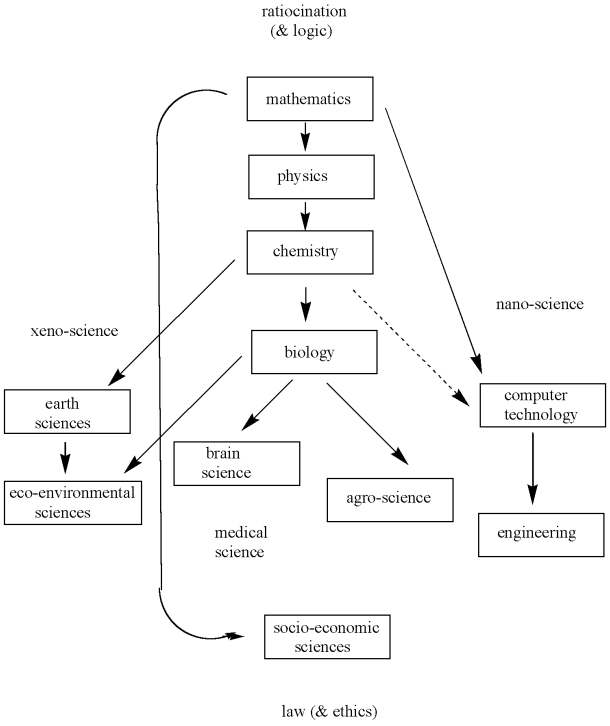


Figure 3. Curricular requirement partial ordering for Texas A&M University (at College Station)

This empirical figure is indeed seen to somewhat closely parallel our earlier Figure 2, with the understanding that a few sciences (medical science, xeno-science, nano-science, logic, and law) are absent for reasons already mentioned. Of the then encompassed sciences the connection from chemistry to computer technology is weaker than indicated in our earlier Figure 2 (and the connection from biology to Earth sciences is stronger, presumably because as already mentioned the division between eco-environmental science and oceanography as one of the Earth sciences has some ambiguity). Finally the socio-economic sciences seem to be rather disconnected from

the other sciences in terms of the requirements in the curricula at Texas A&M University. Perhaps with future developments there will evolve connections to environmental sciences or biology, such as emphasized by Wilson[25] ("sociobiology" being a recent but somewhat controversial field focused on such connections).

A final point of evidence concerning our hierarchy of sciences may perhaps be perceived from the Table 1. The subject areas with a greater degree of requirements within their own subject area seem often to be higher up the hierarchy of fields. Thus requirements within mathematics are very high, with internal requirements in physics following, then chemistry, and biology much less so. A modest exception to this rule is that the engineering curricula have extensive requirements within engineering – though from several of the course titles (such as statics and dynamics, or thermodynamics, or continuum mechanics, or fluid mechanics) many of these courses can be viewed to manifest much of their essence as physics courses, though taught within the engineering departments. A further typical course requirement is a "materials science" or "strength of materials" course, which appears as an engineering course, though one might imagine it would be more appropriate as part of a nano-science (or materials science) department, were there one at Texas A&M University. Overall one sees objective scientometric evidence for the interconnection pattern much as we have now outlined.

## Conclusions

Evidently there are rather different ways to view the interrelations amongst science, differing much from the citation and co-citation connection patterns obtained in earlier decades. These previous scientometric citation-based connection patterns are seen sometimes to miss important historical and on-going interconnections, though they retain an air of objectivity through their numerical analyses of large quantities of data. But such patterns have rather ambiguous choices built into them as to the consideration of which journals to take data from and of how to identify different articles and citations to different fields. In particular we suspect that part of the reason that these earlier studied citation and co-citation connection patterns do not so clearly pick up the deep interconnection between mathematics and physics is that many other connections to mathematics are more properly to "applied" mathematics rather than to "fundamental" mathematics. This reason is further enhanced by the rather singular extent to which physicists are trained in mathematics, so that they frequently might do the applied mathematics themselves. That is, mathematical physics is so thoroughly now developed and populated by researchers that are specialists in this area, that there is much internal referencing, without so extensive a cross-referencing to mathematics – and the same sort of effect may occur in mathematics, with little "back-referencing" to mathematical physics. That is, mathematical physics having developed such an immense structure

may be a little difficult for many "outsiders" to penetrate to do successful research, excepting for some leading "cross-over" authorities, such as Roger Penrose.

Our hierarchical "framework" ordering of Figure 2 here has some favorable basis. One reason that such science-hierarchy diagrams may manifest a different patterning than that based on citations is that the interconnections our figure manifests, may be more a feature of initial foundations which are no longer a focus of current research, which however is clearly the focus of the citation based interconnection patterns, at least so long as they emphasize citations from fairly current literature. But in fact we surmise that many of these interconnections would not be so well manifested in citations from older literature either, as many fields grew up and even extensively developed before their framework foundation in our higher-level sciences was well manifested – e.g. this is certainly true of the relation between biology and chemistry, and also between chemistry and physics. But it is emphasized that we have provided considerable interpretative historical support of our hierarchical diagram, along with an indication of further scientiometrically based empirical support via curricular correlations as considered in the preceding section. Thence there evidently is much more to the inter-relationship amongst the various sciences than is manifested in the citation and co-citation mappings. And in the manifestation of a partial ordering our hierarchy has a neat mathematical structure, thereby indicating further information. As regards our curricular correlations to support the framework hierarchy more thorough investigation would seem to be worthwhile.

And finally we may return to our question of the "centrality" of chemistry. Brown et al.'s argument that chemistry concerns all we can touch and that it has diverse scientific contacts, is evidently further supported by its bridging character between the "hard" and "soft" sciences, and it is supported by our observation that chemistry is the primary initial location of branching in our partial ordering of the sciences. We have also noted that some arguments could ascribe centrality to biology or biochemistry or mathematics from the citation-based mappings. Indeed we have noted that an argument could be made that the math/physics "root" pair is "central" in some sense. Still we see that there are in fact a few different (even scientometric) indications of chemistry's centrality.

*

DJK acknowledges the support (through grant BD-0894) of the Welch Foundation of Houston, Texas. We thank Professor Roald Hoffmann for helpful discussions.

## References


1. T. E. BROWN, H. E. LEMAY, B. E. BURSTEN, *Chemistry – The Central Science*, Prentice Hall, 10th ed., Upper Saddle River, NJ, 2006.
2. J. W. MOORE, C. L. STANITSKY, P. C. JURS, *Chemistry – The Molecular Science*, Brooks/Cole, 2002.
3. J. VAN BRAKEL, Chemistry as the science of the transformation of substances, *Synthese,* 111 (1997) 273–282.
4. H. SMALL, Co-citation in the scientific literature: a new measure of the relationship between two documents, *J. Am. Soc. Information Sci. (JASIS)*, 24 (1973) 265–269.
5. H. SMALL, B. C. GRIFFITH, The structure of scientific literature, I: Identifying and graphing specialties, *Science Studies*, 4 (1974) 17–40.
6. B. C. GRIFFITH, H. SMALL, J. A. STONEHILL, S. DEY, The structure of scientific literature, II: Toward a macro and microstructure for science, *Science Studies*, 4 (1974) 339–365.
7. I. V. MARSHAKOVA, System of document connection based on references, *Nauchno-Tekhnicheskaya Informatsiya, Series II*, 6 (1973) 3–8.
8. INSTITUTE FOR SCIENTIFIC INFORMATION, *ISI Atlas of Science*: *Biochemistry and Molecular Biology. 1978/80* ISI, Philadelphia, 1981; *ISI Atlas of Science*: *Biotechnology and Molecular Genetics 1981/81,* ISI, Philadelphia, 1984.
9. H. SMALL, Visualizing science by citation mapping, *J. Am. Soc. Information Sci. (JASIS),* 50 (1999) 799–813.
10. E. GARFIELD, Mapping the structure of science, *Citation Indexing: Its Theory and Applications in Science, Technology, and Humanities*, Wiley, New York, pp. 98–147.
11. S. AARONSON, The footnotes of science, *Mosaic*, 6 (1975) 22–27.
12. F. MOYA-ANEGÓN, B. VARGAS-QUESADA, V. HERRERO-SOLANA, Z. CHINCHILLA-RODRIGUEZ, E. CORERA-ALVAREZ, A new technique for building maps of large scientific domains based on the cocitation of classes and categories, *Scientometrics*, 61 (2004) 129–145.
13. K. W. BOYACK, R. KLAVANS, K. BÖRNER, Mapping the backbone of science, *Scientometrics*, 64 (2005) 351–374.
14. R. KLAVANS, K. W. BOYACK, Identifying a better measure of relatedness for mapping science, *J. Am. Soc. Information Sci. Technol.,* 57 (2006) 251–263.
15. A. COMTE, *Cours de philosophie positive*, Paris, 1830.
16. P. A. M. DIRAC, Quantum mechanics of many-electron systems, *Proc. Roy. Soc. London* A123, (1929) 714–733.
17. A. KORNBERG, Understanding life as chemistry, *Clin. Chem.,* 37 (1991) 1895–1899; *Int. J. Quantum Chem.,* 53 (1995) 125–130; *Medicine, Science and Society* (K. J. ISSELBACHER, Ed.), Wiley, New York, 1982, pp. 1–17 [based on a lecture delivered at Uppsala on Dec. 8, 1991 in a Symposium on the "Unification of Natural Sciences" as a part of the Nobel Jubilee: 1901–1991].
18. A. KORNBERG, Chemistry – the lingua franca of the medical and biological sciences. *Chemistry and Biology*, 3 (1996) 3–5 [based on a speech given at the "L'Année Pasteur", Symposium on Stereospecificity and Molecular Recognition, held at the Rockefeller University, New York, Sep, 12–15, 1995].
19. A. KORNBERG, The two cultures: chemistry and biology. *Biochemistry*, 26 (1987) 6888–6891 [adapted from a lecture at a meeting of the American Association for the Advancement of Science, Feb. 16, 1987].
20. H. SIMON, *The Science of the Artificial,* MIT Press, Cambridge, MA, 1981.
21. J. NEGGERS, H. S. KIM, *Basic Posets,* World Scientific, Singapore, 1998.
22. W. T. TROTTER, *Combinatorics and Partially Ordered Sets*, Johns Hopkins University Press, Baltimore, 1992.
23. G. BIRKHOFF, *Lattice Theory,* American Mathematical Society, Providence, RI, 1948.
24. D. KNUTH, Mathematics and computer science: coping with finiteness, *Science* 194 (1976) 1234–1242.
25. E. O. WILSON, *Consilience – the Unity of Knowledge*, A. A. Knopf Publishers, NewYork, 1998.

26. G. F. R. ELLIS, Physics and the real world, *Physics Today,* 59 (July 2005) 49–54.
27. J. J. GRAY, *Janos Bolyai, Non-Euclidean Geometry, and the Nature of Space,* MIT Press, Cambridge, MA, 2004.
28. B. RIEMANN'S Habilitationsvortrag "Über die Hypothesen, welche der Geometrie zugrunde liegen", as quoted (and translated) on page 161 of *Concepts of Space* by M. Jammer, Dover, NY, 1993.
29. H. MINKOWSKI, Raum und Zeit, *Phyikalische Zeitschrift,* 10 (1909) 104–111.
30. H. WEYL, *Raum, Zeit, Materie*, Springer Verlag, Berlin, 1923.
31. Y. HARDY, W.-H. STEEB, *Classical and Quantum Computing,* Birkhauser Verlag, Basel, 2001.
32. G. JOHNSON, *A Shortcut Through Time: The Path to the Quantum Computer*, Knopf Pub., NY, 2003.
33. N. LUZIN, Function: Part I and Part II, *Am. Math. Mon.,* 105 (1998) 59–67 and 263–270.
34. V. P. VIZGIN,"On the emotional assumptions without which one could not effectively investigate the laws of science, *Am. Math. Mon.,* 108 (2003) 264–270.
35. D. HILBERT,"Naturkennen und Logik, *Naturwissenschaften* (1930) 959–963.
36. P. A. M. DIRAC, The evolution of the physicist's view of nature, *Sci. Am.,* 208 (5) (1963) 45–53.
37. R. PENROSE, *The Road to Reality*, A. A. Knopf Publishers, NewYork, 2005.
38. M. MONASTYRSKY, Mathematics at the turn of the millennium, *Am. Math. Monthly*, 112 (2005) 832–838.
39. M. DELBRÜCK, *Mind from Matter?* Blackwell Science Pub., London, 1985.
40. E. SCERRI, Has the periodic table been successfully axiomatized? *Erkenntnis,* 47 (1997) 229–243.
41. E. SCERRI, How good is the quantum mechanical explanation of the periodic table? *J. Chem. Educ.,* 75 (1998) 1384–1385.
42. C. A. COULSON, Present state of molecular structure calculations, *Rev. Mod. Phys*., 41 (1960) 170–177.
43. C. L. MORGAN, *Emergent Evolution,* Williams & Norgate, London, 1923.
44. M. H. V. VAN REGENMORTEL, Reductionism an complexity in molecular biology, *EMBO Reports* 5 (2004) 1016–1020.
45. E. NAGEL, *The Structure of Science,* Harcourt, Brace & World, NY, 1961.
46. J. C. ECCLES, K. R. POPPER, *The Self and Its Brain,* Springer International, Berlin, 1977.
47. S. JOHNSSON, *Emergence: The Connected Lives of Ants, Brains, Cities and Software*, Gardners Books, 2002.
48. J. HOLLAND, *Emergence: From Chaos to Order,* Addison-Wesley, Redwood City, CA, 1998.
49. R. B. LAUGHLIN, *A Different Universe,* Basic Books, NY, 2005.
50. J. W. SERVOS, *Physical Chemistry from Ostwald to Pauling*, Princeton University Press, Princeton, 1990.
51. M. J. NYE, *From Chemical Philosophy to Theoretical Chemistry*, Univ. of California Press, Berkeley, 1993.
52. L. P. HAMMETT, *Physical Organic Chemistry,* McGraw-Hill, New York, 1940.
53. D. J. KLEIN, Mathematical Chemistry! Is It? And if so, what is it? *Foundations of Chemistry*, (submitted 2006).
54. A. T. BALABAN, Reflections about mathematical chemistry, *Foundations of Chemistry*, 7 (2005) 289–306.
55. TEXAS A&M UNIVERSITY UNDERGRADUATE CATALOG (2003–2004).

EMERGENCE, 2(2), 24–57


# *Dynamics in Action: Intentional Behavior as a Complex System*

*Alicia Juarrero*

This article is a truncation and summary edited by Alicia Juarrero and Michael Lissack, adapted from Juarrero's book of the same title (MIT Press, 1999). *Note: As it is a truncation, you may wish to read this article twice.*

## *The Problem*

When dealing with hierarchical systems that are self-referential and display inter-level effects, the notion of causality must be reconceptualized in terms other than that of the billiard ball, collision conception that is the legacy of mechanism. Understanding all cause as collision like, and the explanatory ideal as deduction from deterministic laws, are part of a trend that has characterized the history of philosophy for over 2,000 years: the progressive elimination of time and context from metaphysics and epistemology. Aristotle had insisted that formal deduction from universal premises is the logic of reasoning proper to science. Noting that human behavior is temporally and contextually embedded, Aristotle made it the central concern of practical wisdom. Unlike deduction, wisdom varies "as the occasion requires." Modern philosophy, by contrast, insists that (ideally) all explanation is fundamentally deductive in form.

In Plato's dialogue *Phaedo*, which takes place while Socrates is awaiting execution, Socrates worries that earlier philosophers made air, ether,

and water the only causes. What about Socrates' reasons for not escaping from prison? Are they not the true cause of his behavior? Later, and more systematically, Aristotle examined the difference between intentional and involuntary behavior. An adequate explanation of anything, he claimed, must identify those causes responsible for the phenomenon being explained. Aristotle's four causes are final cause (the goal or purpose toward which something aims), formal cause (that which makes anything that sort of thing and no other), material cause (the stuff out of which it is made), and efficient cause (the force that brings the thing into being). Explaining anything, including behavior, requires identifying the role that each cause plays in bringing about the phenomenon. Implicit in Aristotle's account of cause and crucially influential, however, is another of his claims: that nothing, strictly speaking, can move, cause, or act on itself in the same respect.

Contemporary causal theories of action have consistently adhered to Aristotle's principle that nothing moves or changes itself; intentions, volitions, and other alleged causes of action are supposed to be other than the behavior they cause. In addition, by also subscribing to a Newtonian understanding of efficient cause, these theories have also uncritically assumed that intentions, volitions, or agents cause action in the collision-like way that pool cues cause cue balls to move. Following Hume and in opposition to Aristotle, philosophers concluded that deduction from timeless and context-less laws is the ideal, not only of science but of any legitimate form of reasoning. A law of nature—at worst statistical, but ideally strictly deterministic—combined with statements specifying initial conditions must allow that which is being explained (the explanandum) to be inferred. Even human actions must be explained in that manner. By the middle of the twentieth century, the principle that the logic of any serious explanation must adhere to such a "covering-law model" was the received view.

Modern philosophy's understanding of cause and explanation has failed as a general theory. Today, there is no reason to continue to subscribe to this atemporal and acontextual approach. The conceptual framework of the theory of complex adaptive systems can serve as a "theory-constitutive metaphor" that permits a reconceptualization of cause, and in consequence a rethinking of action. A different logic of explanation—one more suitable to all historical, contextually embedded processes, including action—arises from of this radical revision.

## *Complexity As a New Answer*

Several key concepts of the new scientific framework are especially suited to this task. First, complex adaptive systems are typically characterized by positive feedback processes in which the product of the process is necessary for the process itself. Contrary to Aristotle, this circular type of causality is a form of self-cause. Second, when parts interact to produce wholes, and the resulting distributed wholes in turn affect the behavior of their parts, inter-level causality is at work. Interactions among certain dynamical processes can create a systems-level organization with new properties that are not the simple sum of the components that constitute the higher level. In turn, the overall dynamics of the emergent distributed system not only determine which parts will be allowed into the system: the global dynamics also regulate and constrain the behavior of the lower-level components.

The theory of complex adaptive systems can therefore be used as a metaphor for this form of causal relations, which had puzzled Kant as a form of causality "unknown to us." In other words, far from being the inert epiphenomenon that modern science claims all wholes are, complex dynamical wholes clearly—and in a distributed manner—exert active power on their parts such that the overall system is maintained and enhanced. Understanding dynamical systems can therefore revive Aristotle's concepts of formal and final cause by offering a scientifically respectable model of how such causes operate. Since the active power that wholes exert on their components is clearly not the gocart-like collisions of a mechanical universe, the causal mechanism at work between levels of hierarchical organization can better be understood as the operations of constraint.

We can distinguish between two types of constraints: context-free constraints, which take a system's components far from equiprobability, and context-sensitive constraints, which synchronize and correlate previously independent parts into a systemic whole. When organized into a complex, integral whole, parts become correlated as a function of context-dependent constraints imposed on them by the newly organized system in which they are now embedded. Catalysts, feedback loops, and biological resonance and entrainment embody context-sensitive constraints.

From the bottom up, the establishment of context-sensitive constraints is the phase change that self-organizes the global level. Or, to put it differently, the self-organization of the global level is the appearance of context-sensitive constraints on the system's components. Parts hereto-

fore separate and independent are suddenly correlated, thereby becoming interdependent components or nodes of a system. But even as they regulate alternatives, context-dependent constraints simultaneously open up new possibilities. The more complex a system, the more states and properties it can manifest: novel characteristics and laws emerge with the organization of the higher level. Constraints work, then, by modifying either a system's phase space or the probability distribution of events and movements within that space, Since actions are lower, motor-level implementations of higher-level intentional causes, reconceptualizing mental causation in terms of top-down, context-sensitive dynamical constraints can radically recast our thinking about action.

Since the global level of all complex adaptive systems contextually constrains the behavior of the components that make it up, I postulate that behavior constitutes action when self-organized dynamics, as characterized by consciousness and meaning, originate, regulate, and constrain processes such that the resulting behavior "satisfies the meaningful content" embodied in the complex dynamics from which it issued. By serving as the order parameter, those contextual constraints that embody an intention (acting top down) would provide the behavior with continuous, on-going control and direction by modifying in real time the probability distribution of lower-level processes and, as a consequence, the behavioral alternatives available to and implemented by the agent.

Far from representing messy, noisy complications that can be safely ignored, time and context are as central to the identity and behavior of these dynamic processes as they are to human beings. Unlike the processes described by classical thermodynamics, which in their relentless march toward equilibrium forget their past, complex adaptive systems are essentially historical. They embody in their very structure the conditions under which they were created (including the chance events around which each self-organized stage reorganizes). The unrepeatable, random fluctuation or perturbation around which each phase of a sequence of adaptations nucleates leaves its mark on the specific configuration that emerges. The structure of a snowflake, for example, carries information about the conditions under which it was created. Each level is uniquely and progressively individuated, as is the developmental and behavioral trajectory of each organism (Salthe, 1993a).

## *The Role of Narrative*

Far enough from equilibrium, dynamical systems can abruptly and irreversibly undergo a radical transformation. On the other side of this "bifurcation," a system either reorganizes into a higher level of complexity characterized by renewed potential and possibilities, or falls apart. Across phase changes, that is, there are no established dynamics that can serve as the context from which the parts derive their meaning; the change itself in the dynamics governing the system's stable states needs to be explained. How, then, must human action be explained?

I propose an interpretive, narrative model of explanation. In hermeneutical interpretation, the meaning of a complete text is constructed from the relationships among the individual passages. In turn, the meaning of the story's individual passages is derived from the meaning of the entire text in which those passages are embedded. This continual, interpretive "tacking" from parts to whole and back to parts reproduces the way dynamical systems self-organize out of the interrelationships among the parts, and then loop back to constrain those parts. The similarity in the dynamics of self-organization and hermeneutics makes the latter uniquely suited as the logic of explanation for stable states of the former. Since these phase changes are unpredictable, the only way to explain them is with a retrospective narrative that retraces the actual leap. Explaining these individual dramatic transformations, as well as the detailed trajectory that even everyday behavior takes, requires a genealogical narrative that makes ample references to temporal and contextual events. This historical interpretation must provide detailed descriptions of the singular incidents that the agent experienced and that both precipitated the transformation and served as the nucleus around which the bifurcation reorganized.

Narrative, interpretive, and historical explanations of action thus require an expanded appreciation of what counts as "reason" and "explanation," for they explain, not by subsuming an explanation under a generalization and thereby predicting it, as modern philosophy would require, but rather by providing insight into and understanding of what actually happened. They do so by supplying a rich description of the precise, detailed path that the agent took, including the temporal and spatial dynamics (both physical and cultural) in which the agent was embedded and in which the action occurred. Who could have predicted that Ibsen's Nora (*A Doll's House*) would leave her husband and child? And yet, at the end of the play, we understand why Nora slams the door, even if no one

could have predicted it. We understand her behavior by coming to appreciate all the complicated and messy factors that became entangled in her life; the drama shows how they interacted to produce a break. Moreover, if we learn anything from watching the play, we also learn something about the quirks and idiosyncrasies of human psychology, of the circumstances in which humans function, and of how these contribute to the unpredictability of our actions. Historical, interpretive stories might not allow us to predict future behavior, but they do allow us to understand why it is unpredictable.

## *Self-Cause*

In a universe where only point masses and forces are considered real, qualities that are a function of the relation between atoms, or between organisms and the world, were also dismissed as subjective. By the end of the seventeenth century, all relational properties, such as temperature and color, that did not to fit into this scheme were relegated to the inferior status of "secondary" qualities. Galileo's ability to set aside the interference of friction from the equations governing the motion of bodies also suggested that context contributes nothing to reality. Once atomism became the ruling conceptual framework, context and environment were thus left without a role to play in either science or philosophy. Indexicals, such as here and there, this and that, now and then, lost their claim on reality as situatedness and point of view became unimportant. An object's only real properties were its so-called primary properties, characteristics such as mass that, because they are internal to the object, it would exhibit anywhere, anytime. Since only primary properties were essential properties, it was therefore no longer acceptable to explain action as Aristotle had, by embedding the organism in its environment. If real things (atomic particles), were unaffected by time and context, an object's interactions with its environment and the unique trajectory it traced through time and space also became secondary, "accidental" properties of no account to what really makes a thing the kind of thing it is and no other ("anthropological considerations," Kant would have called them).

True self-cause would involve localized parts interacting so as to produce wholes that in turn, as distributed wholes, could influence their components: inter-level causality between parts and wholes. But by following Aristotle in rejecting this possibility, philosophy closed off any avenue for explaining action in that fashion. Having discarded the notions of formal and final cause, moreover, philosophy was left without a way of

understanding nonevent causation. As a result, philosophers who champion agent causation as the distinguishing mark of action have never satisfactorily explained either the identity of agents or their manner of causation.

Danto (1979b: 16) proposes that in the same way that "knowledge ... is a matter of bringing our representations into line with the world ... action is a matter of bringing the world into line with our representations." Guiding a plan of action to completion, however, as implied in the gerundive "bringing the world into line," requires the operation of a cause that, far from disengaging at the onset of the action, persists throughout the performance of the act and monitors and directs the behavior. Inasmuch as that type of cause would identify the action's origin, the label "cause" would still apply (Sosa, 1980). But it would not be a Newtonian forceful push. The intentional content according to which the agent shapes the world is not related to its behavioral effect in the way a Newtonian cause is related to its effect. A proximate intention is an action's origin or source, but it is not a discrete event that precedes the action and yet is not itself part of the action. What is significant for purposes of action is that even someone wanting to defend a causal theory finds something not altogether satisfactory about the traditional view of cause, in which the relation between cause and effect is one of "external coincidence." Searle was right: the intention must be in the action. So we need an account of the metaphysics of that type of cause. Still, the reluctance to countenance any kind of self-cause persists.

Following information theory, it is useful to think of action as informationally dependent and constrained behavior. Those intentions and other cognitive phenomena responsible for action must be robust enough to withstand the mischief of noise and equivocation and to flow into behavior. We can now appreciate, however, that behavior constitutes action if and only if it flows unequivocally from a cognitive structure in virtue of its meaningful content (Audi, 1995; Kim, 1995). And "it must be this content that defines the structure's causal influence on output" (Dretske, 1981, 199).

When the environment that is part of the system's external structure is also taken into account, at least three levels are simultaneously involved: the focal level, the environmental level immediately above, and the component level immediately below (Salthe, 1985)—with feedback paths among them. In control hierarchies with this sort of leakage between levels, a clean dynamic theory referring to one level at a time cannot be formulated. Simon calls systems like these nearly decompos-

able. This apparent design flaw can have remarkable consequences: inter-level leakage makes the system robust to noise, context sensitive, and, in the case of artificial neural networks, able to generalize. Multilevel, dynamic coupling of components, both at the same level and between levels (such as one finds in the cells, tissues, organs, and so forth of biological organisms), "maintains a certain autonomy at all hierarchical levels" (Jantsch, 1980: 247). Components at different levels are not subsumed or fused into the highest level, but they do interact. Given sufficient environmentally imposed structure, they can be labeled as Prigogine calls them, "dissipative."

## DISSIPATIVE STRUCTURES

Once in place, the dynamics of a dissipative structure as a whole "provide the framework for the behavioral characteristics and activities of the parts" (Zeleny, 1980: 20). By delimiting the parts' initial repertoire of behavior, the structured whole in which the elements are suddenly embedded also redefines them. They are now something they were not before, nodes in a network, components of a system. As such, they are unable to access states that might have been available to them as independent entities. Insect colonies are an example of this phenomenon, self-organizing systems whose complexity "permits the division of functions, particularly the division of labor, as well as hierarchical relationships and mechanisms of population control" (Janisch, 1980: 69). The evolutionary advantage of such systematic hierarchical differentiation is that the whole can access states that the independent parts cannot. The overall hive can do much more than the individual bee. The price is that workers in a hive lose the ability to reproduce.

In short, not only individual but interacting parts suddenly correlate to create systematic wholes; once organized, the resulting systems affect their components. In other words, self-organizing systems exhibit that previously unknown inter-level causality, both bottom up and top down. They display bottom-up causality in that, under far-from-equilibrium conditions, their internal dynamics amplify naturally occurring fluctuations around which a phase change nucleates. When this discontinuous and irreversible transition occurs, a qualitatively different regime self-organizes. A new "type" of entity appears, one that is functionally differentiated. In turn, the newly organized hierarchy constrains its components' behavior top down by restructuring and relating them in ways in which they were not related before. Dissipative structures thus

operate on two levels simultaneously: part and whole, which interact in the manner of Douglas Hofstadter's (1979) "strange loops," or Kant's "unknown causality." In Chuck Dyke's (1988) great phrase, they are "structured structuring structures."

## *Objections Considered*

An objection commonly raised against systems theory is worth mentioning at this point. Because of claims such as Dewan's (1976) to the effect that entrainment is an example of an emergent, holistic property of control that has causal potency, Bunge (1979) charges holism with the false claim that wholes act on their parts. Wholes cannot act on their parts, he maintains, since a level of organization "is not a thing but a set and therefore a concept ... levels cannot act on one another. In particular the higher levels cannot command or even obey the lower ones. All talk of interlevel action is elliptical or metaphorical" (13–14). Since there is, on Bunge's account, no ontological (only an epistemological) relationship between levels of organization, there can be no actual control by one over another.

Complex adaptive systems have proven Bunge wrong; their interlevel relationships, however tangled, are real, not just epistemological. The emergence of relatively autonomous levels of organization carries with it the emergence of relatively autonomous qualities; quantitative changes produce qualitative changes (Bohm, 1971). Once a transition point is passed, new modes of being emerge, in particular new modes of causality. "The most essential and characteristic feature of a qualitative transformation is that new kinds of causal factors begin to be significant in a given context, or to 'take control' of a certain domain of phenomena, with the result that there appear new laws and even new kinds of laws, which apply in the domain in question" (53). Aversion to the possibility that wholes might act on their parts betrays both the continuing and uncritical acceptance of philosophy's refusal to countenance self-cause as well as the prevalent philosophical tendency toward reification: an ontological bias that favors concrete things over processes and relations, substances over properties. It is true, of course, that wholes do not act on their components forcefully; but neither are wholes other than or external to the components that make them up. And to claim that they do not causally affect their components at all begs the question by assuming that all cause must be billiard ball like to be causally efficacious at all.

Contradicting Bunge, Zeleny (1980: 20) suggests that the lesson to be learned from the theory of autopoiesis is precisely "the lesson of holism."

Far from being an inert epiphenomenon, the dynamics of the autopoetic whole serve as the orderly context that structures the behavioral characteristics and activities of the parts, a clear formulation of one of Bunge's (1979: 39) characteristics of a holistic point of view: the dynamics of the global level control the functioning of components at the lower level. The whole as whole most assuredly acts on its parts: self-cause—but not, as some would have it, *qua* other—one part forcefully impressing itself on another. Instead, complex adaptive systems exhibit true self-cause: parts interact to produce novel, emergent wholes; in turn, these distributed wholes as wholes regulate and constrain the parts that make them up.

Bunge (1979) also explicitly rejects the concept of hierarchy because, he notes, "hierarchy" implies a "dominance relation," always by the higher level on the lower one. It is true that the word "hierarchy" implies a unidirectional flow of order or authority, always and only from higher to lower (see Dyke, 1988). To counteract this connotation, students of complex dynamical systems have coined the neologism "heterarchy" to allow inter-level causal relations to flow in both directions, part to whole (bottom up) and whole to part (top down).

## *Back to Constraints*

The orderly context in which the components are unified and embedded constrains them. Constraints are therefore relational properties that parts acquire in virtue of being unified—not just aggregated—into a systematic whole. For example, the physical link between the tibia and both the peronei and the knee joint systematically constrains the movement of the lower leg. As a result of the connection, the tibia's physiology is not independent of the knee, the linkage creates an orthopedic system that controls the tibia in ways to which it would not otherwise have been limited. The anatomical tie restricts the lower leg's range of motion. The constraints that the tibia's relationship to the knee places on the tibia limit the number of ways in which the lower leg can move: it can bend backward but not forward, for example. In this example, a constraint represents a contraction of the lower leg's potential range of behavior: the lower leg has less freedom of movement, given its connection with the knee, than it would have otherwise.

Limiting or closing off alternatives is the most common understanding of the term "constraint." But if all constraints restricted a thing's degrees of freedom in this way, organisms (whether phylogenetically or developmentally) would progressively do less and less. However,

precisely the opposite is empirically observed. Some constraints must therefore not only reduce the number of alternatives: they must simultaneously create new possibilities. We need to understand how constraints can simultaneously open up as well as close off options (Campbell, 1982). To do so, it is helpful to examine another usage of the concept of constraint. Let us turn, therefore, to information theory, in which constraints are identified not as in physical mechanics, with physical connections, but with rules for reducing randomness in order to minimize noise and equivocation.

## INFORMATION-THEORETICAL CONSTRAINTS

Lila Gatlin quotes Weaver (of Shannon & Weaver fame) as saying that "this word 'information' in communications theory relates not so much to what you do say, as to what you could say" (quoted in Gatlin, 1972: 48). In a situation of complete randomness where alternatives are equiprobable, you could say anything but in fact say nothing. Random, equiprobable signals are static, thus unable to transmit actual messages. It is true that in situations in which all alternatives are equally likely, potential information or message variety is at its maximum: before the process of selection in the Herman example, any one of the employees could be chosen. Likewise, the equiprobability of static crackle equates with unpredictability and maximum freedom; in short, with the possibility of constant novelty. But a series of totally random or equiprobable signals is meaningless: no pattern or message is extractable from the disorder. There is none.

At equilibrium, message variety is therefore a great but idle potential; actual information is zero. "Capacity is of no value if it cannot be utilized" (Gatlin, 1972: 99). Without contrasts there can be no message; television snow is as meaningless as white noise. Transmitting or receiving a message requires a clear demarcation between message and background noise. The transmitter as well as the receiver must reduce the randomness in the sequence of signals to a "manageable" level. Encoding (and deciphering) the message according to certain rules is one way of doing so. Whether in communications or genetics, therefore, actual information content—a difference that makes a difference—requires an ordering process that harnesses the randomness. Constraining "the number of ways in which the various parts of a system can be arranged" (Campbell 1982: 44) reduces randomness by altering the equiprobable distribution of signals, thereby enabling potential information to become actual infor-

mation. Constraints thus turn the amorphous potential into the definite actual; following Aristotle, constraints effect change—and inform. Constraints embodied in encryption rules also take the signals away from equiprobability and randomness.

The "most random state is ... characterized by events which are both independent and equiprobable" (Gatlin, 1972: 87). When anything is as possible as anything else, and nothing is connected to anything else, however, nothing can signify or communicate anything. Flashes from a lighthouse pulsing regularly three long, three short, three long, on the other hand, can carry information precisely because regular flashes are more improbable than random ones, and can therefore be differentiated from background noise. Even to an extraterrestrial, the improbability of regularly pulsing flashes of light says "signal," "signal," "signal," even if ET cannot tell what it means.

The same is true of language: if all sounds were equiprobable and every letter of the alphabet were as likely to show up as any other, no message could be communicated. Hence in any language, some letters appear more often than others. A number on a fair die has the same likelihood of being thrown as any other. The probability that a particular letter of the alphabet will appear in a word or sentence, however, is not like that. Some letters are more likely than others: in the long run they repeat more often (with increased redundancy) than they would in a random distribution. Each become possible; additional contextual constraints (on top of the contextual constraints that create words) make sentences possible. Systems, systems of systems, and so on can be assembled. By making the appearance of letters in an alphabet interdependent, contextual constraints thus allow complex linguistic structures to emerge.

As is the case in all complex systems, newly synchronized components pay a price for creating a global system: the number of ways in which they can be individually arranged is correspondingly reduced. In English, once "*-t-i-o*" appear toward the end of a noun, the probability of *a*'s appearing next decreases dramatically. But the payoff trumps the cost: the interdependence that context-sensitive constraints impose offers the advantage of permitting unlimited message variety despite limited channel capacity. A contextually coded alphabet yields more i-tuplets than its 26 letters; there are more words than i-tuplets, more sentences than words. To achieve the requisite variety, and because Mandarin Chinese limits words to one or two syllables, for example, the context-sensitive constraints of inflection are sometimes needed. Phonetic, syntactic, and stylistic layers of context-sensitive constraints, added on top of the

context-free constraints on the prior probability of individual letters, thus provide a significant advantage over ideograms, pictograms, or hieroglyphs.

Without contextual constraints on sounds and scribbles, communication would be limited to a few grunts, shouts, waits, and so forth that would be severely restricted in what and how much they could express. Language's increased capacity to express ideas rests not on newly invented grunts and shouts, but on the relationships and interconnections established by making interdependent the sounds in a sequence of grunts or shouts; that is, by making the probability of their occurrence context dependent. Context-sensitive constraints are thus as efficient but not as expensive as context-free ones, for they "can be increased by a reasonable amount without cramping the message source too severely" (Campbell, 1982: 9). By correlating and coordinating previously aggregated parts into a more complex, differentiated, systematic whole, contextual constraints enlarge the variety of states that the system as a whole can access.

## EXAMPLES FROM NATURE

All of this would, of course, be of minimal interest to action theorists or philosophers of mind if it were a mechanism found only in language. That this is emphatically not so is one of the lessons to be learned from complex dynamical systems. I have used examples from language merely as a heuristic illustration of the process. The emergence of Bernard Cells and B–Z chemical waves signals the abrupt appearance of context-sensitive constraints in mutualist-driven, open processes far from equilibrium. This discontinuous change occurs when previously unrelated molecules suddenly become correlated in a distributed whole. A complex dynamical system emerges when the behavior of each molecule suddenly depends both on what the neighboring molecules are doing and on what went before. When components, in other words, suddenly become context dependent.

The same is true of auto catalysis. As physical embodiments of context-dependent constraints, catalysts are therefore one way in which natural processes become subject to conditional probabilities. Because of their geometry, catalysts can take molecules away from independence, not just equiprobability, the way context-free constraints do, by enhancing the likelihood that certain other events will occur. Once the probability that something will happen depends on and is altered by the presence of something else, the two have become systematically and therefore

internally related. As a result of the operations of context-sensitive constraints and the conditional probabilities they impose, *A* is now part of *B*'s external structure. Because *A* is no longer "out there" independent of *B*, to which it is only externally related, the interdependence has created a larger whole, the *AB* system. Insofar as it is part of *B*'s new context or external structure, *A* has been imported into *B*. By making a system's current states and behavior systematically dependent on its history, the feedback loops of auto catalysis also incorporate the effects of time into those very states and behavior patterns. Indeed, precisely what makes these complex systems dynamical is that a current state is in part dependent on a prior one. Feedback, that is, incorporates the past into the system's present "external" structure. Feedback thus threads a system through both time and space, thereby allowing part of the system's external structure to run through its history.

Feedback processes thus embody the context-sensitive constraints of history. By embodying context-sensitive constraints, mutualist feedback renders a system sensitive to (constrained by) its own past experiences. This makes nonlinear dynamical systems historical, not just temporal the way near-equilibrium thermodynamical systems are. Once the system's subsequent behavior depends on both the spatial and temporal conditions under which it was created and the contingent experiences it has undergone, the system is historically and contextually embedded in a way that near-equilibrium systems of traditional thermodynamics are not. Because dissipative structures are not just dropped into either time or space the way Newtonian atoms with only primary qualities are, their evolutionary trajectory is therefore not predictable. Mutualism thus makes a dynamical system's current and future properties, states, and behaviors dependent on the context in which the system is currently embedded as well as on its prior experiences. As a result, unlike the near-equilibrium processes of traditional thermodynamics, complex systems do not forget their initial conditions: they "carry their history on their backs" (Prigogine, Spring 1995, US Naval Academy). Their origin constrains their trajectory.

Operating as enabling constraints (Salthe, 1993b), context-sensitive constraints make complexity possible. The emergence of auto catalytic cycles and slime molds—of self-organized systems in general—is the phenomenological manifestation of the sudden closure of context-sensitive constraints. As mentioned earlier, the new relationship among the components is the establishment of a new context—a new external structure or boundary conditions—for those components. Once particles

and processes are interrelated into a dissipative structure, they become components or nodes of a more highly differentiated whole. By correlating previously independent particles and processes, context-sensitive constraints are therefore one mechanism whereby chemical and biological hierarchies are created.

By taking the organism far from equilibrium and precipitating a bifurcation, the persistent interaction of conditioning establishes context-sensitive interdependencies between the organism and its environment. Parts interact to produce a greater organism–environment whole, which in turn affects (top down) those very parts. Conditioning and learning import the environment into the agent's dynamics by reorganizing and recalibrating those dynamics. In this sense, components are embedded in and not just dropped into an environment, as in an experiment. Once self-organized, the global dynamics of the overall organism–environment system become the control knob of its components—top-down causality, in effect.

The difference in the way individual slime mold amoebas behave while they are independent entities and after they self-organize into the complex slug is not explicable solely as the result of bumping into another amoeba (as mechanics and modern philosophy would have it). The difference is largely a measure of second-order context-sensitive constraints embodied by (in) the whole self-organized slug. So, too, the difference in the way molecules of water behave while they are isolated and independent and after they self-organize into the Bernard cell is a measure of the second-order, context-sensitive constraints embodied by (in) the hexagonal cell. That difference is also a measure of each system's complexity or degree of organization (Brooks & Wiley, 1988). Top-down constraints that begin to weaken cause a system to become unstable. When this happens, the conditional probability that a component will behave in a certain way given the systematic context in which it is embedded begins to alter, and the behavior of the components fluctuates much more widely. The overall system's integrity (identity) and survival are in danger.

## *CONTEXTUAL CONSTRAINTS*

Contextual constraints thus perform double duty. From the combined effects of context-free and first-order contextual constraints, dynamical structures and patterns at a higher level of complexity self-organize. Parts interact to produce wholes, When context-free and first-order contextual constraints correlate flows of matter (reactants) and energy and thereby

take them far from equilibrium and independence, a dynamic dissipative structure of process suddenly emerges. This discontinuous transition to entrainment and hierarchical organization is the sudden establishment of second-order context-sensitive constraints: abruptly, the behavior of an individual cardiac cell, generator, water molecule, or letter of the alphabet is no longer independent of those around it. The renewed repertoire of behavioral alternatives and properties that suddenly becomes available to the emergent system as a whole is the phenomenological counterpart of the sudden appearance of second-order contextual constraints. By coordinating previously independent parts, context-dependent constraints allow a more complex organization to emerge, with novel properties that the isolated parts lacked. Self-organization enlarges a system's phase space by adding degrees of freedom. Enabling constraints thus create potential information by opening—bottom-up—a renewed pool of alternatives that the emergent macrostructure can access.

The explosion of potential message variety available to each new global level is its expanded potential. The coherent laser beam can cauterize flesh; the waves of the individual laser atoms, separately, however, cannot. The emergent level is thus qualitatively different from the earlier one. As an integrated organism, the slime mold has properties the independent amoebas that make it up did not. Tissues (which are organized webs of cells) can do different things than independent cells, organs different things than tissues, proteins different things than amino acids, Bernard cells different things than independent water molecules—all because of homologous dynamics at work. Gatlin (1972) argues that the explosion of phenotypes that took place with the appearance of the vertebrates occurred because vertebrates managed to maintain context-free redundancy constant while allowing context-sensitive constraints to expand.

As a distributed whole, a self-organized structure imposes second-order contextual constraints on its components, thereby restricting their degrees of freedom. As we saw, once top-down, second-order contextual constraints are in place, energy and matter exchanged across an auto catalytic structure's boundaries cannot flow any which way. The auto catalytic web's dynamical organization does not allow any molecule to be imported into the system: in a very important feature of self-organizing dynamical systems, their organization itself determines the stimuli to which they will respond. By making its components interdependent, thereby constraining their behavioral variability, the system preserves and enhances its cohesion and integrity, its organization and identity. As

a whole, it also prunes inefficient components. Second-order contextual constraints are thus in the service of the whole. They are, also therefore, the ongoing, structuring mechanism whereby Aristotle's formal and final causes are implemented (Ulanowicz, 1997).

Organization limits the degrees of freedom of a system's components. Once auto catalytic closure takes place, molecules become components in a system. As such, their behavioral repertoire is selectively constrained (their degrees of freedom curtailed) by the systematic context of which they are now a part. Unlike electrical generators and other allopoielic devices that require externally imposed governors, however, both auto catalytic closure and biological entrainment signal the spontaneous emergence of a field or dynamic network that is the endogenous establishment of second-order, context-sensitive constraints on the components at the first level. As distributed wholes, complex adaptive systems are virtual governors that give orders to themselves—*qua* thing, not *qua* other. The coherent laser beam "slaves" its component atomic waves even though "there is nobody to give orders" (Haken, 1987: 420). That is, one particle does not push another around. The orderly relationships that characterize the structure of the overall laser beam as a whole are the context that "gives orders" to its components, The same can be said of individual cardiac cells: the systematic context of the overall heart confers an otherwise absent stability on individual cardiac cells.

Top-down constraints that wholes exert on their components are inhibiting, selectionist constraints. Components that satisfy the requirements of the higher level will be classified as well-fitting. The constraints that wholes impose on their parts are restrictive insofar as they reduce the number of ways in which the parts can be arranged, and conservative in the sense that they are in the service of the whole. But they are also creative in a different, functional sense: those previously independent parts are now components of a larger system and as such have acquired new functional roles. The newly created overall system, too, has greater potential than the independent, uncorrelated components.

Paradoxically and simultaneously, self-organization also constitutes the appearance of the remarkable and unpredictable properties of the global level: the cauterizing ability of the laser beam, the enzymatic capabilities of a protein—or, I speculate, consciousness and self-consciousness—and their attendant states. These emergent properties of the higher level are the phenomenological manifestation of those dynamic relationships. But I emphasize that they are emphatically not epiphenomenal. Although not in a push–pull, forceful manner, the higher

level of organization is causally effective: as a second-order, top-down constraint.

On the other hand, bottom-up, enabling contextual constraints simultaneously renew message variety by enlarging the overall system's state space. The renewed possibilities of the expanded phase space available to the emergent level of organization more than offset (see Alvarez de Lorenzana, 1993) the local order that top-down contextual constraints effect by limiting alternatives at the component level. It is important to emphasize that the potential behavioral repertoire that the context-sensitive ordering process creates is at a dynamical level of organization different from that on which the selective constraints operate. The higher level of organization—whether thermodynamic, chemical, biological, psychological, or social—possesses a qualitatively different repertoire of states and behavior than the earlier level, as well as greater degrees of freedom. The global level, which in one sense is nothing more than the combined enabling constraints correlating components at the lower level, is at the same time the locus of emergent properties. You can write a book; the blastula from which you developed could not. Increased variety is one way in which greater complexity is identified. Not only can you or I write a book, we can do so carefully, sloppily, easily, and so forth. As the number of options open to the overall system increases, the potential for disorder is simultaneously renewed.

## *Identity and Attractors*

A system's identity is captured in the signature probability distribution of its dynamics. A useful way of visualizing this is as ontogenetic landscapes depicting a "series of changes of relative stability and instability" over time (Thelen & Smith, 1994: 122). If a system accessed every point or region in its phase space with the same frequency as every other (that is, randomly), its ontogenetic landscape would be smooth and flat. A completely flat, smooth initial landscape would portray an object with no propensities or dispositions; that is, with no attractors. It would describe a "system" with no identity, a logical impossibility. (On a graph, such a "system" would look like TV snow.) In contrast, the increased probability that a real system will occupy a particular state can be represented as wells (dips or valleys in the landscape) that embody attractor states and behaviors that the system is more likely to occupy. The deeper the valley, the greater the propensity of its being visited and the stronger the entrainment that its attractor represents.

Topologically, ridges separating basins of attraction are called separafrices or repellers. Sharp peaks are some points representing states and behaviors from which the system shies away and in all likelihood will not access; the probability of their occurrence is lowered or eliminated altogether. These landscape features capture the impact of context-sensitive constraints over time. Separatrix height represents the unlikelihood that the system will switch to another attractor given its history, current dynamics, and the environment. The steeper the separatrix's walls, the greater the improbability of the system's making the transition. On the other hand, the deeper the valley, the stronger the attractor's pull, and so the more entrenched the behavior described by that attractor and the stronger the perturbation needed to dislodge the system from that propensity. The broader the floor of a basin of attraction, the greater the variability in states and behaviors that the attractor allows under its control. The narrower the valley, the more specific the attractor; that is, the fewer the states and behaviors within its basin.

Since a system's external structure can recalibrate its internal dynamics, probability landscapes also incorporate the role of the environment in which a system is embedded. Since a system's prior experience constrains its behavior, that history, too, is embodied in its ontogenetic landscape. Ontogenetic landscapes, therefore, are constantly modified, dynamical portraits of the interactions between a system and its environment over time: they capture, in short, a time-lapse portrait of individual systems. Although complex dynamical systems theory is science, *pace* Aristotle, it can account for the particularity and concreteness of individual cases.

Furthermore, attractors and separatrices of complex systems are neither static givens in the manner of an Aristotelian telos, nor external control mechanisms (as was the temperature cranked up from the outside in the Bernard cell example). Nor are they determinants operating as Newtonian forces. Representing constrained pathways within self-organized space, attractors embody the system's current control parameters (its self-organized knobs), which have been constructed and continue to be modified as a result of the persistent interactions between the dynamical system and its environment. The probability that a system will do *x* next depends on its present location in the current overall landscape, which in turn is a function both of its own past and of the environment in which it is embedded. Attractors thus embody the second-order context-sensitive constraints of the system's virtual governor.

More precisely, attractor basin landscapes describe the effects of those second-order context-sensitive constraints that give a system its

particular structure and identity. They identify regions of equilibrium in a system's dynamical organization. As such, a system's dynamical portrait maps the contextual constraints that its attractors and organization embody. The difference between random behavior on the one hand, and the actually observed behavior on the other, provides evidence that an attractor is constraining the latter. Once again, this difference also measures the system's organization (Brooks & Wiley, 1988), and confers on it its identity.

## *Information Theory*

Approaching the problem of action through the lens of information theory allows us a new way of conceptualizing how intentional meaning flows into action. I previously suggested that behavior constitutes an act-token if and only if it is a trajectory that is dependent on a reduction of possibilities at an intentional source. For behavior to qualify as action, the information generated must then be transmitted uninterruptedly into behavior. The technical concepts of noise and equivocation gave us a way of measuring that dependence of outcome on origin; as such, they also gave us a way of understanding how information can flow without interruption from source to terminus, which Newtonian causality could not. The problems and objections of wayward chains and act individuation, which earlier theories of action had repeatedly encountered, could thereby be circumvented or resolved. Information theory, however, was unable (a) to account for the set of alternatives from which the selection is made, (b) to provide a plausible account of the method people use to settle on a determinate course of action, or (c) to handle meaning.

Complex dynamical systems theory is able to assist in all these tasks. The key to self-organization is the appearance of second-order context-sensitive constraints as a result of the closure of positive feedback. Second-order contextual constraints are sudden changes in the conditional probability distribution of component behavior. By partitioning a system's state space into an ordered subvolume, dynamical self-organization is therefore also analogous to information theory's "reduction of possibilities" at the source. In the case of dynamical systems, the range of alternatives available to a complex structure at any given moment is given by its organization's coordinates and dynamics—its order parameter. Evidence from artificial neural networks also suggests that the very organization of those dynamics can embody a robust sense of semantics. Finally, acting as a system's control parameter, attractors of

self-organized dynamics can serve as a causal—but not efficiently causal—mechanism.

## *INTENTIONS*

Prior intentions and plans of action, Bratman (1987) argues, channel future deliberation by narrowing the scope of alternatives to be subsequently considered. Reparsing cognitive space in this way helps us act. In the language of dynamical systems prior intentions restructure a multidimensional space into a new organization characterized by a new set of coordinates and new dynamics. Since contextual constraints that partition a prior intention's contrast space embody the emergent property of meaning and the laws of logic, it is plausible to assume that the cognitive level of organization will show semantic and logical consistency. Dynamically, that means that once I form the prior intention to greet you, not every logical or physically possible alternative remains open downstream, and those that do are contoured differently: the probability of waving or saying "Hi!" goes up; the probability of turning away goes down.

It is reasonable to stipulate that agents who are aware, however faintly, of their behavior's nonbasic ramifications include them among the alternatives of their contrast space and its dynamical pathways, if only by default. Knowingly not preventing something adverse of which one is (however dimly) aware is tantamount to choosing it, in a derivative sense of "choice." We often assume that the degree of awareness is correlated with the significance of those ramifications and accuse those who fail to take appropriate action to prevent their occurrence of being "in denial." As the medievals consistently remind us, acts of omission are acts nonetheless, but only if the agent is aware of the omission! By definition, if the agent logically, cognitively (or emotionally?) could not even have considered an alternative as such, it is not something he or she could have "omitted." The question "Of what was the agent minimally aware, and when?" thus remains central to the question "What did the agent intend?"

Thinking of prior intentions as parsing a self-organized dynamics in this way also explains why settling on one prior intention (say, the intention to rob) precludes settling on a logically conflicting one (not to rob). Once I decide to perform act-token *A*, non-*A* is no longer a viable alternative (Bratman, 1987): it drops out downstream as one of the coordinates. There is, *a fortiori*, no attractor that will get me there from here. Once I decide to greet you, the probability of my insulting you decreases

dramatically.

Each level is partly decoupled from the specific components that make it up. The higher level is meta-stable despite multiple realizability at the lower level. So, too, with respect to the mental. Once the cognitive, intentional level self-organizes, the fact that any one of several neurological processes can implement the same mental event becomes irrelevant. Under normal circumstances, my intention and subsequent action are indifferent to whether the former is realized in any particular neurological process. The alternatives that matter with respect to whether the behavior "raise my arm" constitutes an act-token are whether I intended to raise it, whether someone else lifted it, and whether it occurred as a result of a muscle spasm, for example. What matters, that is, is whether or not the neural process transmits information as mental. The presence and interference of those possibilities matter because trajectories originating in a spasm, for example, would take place entirely outside semantic space. Since the same neurons can be implicated in trajectories inside or outside semantic space, the neurological processes, as neurological, don't matter. The role that the intended meaning plays in bringing about the behavior is what counts.

Imposing order is what all top-down constraints of dynamical systems do (Pols, 1975, 1982). Whether in auto catalytic cycles or human beings, significance is a result of the interplay between the system's own top-down inhibiting constraints and the alternatives available to its components. In linguistics, constraints supplied by a sentence's context narrow the potential meaning of individual words. In genetics, the overall context in which a particular DNA codon is located affects its phenotypic manifestation. There is no understanding meaning independently of those inter-level relationships.

Robert Ulanowicz (1997) labeled an organized entity "ascendant" insofar as it develops as a focus of influence that grows in cohesion and integrity and thereby withstands the environment's destabilizing influence. The in-house pruning and streamlining of auto catalytic webs are carried out in the service of the higher level's focus of influence. While the system is ascendant, its overarching dynamics function as its formal cause, constraining the lower levels that make it up. In that way, they ensure the continuity and enhancement of the global level. In self-organizing, that is, a complex system partly decouples from the environment, from which it wrests a measure of autonomy. The difference between the behavior of objects while they are independent entities and their behavior once correlated and interdependent provides a measure of

the contextual constraints in effect at the global level (see Brooks & Wiley, 1988). The greater the difference between random and systematic behavior, the more stringent the constraints reducing potential variation must be. That difference measures the system's degree of organization and its autonomy *vis-à-vis* the environment. Dynamical systems are also partly independent of their parts, which, in self-organizing, have become replaceable components. Once organized, a system's attractors serve as its formal and final cause, both preserving its identity and drawing behavior into its overall organization.

Divergence from randomness is a measure of any dynamical system's integrity or cohesion relative to the environment's disintegrating effects; that is, of the system's independence from its environment. That ontology underwrites a particular epistemological stance: behavior constrained top down is always amenable to purposive and intentional characterization (Dennett, 1987). This is why behaviorists described their pigeons teleologically as "seeking food," and why news agencies phrase their reports as: "In an attempt to stave off a takeover, GM today took measures to..." The more robust a system's higher levels of organization, the more they and not external mechanical forces control the output. The more robust a system's higher levels of organization, therefore, the freer the resulting behavior. The system is autonomous; it behaves "from its own point of view." That is one sense in which any behavior constrained top down can be considered free.

## *FREEDOM AND WILL*

Insofar as all self-organizing structures, from hurricanes to ecosystems, act to preserve and enhance the overall global level, even at the expense of the particular components, complex systems are goal directed, if not fully goal intended (Dretske, 1988). By curtailing the potential variation in component behavior, however, context-dependent constraints paradoxically also create new freedoms for the overall system. As we saw, each emergent level of self-organization is nearly decomposable from the one below, and each new order possesses emergent properties absent at lower levels. That ontological feature allows scientists to identify and study chemical processes without having to refer to physical processes. Emergent, high-level psychological properties correspond to complex neurological dynamics constructed as a result of the co-evolution of human beings and the complex social organization that they both structure and are structured by (Artigiani, 1995). Once the neuronal processes

self-organize into a conscious and meaningful space, behavior constrained by that organization can express and carry out an agent's intent. And just as the constraints of syntax allow meaning to be expressed, constraints on behavior thus make meaningful actions possible. At the same time as the intention's meaning and values limit behavioral alternatives, a renewed variety of possible behaviors also opens up.

In humans, there emerges both the remarkable capacity for self-awareness and the sophisticated ability to think of, describe, judge, and act in terms of the meaningfulness of our choices—even in terms of ethical and aesthetic values (Artigiani, 1996). The greater the phase space, the greater the number of alternatives available to the organism. To the extent that higher-level semantic considerations constrain behavior, it has more and different alternatives open to it than before. The enlarged phase space is the novel, emergent capacity to act. The astounding number of dimensions (which dynamical self-organization has made available to human beings) provides a second sense in which human beings are free.

## A DIFFERENT APPROACH—HERMENEUTICS

Because deduction has failed as a generalized model that can explain complex systems' evolutionary sequence, a different logic of explanation appropriate to action is necessary. I propose that from now on the covering-law model of explanation (including its probabilistic incarnation) should be considered the limit of explanation, adequate for those phenomena that can be idealized as atemporal and acontextual. For isolated, linear systems, the covering-law model often works fine. For those phenomena, the lighter the inference, the better the explanation and the more accurate the prediction. For open, complex dynamical phenomena in which context-dependent constraints (both bottom up and top down) create inter-level interactions—that is, for phenomena which that "strange form of causality" progressively individuates and marks as essentially historical and contextual—the deductive model simply won't do, however. Understanding human action must begin from the assumption that people are dynamical entities whose behavior reflects their complexity.

The logic of explanation of hermeneutics is appropriate for explananda whose very nature is a product of that strange circle of whole and part. In contrast to covering laws and algorithms and deductions therefrom, that is, interpretation or hermeneutics reproduces the very logic of nature's open, adaptive dynamics. In textual interpretation, "the

anticipation of meaning in which the whole is envisaged becomes explicit understanding in that the parts, that are determined by the whole, themselves also determine this whole" (Gadamer, 1985: 259). Interpreters must move back and forth: the whole text guides the understanding of individual passages; yet the whole can be understood only by understanding the individual passages. This inter-level recursiveness, characteristic of hermeneutics, is thus "a continuous dialectical tacking between the most local of local detail and the most global of global structure in such a way as to bring both into view simultaneously" (Geertz, 1979: 239). The inter-level tacking of the hermeneutic "circle" reproduces the self-organization of complex dynamical processes. By showing the dynamics of complex adaptive systems, hermeneutical narratives are uniquely suited as the logic of explanation of these strange-loop phenomena.

Like intentional actions, interpretations are characterized by strange-loop, inter-level relations and are, in consequence, essentially contextual and historical. Interpretations therefore explain by showing those nonlinear, inter-level processes at work. Behavior that occurs between phase changes must be explained by a hermeneutical, interpretive reconstruction. First, the agent's mental state that initiated the behavior must be identified. To do this, the explainer must describe both the contrast space of alternatives that embodies the agent's frame of mind and the attractor–separatrix dynamics that govern those coordinates. Unlike that of Newtonian particles, dynamical systems' behavior depends crucially on their history and experience and on the environment they are currently in. Whether the explanandum is a snowflake or a person, explaining any dynamical system's behavior requires that we fill in all that relevant background. Explanations of actions must therefore provide a narrative that interprets and recounts what those cognitive and affective dynamics were like at the time they initiated the action.

Next, because the agent opted for one of the alternatives in the contrast space, the historical and interpretive narrative must describe the specific path the behavior took, mentioning at each step along the way how much was specifically constrained by the intention and how much by the lay of the land or the external structure of the intention's control loop, as well as by the agent's other dynamics. The explainer must also determine how much equivocation, if any, compromised the flow of the intention's content into behavior, as well as how much information available at the behavioral end is extraneous noise unconstrained by intentional attractors.

Narrative hermeneutical explanations are not simple temporal listings

of discrete events; that is, mere chronology, a linear sequence of independent frames on a film. In a true interpretive narrative, the telling of the tale explains by knitting together sequential but interconnected threads, such that it describes a temporal and contextual pattern, the meaningful organization that flows through the singular sequence of events and binds them into a whole. Just as first-order contextual constraints bind individual molecules into an integrated auto catalytic network, hermeneutic explanations of actions must construct a narrative that hangs together as a story. Unlike covering-law explanations of behavior, which abstract away time and space in favor of universalities, hermeneutics explains by highlighting and showing the concrete and temporal, context-dependent dynamical inter-relationships that give the action its unique character.

## THE ROLE OF PHASE CHANGES

But, this can be done only during "stable" periods. Between phase changes, while a system is in a particular dynamical regime, naturally occurring fluctuations and perturbations are damped, and explanation will consist in the back-and-forth interpretive reconstruction of the established dynamics that originated and constrained the behavior, and then in the tracking of the actual trajectory to its terminus in actual behavior. But a phase change itself cannot be explained that way. What needs explaining there is the change itself in the established attractors, the system's transformation into an entirely different dynamical regime. An alcoholic's turnaround cannot be explained in terms of their earlier state of mind. The radical mental transformation itself needs explaining.

A phase change is the qualitative reconfiguration of the constraints governing the previous attractor regime. The shift creates new relationships among the system's components as well as between the system and its environment. Phase changes signal a reorganization of the old dynamics into a new system with renewed relationships among the parts. These new relationships embody new properties and are governed by new laws. Within an established dynamical regime, the components' meaning is given by their contextual setting. There the meaning of individual actions depends on the agent's overall psychological dynamics, in combination with the circumstances in which these are embedded and from which they issued. Determining when or even whether a system will undergo a phase change and switch attractors is even more difficult than reconstructing either the dynamics of an established state or a trajectory

through that established regime. Close inspection, however, can at times reveal that a phase change is imminent. When dynamical systems are taken far from equilibrium to a critical threshold, the pull of the established attractor begins to weaken, The landscape begins to flatten out, so to speak. The system accesses states that it would not ordinarily have visited. In human terms, when a person's behavior begins to fluctuate widely such that previously uncharacteristic behavior becomes commonplace, watch out: a psychological crisis is in the offing.

It is impossible to predict with certainty whether an established attractor regime will be able to damp a naturally occurring fluctuation and stabilize; or, to the contrary, whether the system will reorganize or disintegrate. Which critical fluctuation happens to be the one around which the system will reorganize or which perturbation is the one that will destroy the system is often a chance matter. And not just which: *when* is just as important. Timing is crucial: the perturbation that would have taken us over the edge as a child might have only minimal impact today. A full narrative explanation must include all these details. Interpretive explanations of individual actions are therefore always historical and concrete.

## INTERPRETATIONS AND SENSEMAKING

The threat of relativism lurking in the hermeneutic circle has often encouraged philosophers to reject it. By drawing the explainer and the explanation into its strange loop, hermeneutics appears to forestall the possibility of any claim to truth and certainty. If we live in a dynamical universe, the novelty and creativity such complex systems display do indeed signal the end of eternal, unchanging, and universal certainty. Unlike modern science, however, dynamical systems theory provides an understanding of both the construction and the integrity of wholes that does not dissolve their unity at that level. According to Gadamer (1985), the resolution to the circularity of hermeneutics is found in Heidegger's recognition that "the circle of the whole and the part is not dissolved in perfect understanding but on the contrary, is most fully realized" (261).

We make sense of persons and their actions through an interpretive dialectic between wholes and parts. From descriptions of the dynamics of a particular instance of behavior, it might be possible to reconstruct the agent's character or personality and therefore the intention that constrained the behavior. We can then examine other examples of that person's behavior to see whether the character that these additional exam-

ples suggest corresponds to the personality we inferred earlier. If the test is positive, we reasonably conclude that the first behavior was "in character." As a result, we judge it to have been the agent's action and hold them responsible. From empirically available information of a behavioral output, and taking care to note any interference from noise or equivocation, the explainer attempts to reconstruct the cognitive source from which the behavior issued. The explainer attempts, in other words, to determine the mental dynamics from which a given instance of behavior flowed and the particular intentional attractor that constrained it. The purpose of examining several examples of behavior is to fashion an interpretation of the unknown dynamics (the intention) that constrained the particular behavior in question. The explainer then checks their interpretation of the agent's character by examining whether subsequent behaviors fit that initial (always tentative) interpretation.

## *Providing Explanations*

Between phase changes, a complex system's behavior is governed by both the combined constraints of its own internal dynamics and the initial and on-going conditions in which it finds itself. As a result, in contrast to the covering-law model, the direction of explanatory primacy in interpretation is not the usual downward, reductive direction (Wimsatt, 1976). It moves up and down levels, from whole to parts, from inside to outside, and vice versa. This tacking reproduces the inter-level ontological processes that created the explanandum. In contrast to behaviorist analyses that attempt to bypass the subject's actual mental state, moreover, the hermeneutic explainer tries to reconstruct the particular internal dynamics from which the actual behavior issued, not bypass them. Wright (1976) claimed that only the impact of the mental event must be taken into account; this impact, however, could not be determined through commonalities across stimuli-response patterns of behavior over time.

Because complex attractors are often implicated in human actions, differences in behavior are sometimes more informative than commonalities. Variations across examples of behavior can reveal a particular intention's complexity in a way that similarities cannot. The back-and-forth tacking of hermeneutics can reveal the convoluted structure of those variations in a way that the covering-law model cannot. By respecting the vectorial nature of the trajectory it is reconstructing, an interpretive narrative does not try to reduce the purposiveness of action to nonpurposive elements.

The whole point of hermeneutical interpretation of action is to show

how meaningful intentions emerge and then purposively to constrain the behavior that flows from them. By recognizing that they are dealing with a unique trajectory, interpretive narratives also take for granted that their account need not apply to other behaviors, even those that appear similar. It is important, nevertheless, to emphasize that interpretation can discover only whether a particular instance of behavior was "in character." That sometimes—perhaps often—it is possible to judge someone's behavior accurately as being "in character," however, should not lull us into believing that we have achieved certainty in judging a particular instance of behavior. That is, when we are dealing with complex adaptive systems, surprises are unavoidable. Because of their sensitivity to initial conditions—due, in turn, to their contextual and temporal embeddedness—complex adaptive systems are characterized by unusual twists and novel turns. We saw the havoc that equivocation and noise can wreak in interrupting and compromising an intended trajectory.

Since we will never be able to specify any dynamical system's initial conditions to the requisite (infinite) degree, *a fortiori* we will never be able to capture all the details and circumstances of anyone's life and background. Given this limitation, we must always keep in mind that reconstructing specific instances of behavior will always be, at best, an interpretation and not a deduction—a much more fallible type of explanation than we had previously hoped was available. Interpretations of human action are always tentative (Metzger, 1995). Absolute certainty about either what the agent just did, or what they will do, specifically, a year from now, is impossible, As the title of Prigogine's last book (1996) announces, the dynamics of complex systems signal the end of certainty. The exact trajectory of any stochastic entity captured by a complex attractor, even between phase changes, is impossible either to predict or to retrodict precisely, even in principle. It cannot be predicted with exactitude in part because of the multiple realizability that self-organized systems support and the mischief that initial conditions wreak, but also because open dynamical systems' control loop runs partly through the environment. The dramatic fluctuations in behavior that strange attractors allow make any hope of predicting a dynamical system's specific future trajectory a futile wish. Knowing that someone is ambitious will not tell you what specific path their behavior will take.

Even though agents are usually in a privileged position to determine the alternatives present in their conscious contrast space and the degree of their awareness of each, extensive reflection and probing may be necessary before they can articulate all the relevant content of that aware-

ness. In the end, the subject's own overall mental state (whose coordinates and dynamics identify the contrast space) parses potential behavioral alternatives from others not even contemplated.

I have claimed that a given instance of behavior constitutes an act-token (as opposed to a nonact) if and only if the information available at the behavioral end was constrained by (not merely contingently connected to) the intention's dynamical attractor. Even after the contrast space of alternatives that the agent had in mind has been established, however, the explanatory narrative must still historically reconstruct the behavior's actual trajectory and show that the intentional source constrained it unequivocally. For it to do so, the explainer must describe at each choice point why the agent took this fork rather than that one: what were the available options, and which was chosen.

Explaining why the agent took this path rather than that after forming the prior intention will require reconstructing the agent's background, circumstances, particular frame of mind, and reasoning, whether self-conscious or not. The cautionary tale of complex dynamics is that we can never be absolutely certain that there is no complex attractor constraining the behavior; we just might not have found it yet. It took a long time for science to discover that chaotic behavior is not chaotic at all. Even worse, what looks like behavior constrained by an intention may in fact have been noise all along. Explaining the agent's convoluted path from one point to the next, therefore, will often require identifying many other internal dynamics and external circumstances involved in bringing about the behavior. When all the intertwined attractors (all the entrained emotional, sociological, psychological, and other attractors) that make up a mature person are taken into account, those labyrinthine explanations we often launch on do not seem so preposterous after all.

## *Summary*

Nineteenth-century hermeneutics failed to take into account that the explainer, as much as the phenomenon explained, is embedded in time and space. Twentieth-century students of hermeneutics, in contrast, have finally come to appreciate that interpretation is doubly historical. The phenomenon being explained has a history, and so must be understood within that history; but interpreters, too, are situated within history, within a tradition, which their interpretation both reflects and influences. This double historicity affects the pragmatics of explanation. When the subject is planetary orbits and billiard balls—that is, when interactions

can be ignored—the role of interpreter recedes in importance; not so when the subject is either quantum processes or human actions, Dynamical systems have therefore brought the interpreter back into the pragmatics of explaining action (if not into the metaphysics of explanation, as quantum processes have).

In dynamical terms, the tradition in which interpreters are situated is itself an attractor. As social beings, interpreters are embedded in its dynamics. As Gadamer (1985: 216) notes, "The anticipation of meaning that governs our understanding of a text is not an act of subjectivity, but proceeds from the communality that binds us to the tradition" that frames our interpretation. This fact, on which even the popular media harp, need not lead either to paralysis or to the deconstructionist's conclusion that any interpretation is as good as any other. As Umberto Eco (1990: 21) insists and our discussion of top-down constraints has shown, context constrains the range of plausible interpretations: "A text is a place where the irreducible polysemy of symbols is in fact reduced because in a text symbols are anchored in their context."

I submit that two contexts provide an action's "literal" meaning: the historical background and contextual setting in which the action was performed, and the context established by the "small world" of the action itself. Two contexts likewise frame the meaning of a hermeneutical explanation: the historical background and contextual setting in which the interpretation is offered, and the context established by the "small world" of the interpretation itself. Hermeneutic interpretation, within a narrative framework, thus comes closest to the logic of explanation advocated by David Lewis (1973a). Twenty-five years or so ago, Lewis argued that causes should be analyzed in terms of counterfactuals: if $x$ had not occurred, $y$ would not have. Despite the potential objection—"What underwrites the counterfactual itself if not causality?"—thinking of $y$ in terms of its dependence on $x$ can be helpful in capturing the way meaning flows from intention into action.

Within stable periods, the system's dynamics do much of the causal and explanatory work: the initial conditions account for the particular twists and turns within the behavior's attractor. The less complex the system, the more explanatory work the dynamics do; the more complex the system, the more the initial conditions do. Between phase changes, on the other hand, it is first necessary to reconstruct the process that drove the system far from equilibrium. Did the agent's own internal dynamics drive them to a threshold point? Or was it more a case of external perturbations driving (the agent's) weak internal dynamics to an instability? There is no one-to-one relationship between the dynamics in place

before the phase change and those that appear after. The direction that a stochastic dynamical system's bifurcation will take cannot in principle be predicted even by ideal, omniscient observers. The precise path that the phase change takes can be explained only after the fact.

Such explanation must take the form of a genealogical narrative that reconstructs the bifurcation by painstakingly describing (1) the inter-level and contextual interactions that took the system far from equilibrium in the first place, (2) the particular fluctuation or perturbation that drove it over the edge, and (3) the specific pathway that the bifurcation took (as opposed to other possible alternatives). Phase changes cannot be explained in terms of the dynamics from which they issued. The reason, to repeat, is that phase changes mark a qualitative, catastrophic transformation in the dynamics themselves. Across phase changes, therefore, what requires explanation is how the meaning that governed one stable state is transformed into qualitatively different dynamics governing a different space of possibilities with a different frequency distribution.

Phase changes embody essentially incompressible information. That is, there exists no law or algorithm more concise than the process itself that can capture and describe what happened. That is why fiction and drama, bible stories, fairy tales, epics, novels, and plays will always be better than deductions or formulas for explaining personal transformations of this sort. The rich, vivid descriptions and reenactments that these genres provide represent meaningfully for the reader and spectator the processes that precipitate such personal transformations. They do so by paying special attention to the role played by both the agent's internal dynamics and the particular environmental perturbations that drive a system far from equilibrium. Stories and dramas also show the reader and viewer how random, unrepeatable events and circumstances can be responsible for either destroying people or renewing them at a different level. Often insignificant in themselves, these unique events can be the proverbial straw that breaks the camel's back.

Had the characters in those novels and dramas not been near a crisis point, of course, those unique events would not have had an impact. By interacting with background conditions far from equilibrium, unique events provide the turning points along a singular trajectory. Reenactment, which is what both simulations and theatrical performances offer, is even more explanatory than narrative, because we get to see how the tensions of living with Torvald Helmer, George Tiesman, and Charles Bovary drive Nora, Hedda, and Emma to the edge.

## REFERENCES

Alvarez de Lorenzana, J. (1993) "The constructive universe and the evolutionary systems framework," Appendix to S.N. Salthe, *Development and Evolution: Complexity and Change in Biology*, Cambridge, MA: MIT Press.

Artigiani, R. (1995a) "Self, system and emergent complexity," *Evolution and Cognition*, 1 (2): 137–47.

Artigiani, R. (1995b) "Toward a science of meaning," *Technological Forecasting and Social Change*, 48: 111–28.

Artigiani, R. (1996a) "Contemporary science and the search for symmetry in nature and society," *Symmetry: Culture and Science*, 7 (3): 231–46.

Artigiani, R. (1996b) "Societal computation and the emergence of mind," *Evolution and Cognition*, 2 (1): 2–15.

Audi, R. (1995) "Mental causation: sustaining and dynamic," in J. Heil & A. Mele (eds), *Mental Causation*, Oxford: Oxford University Press.

Bohm, D. (1971) *Causality and Chance in Modern Physics*, Philadelphia: University of Pennsylvania Press.

Bratman, M. (1987) *Intentions, Plans and Practical Reason*, Cambridge, MA: Harvard University Press.

Brooks, D. & Wiley, E. (1988) *Evolution as Entropy* (2nd edn), Chicago: University of Chicago Press.

Bunge, M. (1979) *Ontology II: A World of Systems*, Dordrecht: D. Reidel.

Campbell, J. (1982) *Grammatical Man*, New York: Simon & Schuster.

Danto, A. (1979b) "Causality, representations, and the explanation of actions," *Tulane Studies in Philosophy*, 28: 1–19.

Dennett, D. (1987) *The Intentional Stance*, Cambridge, MA: MIT Press.

Dewan, E. (1976) "Consciousness as an emergent causal agent in the context of control system theory," in G. Globus, G. Maxwell, & I. Savodnik (eds) *Consciousness and the Brain*, New York: Plenum.

Dretske, F. (1981) *Knowledge and the Flow of Information*, Cambridge, MA: MIT Press.

Dretske, F. (1988) *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press.

Dyke, C. (1988) *The Evolutionary Dynamics of Complex Systems*, Oxford, England: Oxford University Press.

Eco, U. (1990) *The Limits of Interpretation*, Bloomington: Indiana University Press.

Gadamer, H.G. (1985) *Truth and Method*, New York: Crossroad.

Gatlin, L. (1972) *Information and the Living System*, New York: Columbia University Press.

Geertz, C. (1979) "From a native's point of view: on the nature of anthropological understanding," in P. Rabinow & W.M. Sullivan (eds), *Interpretive Social Science: A Reader*, Berkeley and Los Angeles: University of California Press.

Hofstadter, D. (1979) *Gödel, Escher and Bach*, New York: Basic Books.

Jantsch, E. (1980) *The Self-Organizing Universe*, Oxford, England: Pergamon.

Juarrero, A. (1999) *Dynamics in Action: Intentional Behavior as a Complex System*, Cambridge, MA: MIT Press, http://mitpress.mit.edu/book-home.tcl?isbn=0262100819.

Juarrero-Roque, A. (1997) "Language competence and tradition-constituted rationality," *Philosophy and Phenomenological Research*, 51: 611–17.

Kim, J. (1995) "Can supervenience and 'non-strict laws' save anomalous monism?" in J. Heil & A. Mele (eds), *Mental Causation*, Oxford, England: Oxford University Press.

Lewis, D. (1973a) "Causation," *Journal of Philosophy*, 70: 556–67.

Metzger, M. (1995) "Multiprocess models applied to cognitive and behavioral dynamics," in R. Port & T. van Gelder (eds), *Mind as Motion*, Cambridge, MA: MIT Press.

Pols, E. (1975) *Meditation on a Prisoner Towards Understanding Action and Mind*, Carbondale: Southern Illinois University Press.

Pols, E. (1982) *The Acts of Our Being: A Reflection on Agency and Responsibility*, Amherst: University of Massachusetts Press.

Prigogine, I. (1996) *The End of Certainty: Time, Chaos, and the New Laws of Nature*, New York: Free Press.

Salthe, S. (1985) *Evolving Hierarchical Systems*, New York: Columbia University Press.

Salthe, S. (1993a) "Development and evolution as aspects of self-organization," in M. Sintonen & S. Siren (eds), *Theory of Evolution: In Need of a New Synthesis?* Philosophical Studies from the University of Tampere, 19.

Salthe, S. (1993b) *Development and Evolution*, Cambridge, MA: MIT Press.

Sosa, E. (1980) "Varieties of causation," *Grazer Philosophische Studien*, 11: 93-103.

Thelen, E. & Smith, L.B. (1994) *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press.

Ulanowicz, R. (1997) *Ecology: The Ascendent Perspective*, New York: Columbia University Press.

Wimsatt, W. (1976) "Reductionism, levels of organization, and the mind–body problem," in G. Globus, G. Maxwell, & I. Savodnik (eds) *Consciousness and the Brain*, New York: Plenum.

Wright, L. (1976) *Teleological Explanations*, Berkeley and Los Angeles: University of California Press.

Zeleny, M. (ed.) (1980) *Autopoiesis, Dissipative Structures, and Spontaneous Social Orders*, Boulder, CO: Westview.

Found Sci (2013) 18:757–780
DOI 10.1007/s10699-012-9304-9

# On the Import of Constraints in Complex Dynamical Systems

**Cliff Hooker**

Published online: 18 October 2012


**Abstract** Complexity arises from interaction dynamics, but its forms are co-determined by the operative constraints within which the dynamics are expressed. The basic interaction dynamics underlying complex systems is mostly well understood. The formation and operation of constraints is often not, and oftener under appreciated. The attempt to reduce constraints to basic interaction fails in key cases. The overall aim of this paper is to highlight the key role played by constraints in shaping the field of complex systems. Following an introduction to constraints (Sect. 1), the paper develops the roles of constraints in specifying forms of complexity (Sect. 2) and illustrates the roles of constraints in formulating the fundamental challenges to understanding posed by complex systems (Sect. 3).

**Keywords** Complexity · Complex systems · Dynamical constraints · Lagrangian dynamics · Organization · Dynamical understanding

## 1 Constraints (An Uneasy Introduction)

A constraint, as the name implies, specifies some limit on independent behaviours. A marble rolling in a bowl, e.g., is confined to the surface of the bowl, so that its position in any one spatial dimension is not independent of its positions in the other two dimensions. In contrast, a small spacecraft has no such constraints on its translational movement. Its *degrees of freedom* include the abilities to move in all three dimensions independently. (It can also rotate independently around an axis in each dimension; so, assuming it is a rigid body, all told it has 6 degrees of freedom.) The marble's bowl leaves it only 2 translational degrees of freedom. The bowl constraint is conveniently expressed by the relationship among positions that describe the bowl's surface. (If it is a spherical bowl of radius r then the constraint relation among the positions x, y, z is $x^2 + y^2 + z^2 = r^2$). This characterisation generalises: a *constraint* on a dynamical process is a reduction of its underlying degrees of freedom arising from the

C. Hooker (✉)
School of Humanities and Social Sciences, McMullin Bld, University of Newcastle, Callaghan, NSW, 2308 Australia
e-mail: Cliff.Hooker@newcastle.edu.au

physical conditions in which the process takes place. A system's effective degrees of freedom are those provided by its inherent variabilities (its dynamical variables) minus those removed through constraints.[1] Constraints are expressed as relationships among system variables.

Constraints play a central role in dynamics. Currently, we work from a general Lagrangian/Hamiltonian dynamical formalism as mathematical paradigm, where dynamical motions are represented as passages along system trajectories (flows) measured (parameterised) by time and driven by energy gradients (forces) under the very general formula 'F = ma'.[2] Typically, the forces are the basic interaction forces operating among basic system components, e.g. gravitational or electromagnetic forces among particles or chemical forces among molecules. If this exhausts the system, its behaviour is said to be free, not under constraint, that is, it has available all of its degrees of freedom, = all of the mutually independent kinds of basic motion its interaction structure permits (translations, rotations, etc.). Constraints then represent additional forces on the system that also contribute to shaping the system dynamical behaviour, but for one reason or another are not specified in detail. Typically this is because they change only on sufficiently longer timescales than the interaction phenomena we wish to study that they can be treated as constants and in that sense external to the dynamics to be studied. In any case, conveniently, their effects for the interaction dynamics to be studied can be sufficiently summarised as the requirement that certain relationships among the interaction variables be maintained invariant across the interaction dynamics.[3]

Given the glorious scientific history of our mastery of fundamental dynamics, here is a remarkable thing: within the general Lagrangian/Hamiltonian dynamical formalism we currently only know how to systematically construct coherent dynamical models for systems that are free or at least that do no work on their constraints, the canonical form for reversible dynamics.[4] If the marble rolls without friction on the bowl's surface, making contact but no more, then this last condition is satisfied. It means that the force exerted by the bowl on the marble is everywhere orthogonal to (that is, perpendicular to) its motion. This is how we normally imagine the bowl situation, but that involves assuming that gravity is the only basic interaction force and the marble and bowl surfaces are each smooth and rigid constraints; then it is easy to suppose that the marble rolls smoothly and reversibly. But suppose instead that the bowl surface is rough and deformable and so there was rolling friction between marble and bowl. Under that constraint the marble loses energy as it does work against friction, generating an irreversible dynamics of different form to that of smooth rolling. A plastic bowl might even soften when heated by the rolling friction, which would alter the constraint itself as the very motion proceeds to which the constraint gave rise, like the banks of a river eroding as the flow they channel wears against them, a radical dynamical irreversibility. Here

[1] It is possible in principle to also constrain a system's parameter ranges—in effect, to constrain the range of quantitative forms its dynamics can take. The net effect of such constraints will still appear as (indirect, often extremely complex) constraints on variables.

[2] For constant m. According to Coleman and Korté (1999) a 2nd order differential equation of this general form is a structural feature of any ('non-pathological') space-time geometry and force field.

[3] In Sect. 3 we will return to the internal/external distinction when we inquire about boundaries and the 'self' in self-organisation—see note 31 and text, cf. Bénard cell formation and internal constraints below and note 14.

[4] Such constraints are referred to as 'ideal' constraints—meaning both that they permit the use of an elegant and powerful mathematical apparatus, Lagrangian/Hamiltonian dynamics, and that it takes its simplest form under these conditions. This last is something of a hopeful wish, as we will see—there may be no extended, more complicated but unified version of which the present formulation is a special, idealised case. Thus in this single use of 'ideal' does science run together two utterly distinct notions, ideal = best in usefulness (value), ideal = idealised (simplified). On the importance of understanding the latter sense of idealisation in complex systems see Hooker (2004, 2011c).

is where the notion of constraint starts to introduce complications to formulating dynamics. Unlike the unvarying general schema of motions and interaction forces that frame a general theory of dynamics, classically 'F = ma', it is a detailed empirical matter what kind of constraints there are and what conditions (like orthogonality) they satisfy, and these differences can make a difference to the basic form of the resulting dynamics, not just to its details.

As well as their motions, each of these different cases makes a distinctive difference to the nature of the mathematical problem posed by their dynamical description. If the constraints are holonomic—roughly, of fixed geometry, like the bowl surface when there is frictionless rolling—then the mathematical equations are analytically solvable and the global motion of the system can be summed up in mathematical functions (the scientific 'gold standard'). But holonomicity easily fails, e.g. any object changing its constraint type (a ball rolling off a table edge, a flowing river) renders the overall motion constraint non-holonomic. But where work done on constraints actually alters them as motion progresses (e.g. river banks) there may be no mathematical formulation of the problem available within the current dynamical Lagrangian/Hamiltonian formalism. This grand lacuna is potentially vast, applying not only for plastic bowls and rivers but also for all living things, since in a group of self-regenerating processes (e.g. cell metabolism) the constraints that produce any one product must be largely constituted by other such products and work must be done to recreate them following their metabolic use and entropic decay (see Sect. 3 below).

The little 'toy' system of bowl + marble is useful for introducing another role for constraints, namely in characterising dynamical bifurcations. Suppose our bowl to be hemispherical, made of a deformable but tough plastic, and sitting evenly on a base equipped with a vertical screw at its bottom point. With the screw not touching the bowl the marble rolls around the bowl's interior in smooth loops at all sorts of angles depending on the marble's initial 'launch conditions'. Now screw in the screw until a small mound appears in the bottom of the bowl. At this point any marble paths passing sufficiently near the bottom of the bowl will deviate from their old paths as they encounter the mound, altering the dynamics in a subtle but profound way that depends on the shape of the mound, e.g. the marble may now temporarily bounce off the bowl surface. As the screw is extended the effect will become more pronounced and widespread. The shift from the original dynamics to the new one is a bifurcation because it represents not only a change in specific behaviours (in trajectory shapes) but also a change in dynamical form: a differently structured dynamical equation is required to model the behaviours and the pattern of all possible trajectories (the flow) changes. This change is brought about by a change in constraints, namely in the shape of the bowl.

This is the simplest birfurcation, where the dynamical form depends on a constraint determined by the value of a single parameter, in this case screw height, and there is a threshold (here 0 screw height) at which the system changes its dynamical form and flow. A related bifurcation is the formation of 'rolling boiling' convection cells to replace horizontally layered conduction (Bénard cell formation) in a fluid heated from below as the applied heat flow (the relevant constraint here) passes a hydrodynamic threshold.[5] The particular pattern of convection cells obtained depends on another constraint: the shape of the fluid container. There are many other kinds of bifurcations but all depend on changes of constraint in these kinds of ways. We can be sure that bifurcations are objective and, e.g., not just mere shifts of pattern (cf. shadow plays), because new kinds of work are done, there is a shift in the dynamical form and thus energy flows.

The Bénard cell bifurcation has a feature common to most bifurcations that the simpler marble bifurcation lacks: the formation of a new internal constraint: the cells of rotating

[5] See e.g. http://en.wikipedia.org/wiki/Rayleigh-Bénard_convection, Bishop (2008).

fluid to which fluid molecules are constrained. Creating these involves a contribution from the liquid molecules themselves, that is, from within the system. For this reason bifurcations of the Bénard cell formation kind are often called self-organising. In the Bénard cell case, e.g., fluid flows along adjoining cell walls must move in the same direction, otherwise fluid flows would collide there, destroying the cell formation; but continuity requires that the flows on opposite sides of each cell must be in opposite directions, thus cell circulations must alternate clockwise/counter-clockwise across the whole fluid. Change one cell's circulation and every other cellular circulation must also change. For this reason the cellular formation is said to show global or generalised rigidity. If this seems an active internal ordering by the system itself, it equally raises the question of how it was achieved. How is this global information generated, disseminated and effectively applied throughout the system? We lack detailed answers beyond gesturing at the idealisation of a sea of competing coherent micro fluctuations one of which was amplified by the dynamics and propagated to global fixation.

This issue of the system 'self' in self-organisation is complemented by another, the shift of constraints through an enlarging system self. The role of an apparent self arises for Bénard cell formation because the internal motions change holistically at the bifurcation threshold, in ways shaped by the container constraint. But why not absorb the container into an enlarged system as a molecular lattice and consider the bifurcation dynamics of this enlarged molecular system? Should not it reproduce the cell formation bifurcation without the need for an explicit external container constraint? This may seem 'unnatural' when the usual always-rigid metal container is involved, but it looks rather more 'natural' if the container can soften under heating so that it undergoes a mutual development in interaction with the fluid, like the river banks earlier, and even necessary if the container can melt and join the original fluid in a joint fluid dynamics. (Anyway, ultimately such intuitions are irrelevant, only the coherent dynamical possibilities matter.) And why not then also absorb sufficient of the heat source as well, so that both erstwhile constraints are absorbed into internal dynamics? And on to absorb the dominant source of gravitation and so on until the cosmos be encompassed. Can we not in this way do away with constraints altogether?

Certainly, there are many cases of no-external-constraint dynamics that show complex behaviour, e.g. the solar system (including a moving sun) can shift from coherent into chaotic motions. But here no external constraints were ever needed. As soon as we consider other systems, the constraints re-emerge, only now as internally generated. In both the usual and enlarged system representations the Bénard cell container's metallic molecular lattice must still constitute a constraint on the motions of its fluid molecules, sufficient constraint to retain the boiling liquid within them and impel a cell shape and a horizontal arrangement of cells. Constraints do not have to be pre-formed: free-falling liquid raindrops form internal molecular lattice constraints as they cool that join to form snow flakes, a lattice that thereafter constrains all their molecular motions, and this holds for all cooling phase changes. And so on. The point is, constraints are as dynamically real as the focal interaction dynamics of a system, because the constraints themselves are generated from the interaction dynamics of the elements that subserve (are the sources of) those constraints and so is their capacity to filter out the thus-constrained motions of the interaction dynamics. One can deduce from the properties of a metal lattice (the constraint) the existence and properties of the Fermi conduction bands (the filtered electrical conduction behaviour). The internal formation of constraints is the most dynamically robust form of emergence since a new dynamical individual is objectively born from the bifurcation, and thus carries new causal (energetic) powers. So external constraints cannot be made to disappear; instead they re-appear as internal constraints, where they also stand at the centre of understanding self-organisation and selfhood.

A final feature of constraints will complete this introduction to the idea. The term 'constraint' implies limitation, and specifically here it refers to limited access to dynamical states or, equivalently, reducing degrees of freedom by limiting dynamical trajectories to sub-sets of the basic interaction state space. This is the common *dis*abling sense of the term. But constraints can at the same time also be *en*abling, they can provide access to new states unavailable to the unconstrained system: equivalently, by coordinatedly decreasing degrees of freedom they provide access to dynamical trajectories inaccessible to the unconstrained system. For instance, low resistence electrical conduction is a state available to a metal lattice but unavailable to the metal molecules as a collection of independent entities (e.g. in a gaseous state). More generally, a skeleton is a disabling constraint, for example limiting the movements of limbs (cf. an octopus), but by providing a jointed frame of rigid components for muscular attachments it also acts to enable a huge range of articulated motions and leverages, transforming an organism's accessible niche, initiating armour and predator/prey races, and so on. It was Pattee who emphasised the importance of constraints, especially of such coordinated constraints, to biological organisation and evolution.[6]

Each of the eight great transitions in evolutionary history (Maynard-Smith and Szathmary 1995), e.g. the emergence of multi-cellular organisms, marks a new coordination of constraints. By permitting reliable cooperation instead of competition and reliable inheritance of the fruits of cooperation, these coordinations created new complexity and opened up vast new possibilities. Moreover, the constraints simultaneously operative in biological systems (a) are typically mechanical and chemical, solid and liquid, and so forth, (b) operate in an equal diversity of ways, e.g. in development they include quasi-equilibrium effects (differential adhesion), non-linear and non-equilibrium processes (chemical oscillation), phase transitions (epithelial-mesenchymal transformation; skeletogenesis), etc., and (c) operate across a variety of levels and spatio-temporal scales (Newman 2011a,b). Coordinated constraints can work their way around physical laws, e.g. while no single pump can lift water higher than 10 metres, trees lift it several times this by linking together many cellular pumps. While it is possible to obtain complex dynamics in simple systems, plausibly the only way in which the complex properties we will shortly introduce can be obtained is through complex coordination of constraints of the kind neural, muscular and skeletal coordinations exemplify. These begin already with the complex coordination of biochemical products and gradients that allow intra-cellular chemistry to support cellular maintenance. We are here far from the holonomic constraints of standard analytical mechanics and deep into the domain of multiple state-dependent, interacting, non-holonomic constraints.

Constraints, external or internal, specify context, the dynamical context in which interaction dynamics take place. But in this role constraints often do not stand aloof from interaction dynamics, as they do when in holonomic form, but rather they so interrelate with the interaction dynamics as to enable new dynamical forms and with that new selves (cf. note 39 on contextual emergence and reduction). Context and content are intimately interwoven, the most basic expression of the antinomy of individuality and wholeness. In Sect. 3 we will inquire a little further both into the notion of 'self' (note 3) and that of 'organisation', each so

[6] See Pattee (1971, 1973). Pattee (1973) also draws attention to the way in which constraints are typically specified as macroscopic effects relative to the dynamical relations involved in their constitution and often also relative to the specification level of the interaction dynamics, thereby implicitly defining the locus of a partition between macro and micro, a partition that could potentially be moved to encompass more and less within the micro level specification. This is, he suggests, directly reminiscent of the 'cut' between (relatively) macro observer and micro system in quantum mechanics. While this is intriguing, Pattee rightly remarks that after years of discussion the latter 'cut' is still not satisfactorily understood and, if my argument of Sect. 3 holds, neither is the partition for systems that alter their constraints by doing work on them.

glibly used in 'self-organisation', and into the equivalence of internal and external constraints discussed here.

## 2 Complexity (A Constrained Review)

With some feel for the notion of constraints in complex systems, it is time to turn to the first objective, to spell out the role of constraints in specifying forms of complexity, many of which tend to be overlooked. The preceding discussion has already thrown up several examples of complex behaviours, e.g. bifurcation thresholds and domains, and self-organised emergence through constraint formation. Unhappily, there is currently no unified theory available that captures all the cases of complexity, neither a fundamental dynamical one nor some higher order characterisation. We simply don't know enough, especially about the relevant mathematics, to tackle such an ambitious goal. Here the aim instead is to illuminate the principal system properties involved in complex dynamics, in whatever combinations, with particular attention to the roles constraints play in their conception and operation.[7]

*Group 1*: *general dynamical properties*. The discussion begins with a group of general dynamical properties that underpin later, more specialised properties. Here basic interaction dynamics dominate their characterisation and constraints play either no role or simply play their general dynamical role. There is, it would seem, just one general property that is necessary to obtain any form of complexity: *non-linearity* of basic interactions and the consequent *failure of additivity* (*non-superposition*).[8] Other properties that might have been considered contenders for necessity, especially *non-holonomic constraints* and *irreversibility*,[9] prove not necessary to complexity, as the example of unconstrained reversible 3-body gravitational systems shows. Neither are any of these properties sufficient for complex behaviour, as respectively shown by the examples of 2-body gravitational motion for non-linearity, a contained gas for non-holonomicity, and a freely expanding gas for irreversibility. We currently have no general account of which kinds and combinations of these properties will yield which classes of complex systems. However, as we approach what are intuitively more complex systems all three properties tend to form a common basis, e.g. for all living systems.

Neat formal criteria for complexity wont help us out of the woods here. The most prominent proposal is that of *algorithmic complexity*, roughly measured by how compressed (shortened) a description can adequately replace a complete detailed description of a system. But while repetitive systems like lattices come out simple, it leaves random gases, that generate no interesting complex behaviour, as hugely complex (since every molecule has to be separately

[7] For a more general discussion of these properties the reader is referred to Hooker (2011b,d), from which parts of this discussion, modified and improved, have been adapted. One advantage to setting out the extended set of complexity-making properties is avoiding a current tendency toward the mis-placed simplicity of focusing on just a select few characteristics, e.g. just chaos, or counter-intuitive/ unpredictable behaviour, or having a long incompressible description (algorithmic complexity).

[8] An interaction is *linear* in some variable v if the interaction force varies proportionately to v (formally, $F(kx) = kF(x)$, k a number) and non-linear otherwise. Gravitational force, e.g., is spatially non-linear since it varies as the inverse square of the interaction distance ($G(kr) = k^{-2}G(r)$). Interaction linearity yields linear dynamical equations describing the system and these are characterised by additivity: any numerical sum of solutions is also a solution. No complex dynamical behaviour is possible.

[9] A constraint is *holonomic* if its local (momentary, nearby) differential form is integrable to yield a global constraint relation that is purely a matter of space-time geometry and independent of the system's dynamical states, e.g. the frictionless bowl as constraint for the rolling marble, and otherwise is non-holonomic (cf. note 37). A process is reversible if its dynamics (its state sequence structure) remains the same under time reversal (and only the time-ordering of the sequence structure is reversed), e.g. if it obeys Newton's laws; and otherwise is irreversible. Irreversible processes involve energy dissipation.

specified) while blurring over all the distinctions in between (see below). The most complex systems described below, organisms, will be of large but intermediate value in algorithmic complexity.

Our next property is *complicatedness of dynamical domains*. The marble in the bowl can roll through the centre of its bowl or not; if it does the motion is an oscillation in a straight line while if not it consists of a loop about the centre of some kind. This yields two dynamical domains. A complex dynamics may generate several different equilibria of various stabilities (attractor basins of various shapes) either directly intersecting, even having fractal common boundaries, or connected by transient paths (where small disturbances may change its destination, cf. rolling on horizontal surfaces) plus other transient paths that do not end.[10] This 'attractor landscape' is a system's dynamical signature, expressing its dynamical form. By contrast, a linear dynamics has no attractor basins, hence no equilibria, simply a 'flat' landscape filled with transients. A system that remains within a single attractor landscape is structurally stable (= autonomous dynamics in mathematical parlance) and otherwise is structurally meta- or un- stable. While mathematical dynamics typically assumes structural stability, many complex systems are structurally unstable, e.g. exhibiting phase changes. Inter-landscape transitions are bifurcations. All this invites introducing both complicatedness of attractor landscape and of inter-landscape transformations as aspects of dynamical system complexity. In both cases constraints combine with interaction dynamics to create these complexity features in the manners already illustrated.[11] In principle, all of the properties to follow correspond to features of landscapes and landscape transitions; but we are far from being able to actually construct such systematic correspondences so it is still most useful to discuss particular properties.[12] Anyway, the roles of constraints are implicitly enfolded into the resulting landscapes rather than being made explicit as is the aim here.

There follow a collection of properties common to many complex systems but which are noted only in passing here since they do not concern new roles for constraints. *Amplification* locally and up to whole system level, especially the amplification of small perturbations and perturbations localised to a few components, is common in non-linear systems. Critical point birfurcations and many others where a new system constraint is formed are thought to occur through amplification of a fluctuation at component level in this way. Similarly,

[10] Equilibria of any sort are stable, meta-stable or unstable. An equilibrium in some aspect A is stable, with respect to some class of disturbances (perturbations) D, if (and only if) its response to any disturbance from D is to soon return near to (including exactly to) its original A condition under its own dynamical processes and remain there. An equilibrium is unstable to a class D of disturbances if it does not return near to its original A condition and it is meta-stable to D if it is stable for some disturbances from D and unstable for others. This formulation holds for both static and dynamical equilibria, it is just that in the static case it applies to a set of state parameters and variables while in the dynamical case it will apply to a set of process parameters and rate variables. The closed set of states a system repeatedly traverses when at equilibrium is its attractor (the marble's rest point is a point attractor, if it circled frictionlessly around that point but up the basin wall it would be a cyclic attractor) and the wider set of states it can transiently pass through while still returning to its attractor is its 'attractor basin' (the bowl provides a literal attractor basin).

[11] Scientists will often speak of stability as itself a constraint. This is potentially confusing since, read literally, the claim is a category error, stabilities are facts characterising a system dynamics in some domain, they cannot of themselves act as constraints within that dynamics. Instead scientists typically intend only the metaphorical claim that unless a system's dynamics is stable it will cease to retain its identity as that system, or perhaps the related metaphor that some system has been selected for its stability under certain conditions, so stability is among the selection constraints; in either case it is just a way of claiming that stability is a materially necessary condition for system existence.

[12] I am using the term 'complicatedness' here, not as a question-begging substitute for 'complexity', but simply as a stop-gap for some specific measure or other of numbers, kinds and connectednesses of attractor basins and landscapes. I doubt there is any one clearly preferable measure and anyway actual measures cannot at present be more than very partially applied (cf. e.g., Wackerbauer et al. 1994).

every inter-landscape shift corresponds to *symmetry-breaking*. The change from conduction to convection in the formation of Bénard cells, e.g., corresponds to the breakdown of previous conductive horizontal symmetry; in consequence the orderedness and constraint coordination increased, delivering an increase in complexity of the system behaviour, and this is typical.

Amplification is where small differences in system state are amplified into large differences in subsequent system trajectory; this is *sensitivity to initial conditions*. Under certain conditions it takes a special form where a *strange attractor* is formed in the attractor landscape in which there is said to be *chaotic motion*. However, the motion remains deterministic and, far from being more disordered than a normal attractor, is best viewed as super-ordered since every point within it may manifest sensitivity to initial conditions. Systems manifesting sensitivity to initial conditions present the problem that small uncertainties (including errors) in knowledge of initial conditions may be amplified into large subsequent uncertainties (and errors) in system state. This yields *limited predictiveness* because system predictability is limited by limited knowledge of initial conditions (even when dynamical form is fully known). How severe a limitation this is in practice, and in what respects, depends on the amplification processes and uncertainties/errors involved. For instance, while prediction that a system's state will remain within a strange attractor is often legitimate, knowledge of location within the attractor can be quickly lost.[13] However, it is also true that systems showing sensitivity to initial conditions can be significantly influenced using only small signals (as perturbations to be amplified). The ideal condition for this behavioural sensitivity is near to the chaotic condition, where there are multiple sensitivities, but not fully chaotic (in a strange attractor) where there are too many of them. This is *edge-of-chaos criticality*, a condition toward which many complex systems move, from sand piles to sensory neural detectors.

*Group 2*: *Constraint altering dynamical properties*. We come next to two properties, already briefly introduced, where the dynamics of constraint change plays a key role in their clearest conception. *Self-organisation* occurs when, through existing constraint coordination, a system bifurcates, sufficiently under its own dynamics, to a form admitting at least one new constraint or (equivalently) new constraint coordination—for instance, in Bénard convection cell formation. There will thus also be new behaviours, sometimes more complex (e.g. Bénard cell formation) and sometimes less complex (e.g. crystal lattice formation). By contrast a reverse bifurcation, where those constraints are instead lost, though also a dynamical landscape transition, would not normally be considered a self-organisation. (It might be considered a self-disorganisation). Since the condensing of molten iron to form a solid iron crystal is also considered self-organisation it is clear that self-organisation has little to do with organisation proper, since an iron crystal is too ordered to be significantly organised (see below).[14]

*Emergence*: When the system outcome of dynamical interaction among system components is surprising or unexpected or too complex to readily understand, scientists are apt to talk about emergent patterns, but this is a vague, shifting and subjective approach. Tightening

[13] But for some of the surprising subtleties surrounding this and like notions, see e.g. Bishop (2011).

[14] All things considered, it would make for clearer communication if 'self-organisation', when used generally, were replaced by 'self-ordering' and 'self-organisation' used according to the definition in the text. Note that what counts as "sufficiently under its own dynamics" to justify the 'self' in self-organisation can be a vexed matter. The pan of heated fluid counts as self-organised because, while the applied heat is key to 'forcing' the dynamical changes, it is applied sufficiently externally to the fluid dynamics. If we deliberately add species to an ecology until it achieves a certain resilience to drought, on the other hand, it would probably not be considered to *self*-organise that dynamical form transition. The system could also not be considered to simply have been organised either, because the outcome may not be increased organisation (though, confusingly, we could say that we 'organised' the change—see note 25). See further Sect. 3 below, note 31 and text, and Hooker (2011d).

it to the appearance of a phenomenon that could not have been predicted from knowing only the pair-wise dynamical interactions of components, is sharper, but still ties the definition of evidently physical properties to a cognitive test (prediction) and, since prediction is so limited, far too much would count as emergent. A better option is (as always) to pursue a dynamical criterion. A clear, wide criterion would be to identify emergence with bifurcation generally, a clear narrower one would be to identify it with just self-organisation. In each case a new dynamical landscape forms or comes into being (intuitively: emerges) through constraint change. On the other hand in the bowl+screw case no new kind of physical work is done; the constraint merely changes its geometric detail. In the self-organisation cases, by contrast, a new existent doing new physical work emerges in consequence of a new constraint emerging.[15] The wider criterion would include some fully explainable cases (e.g. the bowl+screw bifurcation), the narrower criterion would not (see Sect. 3). There do not seem to be any other interesting, clear definitions (cf. Hooker 2011d).

Two further properties can now be characterised. *Path-dependence* is the consequence of positive amplification since then initially nearby dynamical trajectories subsequently diverge as a function of small differences in their initial conditions, so the path taken depends on precisely where the first step began. A notable sub-class of path-dependencies are those where, once begun, development along a certain path itself becomes reinforced, e.g. where an initial fluctuation is amplified and entrenched, especially where that entrenchment involves a bifurcation that makes the development irreversible. Examples include a particular impurity site of first freezing or rolling boiling; a first stage in a developmental trajectory that makes possible a further, perhaps more important, stage; a first genetic mutation that yields a distinctive kind of adaptive advantage; a first oil discovery or shop in a new suburb that transforms a local economy. *Historicity*. These cases also exhibit clear senses of historical possibilities taken up or foregone and their resulting paths are often said to 'fix' their initial historical conditions.[16] By contrast, for stable systems in an attractor basin there is no overall path-dependence since the same outcome occurs (capture by the attractor) for all beginning points (initial conditions) in the basin. Developmental historicity plays an essential role in biology, e.g. whenever there is an earlier stage in a developmental trajectory that makes possible a further, sometimes functionally more important, later stage (Newman 2011a,b). A characteristic historicity to the internal make-up and behaviour of a system, e.g. for an organism, expresses its individuality and so strengthens its sense of self.

And with (relatively) macro constraint formation providing a principled notion of system level,[17] two further properties follow. *Modularity* obtains when system constraint coordination is such that system dynamics can, to a sufficiently good approximation (e.g. to capture essential system functionality), be expressed as an interactive product, the dynamical product

[15] These situations are thus said to show 'top-down' constraints or 'downward' causation, e.g. Bishop (2008), Emmeche et al. (2000), O'Connor and Wong (2002).

[16] However, it would stretch the notion of *physical* constraint to vacuity to call all these initial conditions path constraints, because there is often no cohesive force involved that could ground the constraint.

[17] 'Level' is a loosely (ab)used term. The formation of a new (relatively) macro constraint, however brought about, creates a new level proper in the system, since the constraint now filters out microscopic detail incompatible with it. The iron crystal lattice, e.g., filters out thermal fluctuations and many external perturbations, dissipating their energy as lattice vibrations. (Otherwise the constraint would not be stable against microscopic-originated perturbations and similar external disturbances.) The iron becomes a 2-level system, (1) that below the level of the lattice, the individual ions and electrons, obeying their dynamical interaction laws, and (2) that at the lattice level with its fermi conduction band where electrons stream through the lattice, the lattice collectively vibrates, and so on. This is a dynamically well-defined and grounded notion of 'level', all other uses are for gravitation (e.g. level table) and measurement (e.g. flood level) or are metaphorical (e.g. abstraction level) or confused.

of its intra-modular dynamics and its inter-modular dynamics.[18] Three kinds of modularity can be distinguished, spatial or 'horizontal' modularity at the same macro level (e.g. groups and populations, most buildings and machines), cross-level or 'vertical' modularity (e.g. caste and class social models, business managerial models, scale models of organisms as cells, organs, organism) and process modularity (e.g. models of organisms and complex machines as mechanisms, such as respiration and pulp mill regulation). *Hierarchy* proper is asymmetry of vertical constraint in a sufficiently vertically modular system. It is the exception, the commoner case being mutual constraint both upwards (components constraining their macro level, e.g. ions constraining crystal lattice angles) and downwards (macro level constraining components, e.g. lattice constraining ion vibrations). Modularity reduces system complexity, by decreasing dynamical degrees of freedom, while increasing functional and possibly developmental reliability and ease of repair, but potentially at the risk of removing more subtle but powerful higher order intra-system relationships like multiplexing and multitasking.

*Group 3*: *Global*[19] *functional/organisational constraints*. Here we shift to properties characterised essentially in terms of input/output relationships and only indirectly in terms of the dynamical processes that subserve those relationships. A function F: x −> y takes input x and outputs y; this is as true in mathematics as physics; but in real natural systems, where x, y are real conditions, the transform '−>' must be a real dynamical process that transforms x into y using the basic interaction dynamics and suitable constraints. These constraints, we shall see, can become crucial to the realisation of the function.

Of natural necessity the capacity to carry out functions is central to biology. First, as systems subject to the 2nd law of thermodynamics, they must take in metabolic resources (that is, food and water) and export entropically unusable end-products as wastes; this is their most basic internal functional requirement. Second, as ecological creatures, they must also do this while food-hunting and mating successfully and avoiding predation; these are their most basic ecological functional requirements. Many more functions derive from these basic ones, e.g. respiration as part of metabolism, sensory-motor stimulus-response coordination as part of functioning in an ecology. And of course functionality is not confined to living creatures, almost every artifact is designed to serve some function, from a car engine to a teapot—both of the latter being designed to convert a stored resource (chemical energy, water+tea leaves) to a valuable usable product (mechanical torque at axles, drinkable tea in cups). Another dimension of complexity is the number and subtle interrelatedness of the functions served by a system. The tea pot is a simple device, the car engine much less so.

The car engine holds the clue to the importance of functions for us here. A traditional car engine has many different parts (carbureter, pistons, valves, con-rods, tie-rods,. . .) with each part playing a distinctive and essential role in the whole to bring about the function. It is the coordination of these roles that yields the dynamical process that realises the function of creating output usable torque from input fuel: inserting fuel/air mix, compression, combustion, cam shaft torque, exhausting waste products. Although no constraints appear

[18] Then all system components, at whatever degree of abstraction, are modules, with basic components being those taken to have no internal dynamics and fixed inter-modular dynamical interactions.

[19] 'Global' has been used to mean 'spanning the whole of'; thus a global constraint is one that is applied to the whole system, that spans the whole system, as opposed to constraints that apply more specifically to particular sub-systems or parts. A global constraint is ipso facto a macro constraint, while a macro constraint must have at least a global component to it. Both global and macro constraints, functions, etc. can be relativised to sub-systems or parts of systems and would be called 'relatively global/macro'. Often one wants to say something about all constraints, functions, etc., relatively global/macro or not. For ease of expression, in what follows the relativisations will simply be assumed, except where clarity requires explicitness.

explicitly in this process description, to perform their roles while preserving their exact role interrelationships, requires that the parts operate under a very specific set of constraints, in this case realised in the rigidity of the con-rods etc. and through parts whose only purpose is constraint, including the engine housing and the tie-rods, whose purpose is to lock the whole assembly together. So the global functional constraint of converting fuel to usable torque is itself realised in a carefully interrelated collection of more specific constraints. Without this specific coordination of constraints we cannot dynamically realise the global engine function.

If we turn to biological functional constraints we must find the same. For instance, respiration requires multi-level processes from intra-cellular Krebs Cycles to somatic cardio-vascular provision of oxygen and removal of carbon dioxide, processes that must be made coherent across the entire body and from sub-cellular to organ levels, while also proceeding more or less mutually independently within the coherence requirement. All this requires the intricate coordination of manifold constraints at many levels.[20] A particularly subtle version of these constraints must operate in the special case of intra-cellular function since cellular metabolism manufactures all or most of their required chemicals internally from simpler input components in such a way as to regenerate themselves, including this same self-regenerative capacity. This means that the products of some interactions act as constraints for the production of others in a network that loops on itself to achieve self-reproduction, an organised network of state-dependent, interacting work-constraint cycles (Kaufman 2000). DNA itself is a set of coordinated constraints that catalyse RNA production and RNA similarly catalyses protein production and so throughout cellular biosynthetic pathways. Mobile coordinated constraint sets of these kinds are critical to evolution and development.[21]

Such constraint networks can be anticipated to be structured so as to allow *multiplexing* (many component roles combining to realise a single function) and *multitasking* (the one component playing roles in realising many functions). Multiplexing and multitasking are attractive because they reduce the number of required components while increasing system functionality and adaptability, and possibly evolvability, thereby increasing, while adding additional dimensions to, system complexity (and possibly also increasing system instability

[20] The overall process of respiration is *multi-level*: involving sub-cellular to organism coordination, *multi-dimensional/plexed*: involving organised interactions among many body parameters, *multi-modal/tasked*: involved in many different bodily modes of operation (motor, cognition, stress, etc.), e.g. the cardio-vascular system simultaneously transports resources (oxygen etc.), wastes (carbon dioxide etc.), regulatory hormones, and so on, and *multi-phasic* (asynchronous and non-stationary): respiratory processes occur on many different timescales, with local parameters constantly changing functions of temporary activity while more global respiratory parameters are functions of the longer term developmental and subsequent functional history of the organism. In this conception, slower, more global processes set changing constraints for faster, local processes and their relative mutual independence is achieved if the sets of micro components realising each are largely or wholly separate and those involved in the slower, more global process filter those in the faster, local process from disturbing it, except under specific functional conditions (e.g. to trigger macro actions via macro meta-stabilities). Then global coherence is a result of internal regulation at various, more or less independent, functional levels (intra- and inter- cellular, organ and body). It is processes like these that we need to come to grips with if we are to capture real biological functions, but although the basic interaction dynamics is mostly well known, we are not close to modelling the intricate process and constraint interrelations involved. For further discussion and progress in modelling see, in biology, e.g. Bechtel and Abrahamsen (2011), Newman (2011a,b), and in robotics e.g. Clarke and Proença (2009), Mackworth (2009).

[21] Bird beak development, e.g., is controlled by production of a protein BMP4 that regulates the rate of growth of the underlying mesenchyme (embryonic connective tissue), which in turn forms the species-characteristic upper and lower jaws, its production rate and timing explaining differences in beak shapes from Darwin's finches to chickens and ducks (Newman 2011a). Such mobile morphogens (Turing 1952; Newman and Bhat 2008) offer a powerful way to generate massive variety in form from small alterations in sequence and timing and may prove to be a foundation for embryogenesis (Newman 2011b), one creative enough to confound critics of evolutionary biology on grounds of its explanatory inadequacy (e.g. from so-called intelligent design).

and/or rigidity). Such organisation forces a correlative complex *multi-level spatio-temporal process organisation* (cf. respiration, note 20). Multiplexed, multitasked functions cannot all be realised simultaneously at every location, the resulting interference would render reliable performance impossible. It is necessary then to distribute the realising dynamical activities spatially and temporally so that each local area over each process cycle is restricted to a coherent set of concurrent activities. Moreover, these distributions have to be subtly organised so that each function is realised at convenient locations and times for receiving its inputs and also useful locations and times to contribute its outputs. Currently we often have only sketchy knowledge of how those coordinated constraints are physically realised.

Nonetheless they form the basic organisation characterising all life: *autonomy*, namely the coordination of the internal metabolic interaction cycle and the external environmental interaction cycle so as the latter delivers energy and material components to the organism in a usable form and at the times and locations the former requires to complete its regeneration cycles, including regeneration of the autonomy capacity.[22] And autonomous organisation in turn forms the proper basis for *adaptation*, *adaptiveness, agency* and *learning*.[23]

In sum, it is the specific structuring of coordinated constraints that is the necessary condition for utilising the available basic interaction dynamics so as to produce the dynamical processes that realise global functions and the many more specific sub-functions they in turn require.

That formulation also holds true of the possession of *functionally active boundaries*. At its most general a boundary is simply a division made so that it singles out a unique interior. Mathematics uses the notion in that general, abstract sense, e.g. in boundary value problems. Even within physical systems, some systems have no non-arbitrary physical boundaries, e.g. gravitational systems (sun, solar system, galaxy) since the gravitational force has no shield and no null locations, and this is still more common in social systems, e.g. consider the boundary of a broadcast concert audience in the presence of manifold transmission and playback technologies. This and like complications encourages the standard practice of placing a system's boundary so that it internalises all interactions capable of making a sufficient difference to system behaviour as to be evaluated as significant to the purposes at hand. More narrowly, some physical processes naturally identify boundaries within and of themselves, e.g. fluid boundary layers. While this latter takes us closer to the desired notion, it and all the preceding senses of boundary are too weak for the present notion, which is that

[22] On autonomy see Christensen and Hooker (2000a, 2002), Bechtel (2007), Moreno (2007), Hooker (2011b), Moreno et al. (2011), and references. Self-governance lies at the core of our commonsense conception of autonomy. However, we are most familiar with the idea of autonomy as applied to persons and political governance, but these are sophisticated notions applied to sophisticated systems whose trappings may distract from fundamentals. We need to return to basic principles operating in all living systems to construct a naturalist notion that will 'grade up' across the evolutionary sequence to our sophisticated concept.

[23] An organism is adapted when it possesses an autonomy-satisfying set of traits in its life-environment. Conversely, an organism's ecological niche is comprised of the range of life-environments for which its traits provide satisfaction of autonomy. An organism's adaptiveness or adaptability is its capacity to alter its specific traits in mutually coordinated ways so as to adapt to, that is, satisfy autonomy in, a wider range of life-environments. Learning, understood most generally (e.g. for populations and ecologies), is the application of adaptability to develop adaptations; understood more narrowly (e.g. for individuals), it is this process manifest through internal sensory, memory and motor regulation rather than inherited reaction to stimuli, that is, through neurally modulated behaviour. It can be instructive to inquire how various community groups, from business firms to cities to nations, learn. See Christensen and Hooker (2000a,b), Farrell and Hooker (2007a,b, 2009), Hooker (2009). Entities with a distinctive wholeness, individuality and perspective in the world, whose activities are willful, anticipative, deliberate, adaptive and normatively self-evaluated, are properly treated as genuine agents; and when their internal states and dispositions unfold historically, they are properly treated as full individuals; autonomous systems are inherently all of those things. See Hooker (2009), Skewes and Hooker (2009).

of a physical layer having functional roles necessary for system functioning. Thus while a crowd self-assembled at a free concert has a natural boundary (current behaviour influenced by direct sensory interaction with staged behaviour), the boundary carries out no function necessary for the concert itself; by contrast, a ticketed concert in a concert hall has as its boundary the physical barrier of the concert hall, forcing the interactive crowd boundary into conformity with it, but this boundary also has the necessary economic function of supporting the performance by discriminating audience access on the basis of payment.

The basic point of functionally active boundaries (and of some others) is to sustain an asymmetry (thermodynamic and organisational) between interior and exterior, by excluding some class of stimuli perturbing to internal system functioning (e.g. carapaces excluding blows, epidermises excluding toxins) and in more complex cases by gating the system (selectively admitting and expelling various material parts of the system, e.g. food in/ wastes out). In the simpler exclusion-only cases the boundary can be used to cleanly identify the system in space and time as what is physically internal to it. But it is surprising how quickly gating sets in; e.g. even the engine casing not only performs several different exclusionary functions (gas containment, vibration and shock resistence) but also fuel and exhaust gating. And once gating occurs identity typically cannot so simply be expressed. In the ticketed concert hall, e.g., audience and players may be let out (e.g. to toilet) and return yet remain part of the concert, even play key roles in it.

The ticketing barrier and engine casing gates roughly correspond to the ion gates a single cell has in its walls that selectively intake nutrients and expel wastes, but while there are only a few essential functions played by the former barriers, there are many, many more functional roles played by the cell membrane in maintaining cellular functioning. To this end the cell membrane contains many integral membrane proteins that serve specialised roles in membrane functioning, representing a huge increase in its functionality but also in its internal boundary membrane organisation to sustain those functions. In addition, the boundary itself has to be regenerated as part of the intra-cellular functions that boundary itself helps to facilitate. In biology functional boundaries become ubiquitous, diverse and complicated, ranging from those for chromosomes (nuclear membranes) and cells (membranes and walls), through those for tissues and organs (mesenteries and connective tissues), to those for multi-cellular organisms (epidermises), and finally(?) those for multi-organism groupings, e.g. bacterial colonies, slime moulds, termite nests, human clubs, cities and territories. Across this range boundaries vary greatly in their structure and properties, and also vary increasingly within each category (e.g. the large differences between moulds, fungi and animals within multicellular organisms, and between clubs, cities and territories).

Boundaries represent constraints. The point of the theatre boundary is to constrain entry of persons and exit of sights and sounds and structure communication to others in these terms. These features are true of all boundaries, even permeable ones. (Otherwise there is no identifiable boundary function.) However, in simple cases like the theatre or engine casing this boundary subsists in itself, its strong internal constraints allowing it to perform its functions largely in passive independence from the goings on in its interior. This has been typical of engineered boundaries. But now there is an increasing move toward more interactive boundaries, windows that tint or have covers that close in response to sunlight, doors that open automatically to select ID-scanned entrants, etc., and soon membranes that respond electrically, thermally and colourfully to various internal signals, and so on. This means that constraints themselves have to shift as a function of other processes occurring in the system, and thus that the system has to do work on its constraints. (Neither is a surprise, in highly and rapidly adaptive systems like the brain we face subtly interrelated constraint organisation that shifts on time scales from seconds to decades—cf. Christensen and Hooker (2000b),

note 14 and text.) The point here is that as the boundary takes on increased functionality its internal constraints have to become both more organised and more subtly interconnected to processes and constraints internal to the system (but external to the boundary) for both functional responsiveness and energy. By the time we add the biological requirement that the internal system processes and constraints regenerate the boundary, including its constraint structure, we arrive at a very subtly organised articulation of constraints indeed.

This completes the review of fundamental complex systems properties and it makes the case for the central roles of constraints in characterising them. It also demonstrates the diversity of properties characteristic of complex systems and the internal complexity of many of them. This shows the insufficiency of any of the commoner simple ideas of complexity—number of degrees of freedom (component numbers), algorithmic incompressibility, levels,. . .—to capture the notion of complexity by themselves. To approach a characterisation, first omit all epistemic notions, e.g. those appealing to intelligibility and surprise, and all 'external' notions like controllability, as ultimately derivative considerations. Then at the least complexity looks to have five quasi-independent dimensions to it: cardinality (component numbers), non-linearity (of interaction dynamics), intermediate orderedness (algorithmic compressibility), nested organisation (organisational depth) and global organisation. I do not know of any one measure that can capture all these dimensions and I do not think anything less rich is adequate to the forms complexity can take. I do suggest that insight into the nature and forms of constraints will be important to gaining any deeper insight into the matter.

## 3 Challenges (A Tentative Exploration)

The explosion of complex systems concepts, models and methods across all the sciences and public policy over the past 30 years poses many theoretical and practical challenges to contemporary practices. The first science-wide review of the foundational challenges is found in Hooker's *Philosophy of Complex Systems*, Hooker (2011a), a collection of 27 essays by leading scientists and philosophers spanning the sciences and public policy; readers are directed there for further exploration. In this section three of these challenges, deriving importantly from the roles of constraints, will be briefly discussed to illustrate the range and importance of constraint-driven challenges.[24]

### 3.1 Organisation

Global constraints pose two distinctive sorts of challenge to conceptualising complexity in modern science: they introduce the concept of organisation as a key aspect of complexity, a concept that currently has no satisfying measure, and they challenge the representation of global constraints in mathematical dynamics.

The internal articulation of distinct components so as to be globally functionally coherent, found in the motor vehicle engine, provides a paradigm of what is needed to capture organisation proper.[25] Machines and living things are organised because their parts are relatively unique and each part plays distinctive and essential roles in the whole. In this sense, each global functional constraint requires an underlying organisation of dynamical constraints and processes to realise them. The set of coordinated constraints required to realise organisation

[24] The first and last discussions that follow are adapted, with modification and improvement, from Hooker (2011b,d) and the reader is referred there for further discussion.

[25] There is a wider colloquial usage in which being organised merely means being appropriately prepared, whether or not that preparation involves any significant organisation in the narrower sense in the text.

is less ordered than those of a crystal, which is simply and uniformly ordered, but much more ordered than the null coordination within a gas. Thus orderedness, measured by (the inverse of) algorithmic complexity, cannot capture organisation. We approach more closely the notion of organisation by considering the order of the relationships within the constraint coordination, where a 2nd order relation is a relation among relations, and so on. Then an organisation is characterised by (relatively) high order relations that involve many nestings of relations within relations, that is, of nested correlations within correlations. For example, a car engines pistons and valves are internally correlated and these correlation relations are nested within correlations between cylinder positions and the fuel injection system, etc. A system's organisational depth is measured by the degree of nesting of sub-ordering relations within its global ordering relation (cf. cells within organs within bodies within communities). Living systems especially are deeply organised. However, organisational depth also does not fully capture complexity.[26] Thus the coordination of constraints that marks global functional organisation is somewhat subtler than we can at present fully characterise.

Neither can we properly represent it mathematically. Dynamics is presently represented mathematically as a differentiable flow on a structured manifold, e.g. Classical mechanics can be represented as a flow or field of trajectories in space determined by the system Lagrangian, a representation of energy, the motion represented by differential equations. These modelling resources, powerful though they are for modelling the energetics of processes, do not explicitly describe the physical organisation of the system. For instance, a metabolic cycle and a pendulum may be modelled as equivalent dynamical oscillators. In a phase space only the global dynamical states and their time evolution along a system trajectory of the overall flow, is specified, not the organised processes that produce the dynamics. Note that any differential equation already compresses constraint and interaction information into a flow and in that sense suppresses the explicit details of the interactions and constraints. The globalness and organisation of the constraints then pose a further, particular problem to representation. And the reverse engineering problem of specifying organisation (as distinct from order) from dynamics is currently unsolvable.[27] In this sense our current fundamental representation of dynamics cannot capture organised constraint coordination, and hence also not the mechanisms that realise them.[28]

Recall from earlier discussion that organisation of constraints (i) stands at the heart of all biological entities, (ii) extends to the subtle organisation underpinning functional fluency and

[26] Because, e.g., it does not capture the distinctiveness of nested relations (cf. note 28) and the top-down constraints that modulate them and, in dropping the algorithmic conception, it loses 'horizontal' relational complexity. Gell-Mann Gell-Mann (1994) discusses effective complexity and logical depth (see Bennett 1985, 1992) and Type 2 theories (Marr 1982), as other possibilities for measuring organised complexity, but neither is satisfactory for various reasons he notices—fundamentally for the above reasons. For general discussion of these issues see Collier and Hooker (1999), sections III and VI.

[27] Current reverse engineering programmes presume simple system structures that do not include global constraints and organisation, cf. Bongard and Lipson (2007), Tegnér et al. (2003). This is the counterpart to the severe constraints on programmes for the extraction of causal order in systems, that presume that not even simple feedback loops are present (cf. Shalizi 2006, 2.2.1). However work continues on increasing the discriminatory capacities of reverse engineering—for an interesting development see Schmidt and Lipson (2009)—and it is too soon to pronounce on its limits.

[28] It will of course often be possible to make scientific headway by identifying organised biochemical mechanisms that contribute to realising such constraints, with the Krebs cycle a powerful case in point (see Bechtel and Abrahamsen 2011). And since each individual interaction is derived from basic physical interactions, each can be represented dynamically and the set of them can be modelled as a set of coupled integro-differential equations. But what cannot be represented dynamically, it seems, is their particular nested cyclical organisation; the coupled equations represent any organised collection of interactions whatsoever so long as its net connectivities are represented by these couplings.

effectiveness, e.g. in multi-tasking and multi-plexing, and (iii) underlies virtually all social and much ecological organisation. Thus these limitations form a major foundational challenge to the science of complex systems. Moreover, that challenge extends in concrete form to robotics and other so-called intelligent systems: since none of these as yet has a substantial autonomous organisational basis for its capacities, its intelligences, however skilled in specialised ways they may be, remain unintegrated into their root agency capacities, precluding them from solving the fundamental class of open problems and from open cognitive development more generally.[29]

### 3.2 Boundaries

The challenge here is to general ignorance: I have been able to locate little general knowledge, either theoretical or practical, of when functionally active boundaries are necessary, what their functional roles should be, and of how their constraints and dynamical interactions should be organised to achieve their roles. Despite a slow increase in the interactiveness of some engineered boundaries, we have not yet seen boundaries in engineering of the biological sort and likely will not see them until bioengineering, barely begun, matures. (Most or all current engineering experiments in self-assembly or regenerative re-assembly of which I am aware don't include any significant boundary layer.) However it is the biological cases of organised coordination of boundary constraints in the service of functionality we need to understand if we are to understand the possibilities for complex system function that boundaries offer and biology already exploits. Such understanding is also required to appreciate when, why and how super-systems can form from interacting systems, for instance the differences between the agglomeration and fruiting stages of slime mould assembly, or between interpenetrating DNA-sharing primitive plants and modern plants, or between an ancient walled city and a modern city, and so on.

Conversely, we need to understand when, why and how the biological boundaries we do possess as mammals permit us to construct social organisations with somewhat more open boundaries. Modern cities and regions, e.g., can show significant autonomy, yet have boundaries thoroughly open to all kinds of traffic across them. Threats produce toughened boundaries in response, from quarantining SARS victims (Singapore) to battlefield barriers, but modern cities and regions are in general far more permeable than were ancient walled cities and closed colonies. These modern, more open structures dictate that autonomous systems

[29] See references, note 23. Robotics uses a very limited formal notion of autonomy (something like invariant dynamical form) and limited performance criteria (typically confined to a single task) and an equally limited satisfaction method. There has recently emerged an embodied functionality movement within robotics (see e.g. Nolfi 2011; Pfeiffer and Bongard 2007) where cognitive organisation is strongly shaped by the dynamics of body and environment, in ways that you would expect from an autonomy, interactive perspective. This represents a vast improvement over the computer-on-a-machine approach that had previously dominated. However it is as yet very far from even incorporating normative signals into the body coherence of robots, let alone the complexity required for self-regeneration and the capacity for fluid management of multi-dimensional environmental and internal interaction processes in relation to that. While studies such as that by Nolfi (above) have made progress on the fluid management of environmental interaction, these are still primitive devices when it comes to management of activity in relation to norm-derived goals. The problem in artificial life is still further from solution, since formal reproduction is not regenerative and is not the core of metabolism and thus not the key to metabolism-based action norms. See Moreno and Ruiz-Mirazo (1999); Moreno et al. (2011), cf. Christensen and Hooker (2002, 2004). There is an associated need to bring work on self-assembling, self-repairing robots (e.g. Groß and Dorigo 2007, 2008, http://www.swarmanoid.org/index.php) into relation with attempts to develop artificial autonomous systems where modelling even very elementary cells that are dynamically stable and thermodynamically coherent is proving difficult (e.g. Gánti 2003; Ruiz-mirazo and Moreno 2004; Szathmáry 2005; Barandiaran and Ruiz-mirazo 2008, http://www.ees.lanl.gov/protocells).

are ultimately identified, not by having a specific continuous physical boundary layer, but with the scope of their autonomous self-regulation, that is, with the scope of their organised constraints. Inside = inside the scope of self-regulatory constraint, this is the locus of the key asymmetry between system and environment and what is common across all biological systems. Generalising, the 'self' of self-organisation is what is within the scope of the relevant constraint-forming dynamics.[30] The switch from grounding identity and internality on boundary possession then opens up exploration of the kinds of boundaries that may facilitate autonomous self-regulation in various circumstances. We need to understand where and how autonomy might be expressed within porous boundaries, or perhaps relinquished because reliance on prior functional boundaries can be made to suffice.[31]

Few researchers seem to have addressed these issues. Among the few is Alan Raynor in his provocative *Degrees of Freedom: Living in Dynamic Boundaries*, Raynor (1977).[32] He contends that the general function of boundaries is to co-adapt internal and external environments, and perhaps the boundary itself also along the way (the cell lipid membrane preceded the cell) and that to this end there are 3 universal functional properties of boundaries: permeability, deformability and continuity, that may be realised in varying degrees, sometimes across many intra-boundary layers. This much is not nothing, but may be all the general boundary theory that can currently be achieved. At the least it reminds us of two important factors, internal/external co-evolution and boundary evolution, that should be included in a proper dynamical conception of evolution. But considering the ubiquity of boundaries and the crucial roles they often play in system dynamics—by applying various forms of constraint—we must hope for more general understanding of this subtlest of applications of organised constraints if we are to develop general foundations for a theory of complex systems and valid methods to scientifically investigate them.

30 This in turn provides the foundation for the principle that all systems processes must ultimately be enabled by, and only by, interactions and constraints available internally to the system. See e.g. Bickhard (1993), Christensen and Hooker (2000b), note 4 and text for one among several earlier versions by others, including Newman (1970) (Newman 2011a,b—private communication). This principle removes evolutionary views of proper function, source-based constructs of signal meaning, and so on.

31 Even physiologically, multi-cellular creatures have wider commerce through their epidermic boundaries than do simpler creatures; humans, e.g., do not make nine essential amino acids but import them through eating plants. This issue arises within a larger trend associated with multicellular organisms. Overall, the effect of multi-cellular evolution has been to expand the capacity for interaction with the environment, including both anticipating environmental courses of action and acting to modify the environment to shape its selection pressures. Multi-cellulars differ in at least three important respects from single cells: they have (i) increased substitution of environmental construction for internal construction (e.g. carnivores intake complex molecules, humans rely on environmental production of many essential amino acids), (ii) increased self-regulation of their food acquisition and damage avoidance (e.g. rapid or prolonged migration to track food resources, hiding or hole construction to escape predators) and (iii) increased capacity to self-regulate the modification of metabolism to suit both temporary activity (e.g. heart rate and blood re-direction for running) and permanent change (e.g. callousing, neuro-muscular compensation for injury). Underlying these is a fourth, more basic, way in which they differ: (iv) they have acquired the capacity to communally regulate the birth, specialisation and death of their members (cells). While in neurally more complex species they can show a myriad of forms, every viable community, including human communities, must acquire some form of these latter capacities. (Thanks to Alvaro Moreno for this insight.) Over the last century as human societies have become more developed, they seem to be experimenting with internal regulation of birth (aka the demographic transition), and specialisation (aka education) while decreasing regulation of death (no death penalty, voluntary euthanasia, + some regulation of war). Such shifts require increasing internal self-regulatory capacities.

32 Discussions with Raynor, and reading his work, form an important part of my appreciation of dynamic boundaries, and I thank him for them. But there are also significant ways in which I depart from his 'boundary-centric' approach, perhaps best illustrated in note 31 and text.

### 3.3 Dynamical Representation

Is every behaviour of complex systems equally representable within classical dynamics? The argument from basic interaction answers yes: all of the basic interactions among components, including the constraint forming components, are from classical dynamics (gravity and electro-magnetism) so the behaviour of the systems themselves must be representable in classical dynamics.[33] In fact, the answer is no; the argument from basic interaction is invalid and its failure turns on the kinds of constraints involved. This failure of dynamical representation constitutes the most fundamental challenge of complex systems to standard dynamics.

It is not hard to locate the fault line in the argument from basic interaction: it assumes that once the basic interactions are given, all other interactions will have thereby been fully captured. But some complex systems defeat this assumption. The difficulties come in two forms.

*First*, there is the problem of the representation of integrated multi-phasic systems. In Sect. 1 it was pointed out that most biological systems have simultaneously operating constraints of many different kinds, mechanical and chemical, solid and liquid, and so forth, operating in an equal diversity of ways: quasi-equilibrium, non-equilibrium oscillatory, phasic, etc. How is this diversity to be captured within an integrated dynamical representation? Solid and liquid models, e.g., are very different in form, any direct conjunction of the two will simply produce a contradictory model. Sometimes a hybrid model can be constructed by cleverly shaping a coherent amalgam of a very few selected principles from each that nonetheless capture the phenomenal domain of interest, but this is rare and certainly not to be presumed. Combining mechanical and chemical processes faces similar (if perhaps less severe) problems. In fact such problems can arise within a single kind of dynamical domain. In the case of constrained viscous fluid flow, e.g. through a pipe, the dynamical determinants of motion near a wall are very different to those near the centre and each requires a distinctive model incompatible with that for the other (Rueger 2005). These dynamical representations cannot be unified. For systems like organisms and cities there is no workable option but to investigate partial aspects of them one at a time, often of necessity using knowingly idealised models, e.g. constrained focal-process + process-perturbation and constrained focal-level plus cross-level interaction models.[34] Were this the only integration problem it could be seen as an unavoidable, but ultimately pragmatic, consequence of the limits of mathematical representation, at least in the hands of finite agents; but it is not the only problem.

*Second*, for a certain class of constraints, there is a further, deeper problem for dynamical representation. To understand what is involved it is necessary to return to the fundamentals of classical dynamics. There is a generalised analytical framework for classical dynamical analysis – the Lagrangian/Hamiltonian formalism – that directly or approximately covers a wide range of cases and serves as the core of analytical classical dynamics.[35] In this scheme, recall, basic interaction dynamics and constraint forces combine to produce the system dynamics.

[33] In fact the challenge to this conclusion will not depend on whether the basic interactions are classical, relativistic or quantal, while most examples are classical.

[34] The unified representation problem arises when each model is a degenerate idealisation of the full interaction dynamical flow equations—that is, one that collapses out structure that cannot be regained by any subsequent conjunction of additional detail, as with the two fluid flow approximation models. The one system may have many different, mutually incompatible, degenerately idealised dynamical models (Hooker 2011c, on degenerate idealisation see also Hooker 1994, 2004). However, where the full flow equations are not workable, in principle or in practice, there is no choice but to use degenerately idealised models, and this applies still more forcefully to biological systems.

[35] For a nice introduction see Butterfield (2004a,b). Classic texts here are Goldstein (1950), Arnold (1978). Butterfield (2004a), note 6 offers a brief guide to others. Bloch (2003) offers an introduction to the treatment

External constraints typically introduce unknown forces into the dynamics, so that a determinate Newtonian dynamics cannot be specified, and they result in interdependencies among the intrinsic dynamical variables that have to be accommodated, so that an unambiguous representation of the dynamical possibilities cannot be formulated. To resolve these problems it is sufficient, and arguably necessary, to restrict consideration to those systems where the external constraint forces act orthogonally to all allowed system motions,[36] so that the system does no work against external constraints (constraint force orthogonality). This defines a constraint (hyper) surface in the system configuration space to which the constraint forces are everywhere perpendicular (orthogonal). Thus motion on this surface is effectively constraint-free – this is expressed in D'Alembert's principle. If in addition the external constraints are holonomic – literally: express a whole or single law[37] – then the system dynamics may be re-formulated on their D'Alembertian constraint surface in terms of new generalised, independent variables, the dynamics now having the form of a free (unconstrained) system. Lagrange equations of motion can then be formulated for the system. This resolves the variable interdependency problem introduced by constraints. The method of Lagrange multipliers then formulates D'Alembert's extremal least action principle and permits solving the system dynamics (that is, specifying the action geodesics) on the constraint surface, without knowing the external constraint forces. Rather, once the dynamics is known, the external constraint forces can be reconstructed as the forces they need to be to maintain the external constraints during the system motion. This resolves the problem of their being initially unknown.

Many complex systems do work on their constraints, physically altering them over time. Examples include (i) a river altering its own banks, an accumulative process where the current constraints (banks) are a function of the history of past flows (currents), (ii) intra-cellular biochemical reaction processes where molecular structures constraining some processes are the products of other processes and vice versa; (iii) any self-organisation where the constraint formed becomes an external constraint for subsequent processes (Bénard cell and iron bar formation, etc.). In all these systems constraint orthogonality fails. With this failure the most basic precondition for achieving the core analytic construction fails. There is then no general, analytical mathematical formalism available for dynamical behaviour. The incapacity to construct conditions where D'Alembert's principle holds undermines the applicability of the very variational apparatus that we take to underlie all fundamental dynamics. In this way,

Footnote 35 continued
of D'Alembertian but non-holonomic constraints, see also Flannery (2005). While, strictly, Newton's Laws themselves can in principle be applied to any system, with holonomic or non-holonomic, conservative or non-conservative constraints, the Lagrangian analytical apparatus not only captures the core cases and provides powerful analytical tools for analysing dynamics, it is also the basis for treating continua and force fields, especially the electromagnetic field, and its founding variational principles form the basis for the successful generalisations of mechanics to relativistic and quantum mechanical forms.

36 These are the 'virtual' displacements of a system, as opposed to actual displacements over some small time interval occurring under the influence of the intrinsic forces as well.

37 Holonomic constraints may be written as some function of the space-time geometry in which the system moves (note 9). Specifically, they satisfy an equation of the form $f(r_1, r_2, \ldots r_n, t) = 0$, where the $r_i$ are system coordinates and t is time. This expresses the effect of the constraint forces while not specifying the forces themselves. (The forces are often known only after the main problem is solved.) While smooth (frictionless) sliding under gravity on a sloping plane is a case of holonomic constraint, a spherical bead rolling smoothly on the outside of a cylinder is not because the constraint alters its basic character when the bead falls off. Essentially, for the constraints to be holonomic means that they may be expressed purely geometrically, so that they are independent of the behaviour of the system. Independence fails in the case of the bead on the cylinder, there is a change of constraints at a space-time location determined by the bead's motion. (Note that the reverse relation does not hold, e.g. though independent of system behaviour, containment walls do not form holonomic constraints.)

complex systems challenge the reach of our deepest analytical understanding of dynamics and thus present a fundamental dilemma about how to approach dynamics: retain the present approach and exclude complex systems or search for some new, more generous foundations for dynamics.[38]

The response of those favouring the argument from basic interaction (also the textbook response, when there is one) is to argue as follows: the system can always be expanded so as to include the matter that is the source of the constraint forces and when that is done we discover only further basic interactions of the usual kinds, so the enlarged system undergoes standard constraint-free motion and the problem vanishes. If we cannot solve these systems then it is simply because there are too many components involved, a pragmatic rather than a principled difficulty. This appears persuasive at first blush (e.g. it satisfied no less than Goldstein 1950, p.14), constraints really do arise from basic interactions, otherwise we could not understand dynamically their transformation by the system. But we should look further. Recall that the argument as it stands is at least incomplete: the leap to the conclusion again requires assuming that once the basic interactions are given, all other interactions will have thereby been fully captured, begging the point at issue. It should be considered that in all the cases where there are dynamical constraints that do have work done on them in the course of the dynamics this work can be objectively measured, so it cannot simply be claimed that a constraint-free representation is always available, it has to be shown.

As the point was put in Sect. 1 discussing Bénard cell formation, the constraints don't disappear, they must re-appear internally. Nor therefore is it enough to point out that the work done on constraints must ultimately itself be constituted in basic dynamical processes; this is true, but not the issue. The issue is the reality of supra-basic dynamical formations that can filter other motions.[39] The reality of these is itself guaranteed by the constraint-forming basic interactions, wherever a suitable invariant level suffices explanatorily.[40] Those who deny such compound dynamical entities owe it to us to show how to both faithfully represent the internal presence of the constraints—at all, and if admitted, show the work done on and

[38] There is also a whiff of scandal here as well, namely the scandal, unfortunately increasingly common, of dynamics textbooks simply ignoring these deep problems, or implying that there is only a pragmatic issue of mathematical resources involved.

[39] Bishop expresses this as follows: The properties and behaviors of a system at a particular level (including its laws) offer necessary but not sufficient conditions for the properties and behaviors at a higher level. He calls this contextual emergence. (See Bishop 2005, 2008, 2011). This is apt, since it points to an irreducible wholeness (emergence) under specific dynamical conditions. Hooker (2004) makes the complementary point that each specific reduction of function to dynamics in complex systems requires a dynamically specified context, so that we also only have contextual reduction—and, since the context must specify appropriate constraints, it must include dynamically irreducible features, so that reduction and emergence are intertwined.

[40] Bénard cell formation involves only behavioural change (in a fluid), and so is clearly a supra-basic dynamical formation over fluid elements, a dynamical wholeness, not any new fundamental kind of entity. However typical phase transitions show qualitative as well as behavioural changes, e.g., from solid to fluid; but these too are representable as supra-basic dynamical formations if there is a component level at which the components are invariant through the transition, e.g. molecules for most phase transitions. This idea then extends to those transitions that also involve the erstwhile basic components also changing qualitatively, e.g. undergoing chemical change, becoming ionised (in transitions to plasma) or undergoing nuclear change (e.g. in gravitational condensations), so long as the level of (respectively) atoms, electrons, and nucleons suffices explanatorily (other than for Lagrangian descriptions of the transitions). If so, all these changes can be understood as purely dynamical formations over invariant basic components under constraint changes. If not, then we have to contemplate a mechanics of basic transformation, perhaps equipped with quantum-like creation and annihilation processes, something also beyond Lagrange theory. Thus getting the dynamical specifications right is what is ultimately important. Thanks to Gil Costa Santos now, and Parker English in 1973, for pointing to these telling metaphysical issues.

by them—and still achieve an analytical Lagrangian model.[41] This is prima facie not possible. Constraint forming systems, like the iron bar from cooling, create similar challenges to Lagrangian representation. All these challenges may be restated as follows: since precisely in all such processes the system changes dynamical form, hence would change Lagrangian form, it is unclear how the Lagrangian apparatus in itself could accommodate that requirement. In short, it is difficult to see how systems showing these phenomena could be reduced to presenting merely pragmatic barriers to standard knowledge of solutions.[42]

**Acknowledgments** Discussions with Robert Bishop, Thomas Brinsmead and Stuart Newman of an earlier draft enriched the presentation at several points; their time and insightfulness are appreciated. Blemishes of all kinds remain my own.

## References

Arnold, V. I. (1978). *Mathematical methods of classical mechanics*. Berlin: Springer.

Barandiaran, X., & Ruiz-mirazo, K. (Eds.) (2008). Modelling autonomy: Simulating the essence of life and cognition. *BioSystems, 91*(2).

Bechtel, W. (2007). Biological mechanisms: Organised to maintain autonomy. In F. Boogerd, F. Bruggeman, J.-H. Hofmeyr, & H. V. Westerhoff (Eds.), *Systems biology: Philosophical foundations* (pp. 269–302). Amsterdam: Elsevier.

Bechtel, W., & Abrahamsen, A. (2011). Complex biological mechanisms: Cyclic, oscillatory and autonomous. In C. A. Hooker (Ed.) Philosophy of complex systems. *Handbook of the Philosophy of Science* (Vol. 10, pp. 259–288). Amsterdam: North Holland/Elsevier.

Bennett, C. (1985). Dissipation, information, computational complexity and the definition of organization. In D. Pines, *Emerging syntheses in science, proceedings of the founding workshops of the santa fe institute*. Redwood California: Addison Wesley.

Bennett, C. (1992). Logical depth and other algorithmically defined properties of finite objects. In *Proceedings, workshop on physics and computation*, IEEE, pp. 75–77.

Bickhard, M. H. (1993). Representational content in humans and machines. *Experimental and Theoretical Artificial Intelligence, 5*, 285–333.

[41] Of course we can construct little Newtonian models of work done on and by some specific constraints, e.g. of flowing water eroding a bank, and gesture at others. This too is true but not germane. The requirement is to show how to coherently incorporate these into a single Lagrangian model of the whole system, as the pragmatic claims presume. Nor is it germane to point out that D'Alembert's principle requires only that the *net* work done in virtual displacements is zero (e.g. Subhankar and Shamanna 2006) and that in a basic dynamical model of system+constraints this is satisfied if Newton's Third Law operates to ensure balanced forces by the system on the constraints and by the constraints on the system, since the point is to capture the internal dynamical reality of the filtering constraints. Similarly, we can try to approximate each real constraint as a time-slice sequence of holonomc constraints for purposes of analysis, but this is again just to concede the point at issue. Finally, controlling systems with non-holonomic constraints does not by itself count either, since this can be achieved without needing a full analytical model of the dynamics (e.g. through local curve tracing, cf. Bloch 2003).

[42] There are further technical issues listed as outstanding in Hooker (2011d), most importantly the extent to which the applicability of the analytical Lagrangian model can be pushed into the domain of non-holonomic constraints. The extension can be made specifically to semi-holonomic and exact linear constraints Flannery (2005) and those meeting various other convenient (but not necessarily well-grounded) conditions, like the Chetaev condition, but this did not leave a clear picture of the situation. Recently Flannery (2011a), in a helpfully clear historical review (cf. Soltakhanov et al. 2009) and new generalised analysis, has succeeded in extending the analytical Lagrangian model to all ideal non-holonomic constraints, i.e. to all those constraints that act orthogonally to the motions and so have no work done on them during motion. (See also Flannery 2011b). For the reasons noted in the text, it remains that this analysis cannot be extended to non-ideal constraints, i.e. to systems that form constraints or do work on their constraints. Another outstanding issue was when and why the consequent extremal action paths (all Hamilton's Principle requires) are least action paths. This has been interestingly settled for one class of cases in Gray and Taylor (2007), but cf. Dewar et al. (2011) for some of the complications involved elsewhere.

Bishop, R. C. (2005). Patching physics and chemistry together. *Philosophy of Science, 72*, 710–722.
Bishop, R. C. (2008). Downward causation in fluid convection. *Synthese, 160*, 229–248.
Bishop, R. C. (2011). Metaphysical and epistemological issues in complex systems. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 107–138). Amsterdam: North Holland/Elsevier.
Bloch, A. M. (2003). *Nonholonomic mechanics and control*. Springer, New York, See http://www.cds.caltech.edu/mechanics_and_control/about_this_book/ for corrections and additions, incl. internet supplement.
Bongard, J., & Lipson, H. (2007). Automated reverse engineering of non-linear dynamical systems. *Proceedings of the National Academy of Science, 104*(24), 9943–9948.
Butterfield, J. (2004a) Between laws and models: Some philosophical morals of Lagrangian mechanics. Available at http://philsci-archive.pitt.edu/archive/00001937/. Accessed Apr 2009.
Butterfield, J. (2004b). On Hamilton-Jacobi theory as a classical root of quantum theory. In A. C. Elitzur, S. Dolev, & N. Kolenda (Eds.), *Quo Vadis quantum mechanics? Possible developments in quantum theory in the 21st century*. New York, Springer. Available at http://philsci-archive.pitt.edu/archive/00001193/. Accessed Apr 2009.
Christensen, W. D., & Hooker, C. A. (2000a). Organised interactive construction: The nature of autonomy and the emergence of intelligence. In: A. Etxebberia, A. Moreno, & J. Umerez (Eds.), *The contribution of artificial life and the sciences of complexity to the understanding of autonomous systems. Communication and Cognition*, Special Edn (Vol. 17, No. 3–4, pp. 133–158).
Christensen, W., & Hooker, C. A. (2000b). An interactivist-constructivist approach to intelligence: Self-directed anticipative learning. *Philosophical Psychology, 13*, 5–45.
Christensen, W. D., & Hooker, C. A. (2002). Self-directed agents. In J. MacIntosh (Ed.), *Naturalism, evolution and intentionality. Ottawa, Canadian Journal of Philosophy*, Special Supplementary Vol. 27, pp. 19–52.
Christensen, W. D., & Hooker, C. A. (2004). Representation and the meaning of life. In H. Clapin, P. Staines, & P. Slezak (Eds.), *Representation in mind: New approaches to mental representation* (pp. 41–69). Sydney: Elsevier.
Clarke, D., & Proença, J. (2009). Coordination via interaction constraints I: Local logic. In F. Bonchi, D. Grohmann, & P. Spoletini, E. Tuosto (Eds.), *ICE'09 structured interactions*, EPTCS, 12, pp. 17–39. doi:10.4204/EPTCS.12.2.
Coleman, R. A., & Korté, H. (1999). Geometry and forces in relativistic and pre-relativistic theories. *Foundations of Physics Letters, 12*(2), 147–163. doi:10.1023/A:1021609022945.
Collier, J. D., & Hooker, C. A. (1999). Complex organised dynamical systems. *Open Systems and Information Dynamics, 6*, 241–302.
Dewar, R. L., Hudson, S. R., & Gibson, A. M. Action-gradient-minimizing pseudo-orbits and almost-invariant tori (preprint, submitted).
Emmeche, C., Koppe, S., & Stjernfelt, F. (2000). Levels, emergence, and three versions of downward causation. In P. Anderson, C. Emmeche, N. Finnemann, & P. Christiansen (Eds.), *Downward causation: Minds, bodies, and matter* (pp. 13–34). Aarhus: Aarhus University Press.
Farrell, R., & Hooker, C. A. (2007a). Applying self-directed anticipative learning to science: agency and the interactive exploration of possibility space in ape language research. *Perspectives on Science, 15*(1), 86–123.
Farrell, R., & Hooker, C. A. (2007b). Applying self-directed anticipative learning to science: Learning how to learn across revolutions. *Perspectives on Science, 15*(2), 220–253.
Farrell, R, & Hooker, C. A. (2009). Applying self-directed anticipative learning to science: identifying error and severe testing through interactive exploration of possibility. *Foundations of Science, 14*(4), 249–271.
Flannery, M. R. (2005). The enigma of nonholonomic constraints. *American Journal of Physics, 73*(3), 265–272. Available at www.physics.gatech.edu/people/faculty/flannery/publications/AJP73_March2005_265-272.pdf
Flannery, M. R. (2011a). D'Alembert–Lagrange analytical dynamics for nonholonomic systems. *Journal of Mathematical Physics, 52*(3), 032705. doi:10.1063/1.3559128.
Flannery, M. R. (2011b). The elusive d'Alembert-Lagrange dynamics of nonholonomic systems. *American Journal of Physics, 79*(9), 932–944.
Gánti, T. (2003). *The principles of life*. New York: Oxford University Press.
Gell-Mann, M. (1994). *The Quark and the Jaguar: Adventures in the simple and the complex*. New York: Henry Holt.
Goldstein, H. (1950). *Classical mechanics*. Reading, Mass: Addison-Wesley.
Gray, C. G., & Taylor, E. F. (2007). When action is not least. *American Journal of Physics, 75*(5), 434–458.

Groß, R., & Dorigo, M. (2007). Fifty years of self-assembly experimentation. In W.-M. Shen, H. Lipson, K. Stoy, & M. Yim, *Proceedings of the workshop on self-reconfigurable robots/systems and applications*. USA: USC Information Science Institute.
Groß, R., & Dorigo, M. (2008). Self-assembly at the macroscopic scale. *Proceedings of IEEE, 96*(9), 1490–1508.
Hooker, C. A. (1994). Idealisation, naturalism, and rationality: Some lessons from minimal rationality. *Synthese, 99*, 181–231.
Hooker, C. A. (2004). Asymptotics, reduction and emergence. *British Journal for the Philosophy of Science, 55*, 435–479.
Hooker, C. A. (2009). Interaction and bio-cognitive order. *Synthese, 166*(3), 513–546.
Hooker, C. A. (Ed.). (2011a). *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10). Amsterdam: North Holland/Elsevier.
Hooker, C. A. (2011b). Introduction to philosophy of complex systems. Part A: Towards framing philosophy of complex systems. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 3–92). Amsterdam: North Holland/Elsevier.
Hooker, C. A. (2011c). Introduction to philosophy of complex systems. Part B: Aninitial scientific paradigm + philosophy of science for complex systems. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 843–912). Amsterdam: North Holland/Elsevier.
Hooker, C. A. (2011d). Conceptualising reduction, emergence and self-organisation in complex dynamical systems. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 197–224). Amsterdam: North Holland/Elsevier.
Kaufman, S. (2000). *Investigations*. New York: Oxford University Press.
Mackworth, A. K. (2009). *Agents bodies constraints dynamics evolution* (pp. 7–28). Spring: AI Magazine.
Marr, D. (1982). *Vision*. Cambridge, Mass: MIT Press.
Maynard-Smith, J., & Szathmary, E. (1995). *The major transitions in evolution*. New York: Freeman.
Moreno, A. (2007). A systemic approach to the origin of biological organization. In F. Boogerd, F. Bruggeman, J.-H. Hofmeyr, & H. V. Westerhoff (Eds.), *Systems biology: Philosophical foundations* (pp. 243–268). Amsterdam: Elsevier.
Moreno, A., & Ruiz-Mirazo, K. (1999). Metabolism and the problem of its universalization. *Biosystems, 49*, 45–61.
Moreno, A., Ruiz-Mirazo, K., & Barandiaran, X. (2011). The impact of the paradigm of complexity on the foundational frameworks of biology and cognitive science. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 313–336). Amsterdam: North Holland/Elsevier.
Newman, S. A. (1970). Note on complex systems. *Journal of Theoretical Biology, 28*, 411–413.
Newman, S. A. (2011a). Complexity in organismal evolution. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 337–356). Amsterdam: North Holland/Elsevier.
Newman, S. A. (2011b). The animal egg as evolutionary innovation: A solution to the 'embryonic hourglass' puzzle. *Journal of Experimental Zoology, B: Molecular and Developmental Evolution, 316*(7), 467–483.
Newman, S. A., & Bhat, R. (2008). Dynamical patterning modules: Physico-genetic determinants of morphological development and evolution. Physical Biology, 5, 015008. Available at http://stacks.iop.org/PhysBio/5/015008.
Nolfi, S. (2011). Behavior and cognition as a complex adaptive system: Insights from robotic experiments. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 445–467). Amsterdam: North Holland/Elsevier.
O'Connor T., & Wong, H. Y. (2002) Emergence. Stanford Encyclopedia of Philosophy. http://plato.stanford.edu/entries/properties-emergent/
Pattee, H. H. (1971). Physical theories of biological co-ordination. *Quarterly Reviews of Biophysics, 4*, 255–276. Reprinted in M. Grene, E. Mendelsohn (Eds.), *Topics in the philosophy of biology*, Reidel, Amsterdam, 1976. doi:10.1017/S0033583500000640 (published online) 2009.
Pattee, H. H. (1973). The physical basis and origin of hierarchical control. In H. H. Pattee (Ed.), *Hierarchy theory: The challenge of complex systems* (pp. 71–108). New York: Geo. Braziller.
Pfeifer, R., & Bongard, J. C. (2007). *How the body shapes the way we think: A new view of intelligence*. Cambridge, Mass: MIT Press.
Raynor, A. (1977). *Degrees of freedom: Living in dynamic boundaries*. London: Imperial College Press.
Rueger, A. (2005). Perspectival models and theory unification. *British Journal for the Philosophy of Science, 56*(3), 579–594.

Ruiz-mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*(3), 235–259.
Shalizi, C. R. (2006). Methods and techniques of complex systems science: An overview. In T. Deisboeck, J. Kresh (Eds.), *Complex systems science in bio-medicine*. Springer, New York. Also available at http://arXiv.org/abs/nlin/0307015.
Schmidt, M., & Lipson, H. (2009). Distilling free form natural laws from experimental data. *Science, 324*(5923), 81–85.
Skewes, J., & Hooker, C. A. (2009). Bio-agency and the problem of action. *Biology and Philosophy, 24*(3), 283–300.
Subhankar R., & Shamanna, J. (2006) Understanding D'Alembert's principle: System of pendulums. Available at http://arxiv.org/PS_cache/physics/pdf/0606/0606010v2.pdf.
Soltakhanov, S. K., Yushkov, M. P., & Zegzhda, S. A. (2009). *Mechanics of non-holonomic systems*. Heidelberg: Springer.
Szathmáry, E (2005). Life: In search of the simplest cell. *Nature, 433*(3), 469–470.
Tegnér, J., Yeung, M. K. S., Hasty, J., & Collins, J. J. (2003). Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling. *Proceedings of the National Academy of Science, 100*(10), 5944–5949.
Turing A. (1952) The chemical basis for morphogenesis. *Philosophical transactions of the Royal Society of London, B: Biological Sciences, 237*(641), 37–72. Available at http://www.jstor.org/stable/92463.
Wackerbauer, R., Witt, A., Atmanspacher, H., Kurths, J., & Scheingraber, H. (1994). A comparative classification of complexity measures. *Chaos, Solitons & Fractals, 4*(1), 133–173.

## Author Biography

**Cliff Hooker,** Ph.D. [Physics] Sydney University, Australia, 1967, Ph.D. [Philosophy], York University, Canada, 1970, Fellow of the Australian Academy of Humanities, is Professor Emeritus of Philosophy and Director of the Complex Adaptive Systems Research Group, University of Newcastle, Australia (http://www.newcastle.edu.au/school/hss/research/groups/complex-adaptive-systems-research-group/). He is author of 150+ research papers and author/editor of 20 books in foundations and philosophy of physics, complex systems and scientific method. He has researched and taught foundations of physics with physicists, systems analysis, policy and professional ethics with engineers, business managers and environmental scientists, foundations of bio-cognition and scientific method with psychologists, and all these and more to philosophy students. He recently brought together a group of leading scientists and technically trained philosophers to produce analyses of the foundational challenges posed by the introduction of complex systems models and methods across the sciences: C. A. Hooker (Ed.). Philosophy of Complex Systems. North Holland/Elsevier, Amsterdam, 2011.

Studies in History and Philosophy of Biological and Biomedical Sciences 61 (2017) 20–34

Contents lists available at ScienceDirect

Studies in History and Philosophy of Biological and Biomedical Sciences

journal homepage: www.elsevier.com/locate/shpsc

# Biology meets physics: Reductionism and multi-scale modeling of morphogenesis

Sara Green [a,*], Robert Batterman [b]

[a] Department of Science Education, University of Copenhagen, Øster Voldgade 3, 1350 Copenhagen, Denmark
[b] Department of Philosophy, University of Pittsburgh, 1028-A Cathedral of Learning, Pittsburgh, PA 15260, USA

ARTICLE INFO

*Article history:*
Received 18 January 2016
Received in revised form
8 December 2016
Available online 23 December 2016

*Keywords:*
Multi-scale modeling
Boundary conditions
Explanatory reduction
Developmental biology
Biomechanics
Reductive explanations

ABSTRACT

A common reductionist assumption is that macro-scale behaviors can be described "bottom-up" if only sufficient details about lower-scale processes are available. The view that an "ideal" or "fundamental" physics would be sufficient to explain all macro-scale phenomena has been met with criticism from philosophers of biology. Specifically, scholars have pointed to the impossibility of deducing biological explanations from physical ones, and to the irreducible nature of distinctively biological processes such as gene regulation and evolution. This paper takes a step back in asking whether bottom-up modeling is feasible even when modeling simple physical systems across scales. By comparing examples of multi-scale modeling in physics and biology, we argue that the "tyranny of scales" problem presents a challenge to reductive explanations in both physics and biology. The problem refers to the scale-dependency of physical and biological behaviors that forces researchers to combine different models relying on different scale-specific mathematical strategies and boundary conditions. Analyzing the ways in which different models are combined in multi-scale modeling also has implications for the relation between physics and biology. Contrary to the assumption that physical science approaches provide reductive explanations in biology, we exemplify how inputs from physics often reveal the importance of macro-scale models and explanations. We illustrate this through an examination of the role of biomechanical modeling in developmental biology. In such contexts, the relation between models at different scales and from different disciplines is neither reductive nor completely autonomous, but interdependent.


## 1. Introduction

An important reductionist assumption is that multi-scale systems can be described "bottom-up", if only sufficient details about the states of the components are available. Historically, this assumption has been debated in philosophical discussions about whether biology is reducible to physics. The positivist ideal of a unity of science pictured the relations between scientific disciplines in a "layer-cake" hierarchy where theories from respective disciplines target a specific level or scale of phenomena (Oppenheim & Putnam, 1958).[1] Physics was considered the most fundamental "model discipline" targeting the lowest organizational level, and progressive reduction was considered an important aspect of scientific development (see also Hüttemann & Love, 2016).

The view that an ideal or fundamental physics would be sufficient to explain all macro-scale phenomena has been met with criticism from philosophers of biology. Scholars have stressed the irreducibility of biological features, such as gene regulation or evolution, and argued that biological explanations are irreducible to physical laws and principles (e.g., Bechtel & Richardson, 1993; Bertalanffy, 1969; Burian, Richardson, & Van der Steen, 1996; Dupré, 1993; Machamer, Darden, & Craver, 2000; Mayr, 1988, 2004; Winther, 2009). These contributions have offered important insights to distinctive features of living systems and biological research. However, an important question that is rarely addressed is whether the ideal of progressive reduction of higher-level explanations is supported in physics, i.e., in the discipline that was taken as a model for the reductionist ideal. We argue that lessons from multi-scale modeling offer resistance to reductionism which cross-cut discussions in philosophy of biology and philosophy of physics.

* Corresponding author.
*E-mail address:* sara.green@ind.ku.dk (S. Green).

[1] We use the term "level" when referring more explicitly to part-whole relations in a hierarchical description or a functional system (demarcated by boundaries such as the cell membrane), but we prefer the term "scale" when referring to spatial scaling because biological "levels" are often not straightforwardly distinguished (see also Noble, 2012). For a more detailed discussion of biological levels and part-whole relations, see (Kaiser, 2015).

http://dx.doi.org/10.1016/j.shpsc.2016.12.003
1369-8486/

We focus on what Mayr (1988) calls *explanatory reduction,* which involves explaining phenomena at higher scales in terms of processes at lower scales or levels of organization (e.g., molecules or genes).[2] Other important aspects of reductive explanations are that they typically analyze biological parts in isolation from their original context and give explanatory priority only to factors internal to the system (Kaiser, 2015). In recent discussions on explanatory reduction it is debated whether the constitution of macroscale systems by microscale components allows the researcher to explain the system only with reference to properties of the lower scale constituents (Brigandt & Love, 2012). For instance, although the composition of polypeptides is reducible to a sequence of amino acids, it has been argued that it is not possible to explain protein folding from physical laws and knowledge about amino acids alone (Love & Hüttemann, 2011). The prospect of reductive explanations in biology and physics is, however, an ongoing issue of debate.

This paper sheds further light on the debate on reductionism by clarifying how lessons from multi-scale modeling in both physics and biology offer resistance to the idea that multi-scale systems can be modeled and explained "bottom-up". Secondly, unlike what one might expect from physical science approaches, we argue that work within biomechanics brings attention to the problems of understanding biological processes and parts in isolation from their original context in cells or tissue structures (Kaiser, 2015). Thus, rather than enforcing reductionism, physical science approaches can help reveal the limitations for reducing explanations in developmental biology to genetics. Accordingly, we argue that the role of physical science approaches in biology with respect to reductionism should be revisited.

Our aim is to bring attention to the *tyranny of scales problem* that has so far mainly been discussed in the context of physics (Batterman, 2012; Oden, 2006; see however; Lesne, 2013).[3] The problem refers to the scale-dependency of physical behaviors that presents a hard challenge for modeling and explaining multi-scale systems. No single mathematical model can account for behaviors at all spatial and temporal scales, and the modeler must therefore combine different mathematical models relying on different boundary conditions. Fig. 1 illustrates the interplay of models describing processes at different scales. The expression h = H(r) indicates how macroscale features and properties arise from the collective behavior of microscale variables.[4] However, the expression at the left side of the figure, r = R(h), indicates how microscopic elements are affected by macroscopic variables *h* through

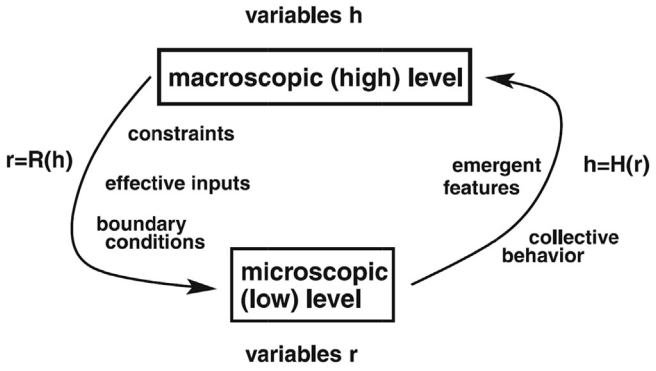


**Fig. 1.** Illustration of the interplay of "microscopic" and "macroscopic" modeling. From *Acta Biotheoretica*, Multiscale analysis of biological systems 61, 2013, p. 14, A. Lesne, reprinted with permission of Springer.

the influences of constraints, effective inputs, and boundary conditions.

*Constraints* in this context are understood broadly as conditions that limit and enable certain behaviors, such as tissue stiffness that influences the bending properties of biological structures (see also Hooker, 2013). Modelers often express physical constraints mathematically as *boundary conditions*, i.e., as definite mathematical parameters. Boundary conditions are often indispensable to the modeling procedure, because the equations cannot be solved without imposing limits on the domain of the model. In this paper we describe how the scale-dependent behavior of physical and biological systems forces researchers to combine different experimental and representational strategies targeting specific scales.[5] We illustrate this with examples from both physics and biology.

Since we are drawing on contemporary cases of multi-scale modeling, the reductionist may object that all we point to are practical limitations of *current* science. That is, one may object that it *in principle* should be possible to explain macro-scale systems with reference only to lower scale molecular details. The specific methodologies will indeed develop and change over time. However, it should be noted that the need to combine different approaches arises due to the fundamental challenge that concepts used to characterize systems and their behaviors can change as one changes scale: They are *multi-valued* across scales (Wilson, 2012). For example, not only does the concept of "surface" change as one moves toward the nanoscale (where there is typically more "surface" than bulk material) but so do the kinds of concepts we need to characterize the dominant behaviors at the different scales (Bursten, 2015; see Section 2).[6] Against the background of this complexity, we find appeals to in principle derivations empty without suggestions of how to make such inferences (see also Batterman, 2016). Rather than logical possibilities for explanatory reduction, this paper is concerned with the challenges faced in

[2] Explanatory reduction is distinguished from *Constitutive reduction* and *Theory reduction*. Constitutive reduction (also called ontological reduction or (token) physicalism) refers to the acceptance that biological systems are nothing but physical-chemical systems. *Theory reduction* considers the possibilities of reducing (in a logical sense) higher-level theories in special sciences to more fundamental ones (cf., Rosenberg & Arp, 2010; Sarkar, 1998; Schaffner, 1993; Winther, 2009). More recently, philosophers of biology have discussed this kind of reduction in terms of explanatory relevance, e.g., whether the explanatory power in biology is constituted by physico-chemical principles or biological mechanisms (Machamer et al., 2000; Weber, 2008). A separate kind of reductionism, *methodological reductionism,* considers heuristic strategies that simplify the problem space for scientific analysis (Bechtel & Richardson, 1993; Brigandt & Love, 2012; Green, 2015).

[3] To be sure, mechanistic accounts in philosophy of biology (e.g., Bechtel & Richardson, 1993; Machamer et al., 2000) have taken issue with reductionism in arguing against reducibility of biology to physics and in allowing for interlevel explanations. However, mechanistic accounts have so far not attended to the challenges for reductionism provided by the scale-dependency of physical behaviors (see Skillings, 2015 for further discussion).

[4] As indicated on the figure, such features are often labelled as "emergent". We shall not go into the question about emergence in this paper (see Boogerd, Bruggeman, Richardson, Achim, & Westerhoff, 2005 for a detailed discussion of emergence in biology).

[5] This paper focuses on the adequacy of explanatory reduction through a demonstration of the requirement of multiple models. As an anonymous reviewer pointed out, it is possible to agree with the inadequacy of explanatory reduction but argue that a single higher-level model is adequate. Our account would offer resistance also to a monistic anti-reductionist view of this kind but we do not develop such an argument in the paper.

[6] Already Galileo (*Discorsi*, 1638) pointed to the importance of scale when considering the disproportional relation between the minimal thickness of bone structures and animal size. Biologists investigating morphological constraints on animal form have similarly stressed that macroscale physics does not apply to microorganisms. At this scale, gravity is a weaker force whereas surface properties and Brownian motion are central parts of the analysis (see e.g., Purcell, 1977; Vogel, 2009). Similarly, the models most useful to model molecular behavior are rarely the most useful for modeling tissues. See Section 2 for further clarification.

scientific practice. In our view, explanatory reduction fails if macroscale models, measurements, and concepts are indispensable for explanations of multi-scale systems. Showing that these are indeed indispensable is the main aim of this paper.[7]

Considering the challenges for bottom-up approaches in physics also calls for us to revisit the relation between physics and biology. Our analysis does not imply or support any sharp distinction between contributions from biology and physics. Both disciplines are highly diverse and often intertwined in interdisciplinary fields.[8] A weaker distinction of disciplinary inputs is, however, useful for revisiting the relation between physical science approaches in biology and reductionism. As mentioned above, physics has often been pictured as a discipline solely targeting the most fundamental "lower levels", with a preference for simple deterministic models. Interestingly, appeals to physical science approaches in the examined cases of multi-scale modeling in developmental biology do not support this view. Rather, they show that macro-scale features (i.e., those at cell and tissue level) are indispensable and irreducible to lower-scale explanations. Moreover, we propose that the requirement of macroscale parameters (e.g., tissue stiffness) as boundary conditions for models at lower scales (Fig. 1) provides a concrete instantiation of top-down effects (Section 4.1). We highlight how recent insights to biomechanical aspects of morphogenesis challenge deeply entrenched presuppositions about the explanatory priority of lower scales, e.g., of the special priority of the molecular or genetic level in developmental biology (cf. Rosenberg, 1997). These challenges are not specifically directed at developmental biology, and we shall comment more briefly on how similar insights can be derived from studies of multi-scale cardiac modeling and cancer research.

We shall proceed as follows. Section 2 introduces the tyranny of scales problem in physical and biological contexts. Section 3 examines the application of physical science approaches in developmental biology and highlights the importance of tissue-scale mechanics for embryo development. Section 4 elaborates on the specific challenges posed for explanatory reduction in the context of multi-scale modeling in biology in light of Sections 2 and 3. We describe the relation between models at different scales (and from physics and biology) as non-reductive and interdependent. Section 5 offers concluding remarks.

## 2. The tyranny of scales in physics and biology

One of the hardest problems in modeling the behaviors of physical systems is to deal with structures that exist across different spatial scales. Generally, relying on a single mathematical model to describe the behavior of a physical system at all scales is not possible, because dynamical and material properties are scale-dependent (Wilson, 2012). Even successful modeling of a relatively simple multi-scale system such as a steel beam requires different models.[9] At atomic scales, steel has a regular lattice structure but at higher scales it exhibits elastic behavior that is well-described by the Navier-Cauchy elasticity equations (see Fig. 2). These equations model the material as a *continuum* and completely ignore atomic structure. Additionally, and very importantly, at intermediate (meso) scales, steel presents a host of other structures such as lamellar inclusions of pearlite, cracks, grain boundaries, etc. To fully understand the behavior of bending steel requires that one bridges across these widely separated scales, i.e., that one can combine models at different scales that inform each other. The problem is hard because "the principal physics governing events often changes with scale, so that the models themselves must change in structure as the ramifications of events pass from one scale to another" (Oden, 2006, p. 2930).

Modeling in biology must also confront the tyranny of scales. Like the steel example, some aspects of biological structures require continuum models whereas others have to take into account the structural diversity of and stochastic relations between the discrete interacting cells and molecules (Lesne, 2013). In biology, researchers face the additional challenge that different integrated processes operate also on different time-scales, from milliseconds to hours, days or even years (Davidson, von Dassow, & Zhou, 2009).

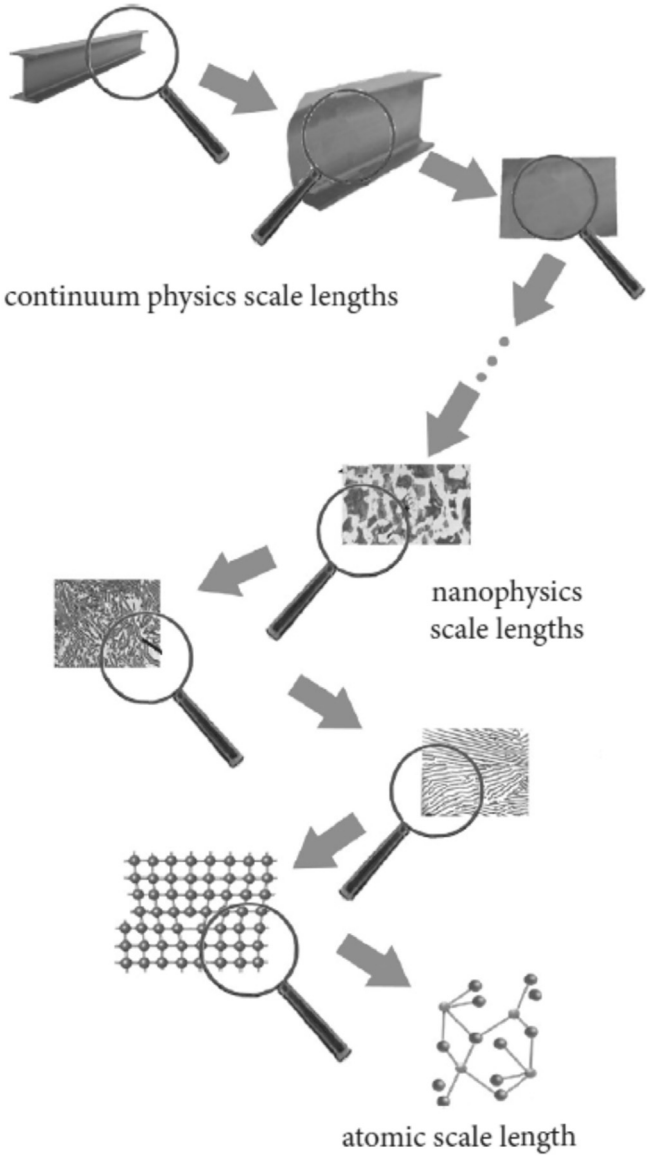


**Fig. 2.** Macro and microstructures of steel. Source: Batterman (2012).

[7] We provide a clarification of what is meant by indispensability at the end of Section 3, derived from case examples.

[8] The intertwinement of physics and biology is explicit in biophysics that comprises a range of important research areas such as membrane physics, biomechanics, as well as research on the energetics of protein folding, molecular motors, and mechanosensors (Dunn & Price, 2015; Morange, 2011). Many of these developments also benefit from engineering approaches, and there are important differences between physics and engineering. But because the models we examine are developed in physics, we shall not in this paper distinguish between physics and bioengineering.

[9] The steel beam is actually incredibly hard to model and is far from simple. Simple should here be understood in comparison to biological systems.

As Newman et al. (2011, p. 313) clarify, developmental systems display both discrete and continuous aspects, depending on the specific spatial and temporal scale of the specific developmental processes. These aspects make the modeling strategy "inescapably hybrid, mathematically and computationally" (Newman et al., 2011, p. 313).

Continuum models treat discrete and diverse entities that exist in finite numbers as a continuous variable. These models are used to model macroscopic behaviors that are relatively independent of smaller-scale properties and local dynamics of the system components (Batterman, 2016). This situation can be seen as analogous to how many applications of the ideal gas law are independent of information about the specific dynamic trajectories of individual molecules of the gas because microscopic fluctuations average out. In biology, such fluctuations can also be buffered by regulatory circuits, yielding robust functions despite perturbations at lower levels (Lesne, 2013). For instance, when modeling cell motion at tissue scales or at that of the whole embryo, developmental biologists often rely on coupled partial differential equations that ignore the stochastic properties of interactions between individual molecules and cells. Similar to mean-field approaches in physics, they study the collective dynamics of the population of cells rather than the individual components (Lesne, 2013). These modeling choices are not just motivated by considerations about tractability but also reflect the problem that some features cannot be modeled at all scales. Just like one cannot attribute features like temperature and pressure to an individual gas molecule, so are some features of cell populations not possible to derive from models of isolated cells. That is, macro-scale phenomena require alternative modeling frameworks. Tissue-scale models of cell movement typically rely on reaction-diffusion equations (modeling e.g. chemotaxis responses or the mechanical influence of the ECM) or integro-differential equations that model cells as flows, akin to processes in fluid dynamics. This can be done via Navier-Stokes equations, or via mass action laws describing chemical kinetics (Hatzikirou & Deutsch, 2011).[10]

The most appropriate model for capturing biological phenomena also depends on temporal scaling. Shawky and Davidson (2015, p. 154) clarify how engineering models of embryonic tissues "range from elastic solid-like to viscous liquid-like, depending on the time-scale of measurement".[11] When modeling complex phenomena like a developing embryo or the human heart, researchers must combine deterministic and stochastic models to couple processes across spatial and temporal scales. At tissue-scales the relevant measurements span over longer times, allowing one to ignore fluctuations in biomolecular species. This often gives more accurate predictions. In contrast, at the molecular scale the dynamics is "dominated by random and short time fluctuations" in the concentrations of ions or proteins (Qu, Garfinkel, Weiss, & Nivala, 2011, p. 22). Combining different kinds of models in large-scale simulations spanning multiple spatial and temporal scales is far from trivial because the models often make different predictions about what will happen with the same system over time (Qu et al., 2011). But modeling the whole system using only one modeling framework is not possible because different aspects of the system dominate the behavior at characteristic scales.

One consequence of the "multi-valuedness" of multi-scale systems is that different details must be ignored by models operating at different spatial and temporal scales. Just as any useful model in science should ignore the degrees of freedom irrelevant for the specific modeling task, "a multiscale model should not intend to keep track of all details at all scales but only of the relevant details, whatever their scales" (Lesne, 2013, p. 17). As we shall clarify in the following section, boundary conditions play a crucial role for the purpose of representing physical and biological constraints and for combining models that account for complementary aspects of the system.

### 2.1. Boundary conditions

We begin again with a simple example from physics and then move to the biological context. An example where boundary conditions play an essential role is in the modeling of the harmonic structure of a violin string. One can determine the modes of the standing wave of a vibrating string by solving a wave equation. To solve the *partial* differential equation requires the imposition of mathematical boundary conditions that fix the endpoints of the string (Batterman, 2012; Wilson, 2012). Thus, it requires that the string at the bridge and nut of the violin remains absolutely stationary as the string vibrates. Unfortunately, while essential for the determination of the modes, if such strict conditions on the string's endpoints were actually imposed on the violin, it would make it impossible for the vibrations in the string to be amplified by the sound box of the violin and we would be unable to hear the instrument. Modeling the behavior of the vibrations that get amplified via the sound box requires that one completely shift scales and focuses on the molecular and sub-molecular interactions between the string and the bridge. Here the equations are of a completely different mathematical type. This is the realm of molecular dynamics governed by *ordinary* differential equations. The lesson here is that sometimes (quite often in fact) one needs to impose boundary conditions in order to efface physical details that will not allow one to model the behavior of interest at a given scale.[12] As mentioned, this is also the case for multi-scale modeling in biology where researchers combine discrete and continuous models, depending on the spatial scale of the biological phenomenon.

For modeling processes at the cell and tissue level, such as formation of the vertebrate limb, continuum models work well and are justified by the quantity of the elements modeled, the robustness of the regulatory dynamics, and the scale for which the phenomenon can be observed (Newman et al., 2011). As in the steel and violin examples, homogenizing heterogeneous entities as a continuum is a necessary requirement for the modeling procedure. For instance, solving a set of partial differential equations for morphological deformations involved in limb development requires that cells or cell layers be treated as fixed in space and time (Newman et al., 2011, pp. 320–221). Ignoring the microscopic processes of cell-cell interactions and subcellular mechanisms can, however, be problematic in contexts where a small number of elements have a large impact on a system. In such cases, upper-scale models must be combined with lower-scale discrete models (e.g., Langevin equations) that capture individual (microscopic) details such as kinetic rates of particular proteins. These models, in turn, often require that macroscale properties, including mechanical properties and dynamics of environmental inputs, are fixed or ignored as boundary conditions. A hard challenge in multi-scale modeling is therefore to connect discrete to continuous models to

[10] These models can be extremely complex and it is often necessary to discretize the continuum equations using numerical strategies such as Finite Element Methods (see Section 3.1).

[11] Consider how water in a swimming pool can seem to be very solid on the short time scale in which a diver belly flops. On the other hand, as one wades in, the water seems quite un-solid like.

[12] Mark Wilson (2012) calls this effacement "physics avoidance."

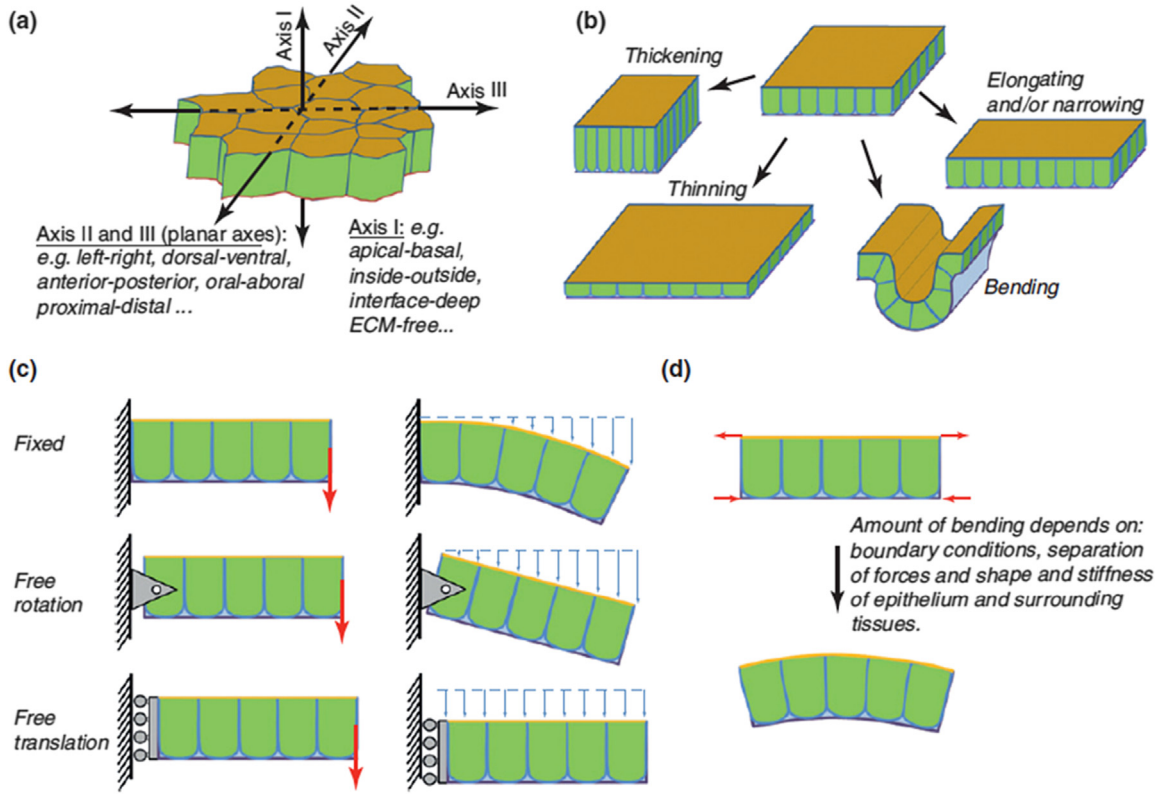


**Fig. 3.** Mechanics of epithelial sheets. Reprinted from *Trends in Cell Biology*, 22(2), Davidson, L.A., Epithelial machines that shape the embryo, 82-87, Copyright (2012), with permission from Elsevier.

bridge the gap between modeling frameworks targeting different spatial scales.

The role of boundary conditions in modeling of biological multi-scale systems can be further clarified through concrete examples from developmental biology. In such examples, boundary conditions are used to represent biomechanical constraints on morphogenic movements of epithelial sheets. The establishment of tissue boundaries and geometrical structures during morphogenesis is mediated and stabilized by interconnected adhesions between cells and the extracellular matrix (ECM). Adhesions fix cells and cell populations in structures with varying degrees of freedom for bending and motility (Davidson et al., 2009). The proteins involved in adhesion serve both mechanical and signaling roles through force-transmission and mechano-transduction, i.e. conversion of a mechanical stimulus into chemical cues that influence biochemical pathways (see also Section 3.2). One focus area in developmental biology examines the ability of cells to undergo changes in shape that impose apical-basal asymmetries and yield bending as displayed in Fig. 3a and b.

The densely packed structure of epithelial tissues can change shape in response to unbalanced mechanical stresses, and understanding the production and propagation of mechanical forces is therefore important for understanding development (Shawky & Davidson, 2015). Mechanical modeling in this context relates physical forces applied to an object (*stress*) to the resulting changes in the shape of the object (*strain*).[13] *Stiffness* describes the bending properties of the material, including the ability of the material to resist deformation.[14] The degree of strain a material exhibits when a defined stress is applied is expressed in the material's elastic modulus. Determining the stiffness of the epithelium and surrounding tissues (Fig. 3d) involves finding values for material parameters, such as Young's Modulus and various tensor fields. Importantly, defining such parameters for the viscous materials of biological tissues involves multi-scale analysis, because the characteristic of macroscale physical forces acting on the integrated effects of the dynamics of cell populations are practically invisible at the molecular scale (Davidson et al., 2009). Tractable models of various epithelial movements, however, require ignoring extreme lower-scale (molecular or genetic) details.

In the modeling of the violin's harmonic (continuum-scale) structure, the purpose of the modeling strategies is to crush lower-scale detail into boundary conditions or mechanical constraints. In biology similar strategies are employed. Here is biophysicist Lance Davidson:

"Mechanical boundary conditions allow engineers and physicists to simplify problems of mechanics by abbreviating complex structures with simpler structures that have limited degrees of motion … [B]oundary conditions are mathematical statements that can indicate restrictions on movement or rotation in any direction. Thus, biomechanical analyses of embryos do not necessarily need to recreate the entire embryo but rather simulate parts whose movement or margins are restricted by explicit boundary conditions", (Davidson, 2012, p. 83, p. 83).

[13] The concepts are here used in the technical sense of mechanics where *stress* is force per cross-sectional area of a material, and *strain* refers to the amount of deformation.

[14] In the biological context, the stiffness, or elastic response of a material to an applied force, depends not only on the material properties of the body, but also on its geometrical properties, and how it is held in place by the surrounding materials such as the ECM and protein structures connecting the cells.

As in the violin case, boundary conditions allow researchers to simplify problems by effacing lower scale details, but this effacement is not just a matter of mathematical convenience. Rather, it is required for the localization and identification behaviors at a given scale. Ideally, from a reductionist perspective, one would like to be able to determine the values for the material parameters at a more fundamental scale and model the system "bottom-up" in one coherent model capturing the "sum of the parts". But it is virtually impossible to do this even for physical modeling of biomechanical properties of adhesion because not all relevant processes can be measured and modeled in the same way. Shawky and Davidson (2015) review a number of different experimental techniques to measure mechanical properties of relevance for understanding adhesion and argue that multi-scale analysis is unavoidable. *Tissue scale* measurements, e.g. of bulk mechanical responses, can account for biomechanical properties of the whole system but retain many uncontrollable variables and cannot account for feedback in cell-signaling and molecular pathways in response to stress. Measurements at the *molecular* and *cell scale,* in turn, can account for finer-grained interactions between cells and molecules, but these techniques require that individual cells or molecules be removed from their native environment. Accordingly, experiments and models targeting the molecular scale cannot account for the constraints imposed by the system as a whole on the degree of freedom of microscale processes.

Thus, the possibility of bottom-up modeling is blocked by the need for boundary conditions imposed at higher scales. These limitations to the reductionist approach are also stressed by investigators in the Cardiac Physiome project, initiated by Denis Noble and Jim Bassingthwaighte, as follows:

"[C]omplex systems like the heart are inevitably multiscalar, composed of elements of a diverse nature, constructed spatially in a hierarchical fashion. This requires linking together different types of modeling at the various levels. It is neither possible nor explanatory to attempt to model at the organ and system levels in the same way as at the molecular level and cellular level … [I]f we did not include the constraints that the cell as a whole exerts on the behavior of its molecules [we would be lost in a mountain of data]", (Bassingthwaighte, Hunter, & Noble, 2009, p. 597.).

The reliance on boundary conditions in multi-scale modeling highlights the importance of system-level constraints and how some details can be *irrelevant* for modeling a specific process at a characteristic scale (Batterman, 2012; 2016). To understand how the system functions as a whole, different models must be combined through careful attention to the boundary conditions imposed for each description. The combination of models at different scales is particularly challenging in the context of developmental biology because the cytoskeleton and adhesions are not just coupled mechanically across scales, but are also involved in complex intra and intercellular signaling pathways. In the following, we shall further clarify the interdependency through an examination of insights from mechanical modeling in developmental biology.

## 3. Tissue mechanics in embryo development

Much research in developmental biology for the past decades has focused on gene regulatory networks and cell signaling pathways (Peter & Davidson, 2015; Rosenberg, 1997). But researchers are increasingly realizing that many developmental processes can only be studied through multi-scale modeling (Brodland et al., 1994; Davidson, 2012; Newman et al., 2011; Wyczalkowski, Chen, Filas, Varner, & Taber, 2012). In the following we examine in further detail how mechanical force production and propagation of stress and strain contribute to the shaping of the early embryo. We start with a case that illustrates how mechanical modeling brings insight to the importance of macro-physical properties of cells and ECM layers for developmental processes.

### 3.1. Mechanical modeling of gastrulation

We begin by examining the use of mechanical modeling in research on the gastrulation process, i.e., the period in embryogenesis where morphological complexity and cell patterns are established from simple multicellular systems. Gastrulation involves radical spatial transformations where the three germ layers (ectoderm, mesoderm and endoderm) are established and take up specific topological positions through highly coordinated cell movements. The different germ layers later give rise to different tissue types. Sea urchin embryos have for many years been used as a model organism to study gastrulation because of their simple organization, optical transparency, and lately also because of discovered commonalities between sea urchin genomes and that of vertebrates (Rast, Smith, Loza-Coll, Hibino, & Litman, 2006). The first steps in sea urchin development involve radial cleavage resulting in the formation of a hollow sphere called a blastula. Sea urchin gastrulation is traditionally divided into two phases called primary and secondary invagination. During primary invagination, a flattened epithelial sheet called the vegetal plate thickens, bends inwards and gives rise to a gut rudiment (archenteron) that elongates over a couple of hours. In the second step, the tip of the invaginating area reaches the inner surface of the apical plate (opposite the base of the organism) and crosses the blastocoel (Fig. 4).

In response to controversies about the mechanisms of primary invagination, Davidson and colleagues developed a mechanical model representing the relations between five candidate mechanisms and biomechanical constraints (Davidson, Koehl, Keller, & Oster, 1995). The mechanical model represents the biomechanical properties of the elastic structures in the embryo that resist and direct forces of invagination. Examples include parameters for, the stiffness of the ECM, as well as cytoskeletal and extracellular fibers. The researchers used finite element mechanical models to simulate the spatio-temporal process of primary invagination as proposed by five different mechanisms (Fig. 5). Finite element methods are commonly used in engineering and biomechanical analysis to find approximate solutions for partial differential equations representing spatial deformations of complex structures. They divide complex physical structures (cells, cell layers, proteins etc.) into finite element subunits that represent a block of material. In this study, the simulation consisted of a 3D representation of the mesenchyme blastula where the geometry of the finite elements was based on data from imaging measures (transmission electron microscopy) of living embryos. The blastula was modeled as a system of three cell layers with associated values for mechanical and morphological parameters such as thickness, elasticity and strain of the elements based on experiments and estimations.

Fig. 5A illustrates the inner cell layer, the apical lamina and hyaline layer of the blastocoel. The cell functions involved in gastrulation (adhesion and mechanotransduction) are mediated by a fibrous meshwork of proteins (e.g., contractile protusions). Fig. 5B–F show the alternative mechanisms proposed for primary invagination. Since our main focus will be on the role of the mechanical model, we will just briefly summarize characteristics of the proposed mechanisms in the figure text.

The simulations based on the mechanical model showed that each mechanism can only operate within a specific range of physical properties of the epithelial sheet, related to the relative stiffness of the cell layers and ECM layers. For each mechanism, physical constraints define the range of possible parameter values for the

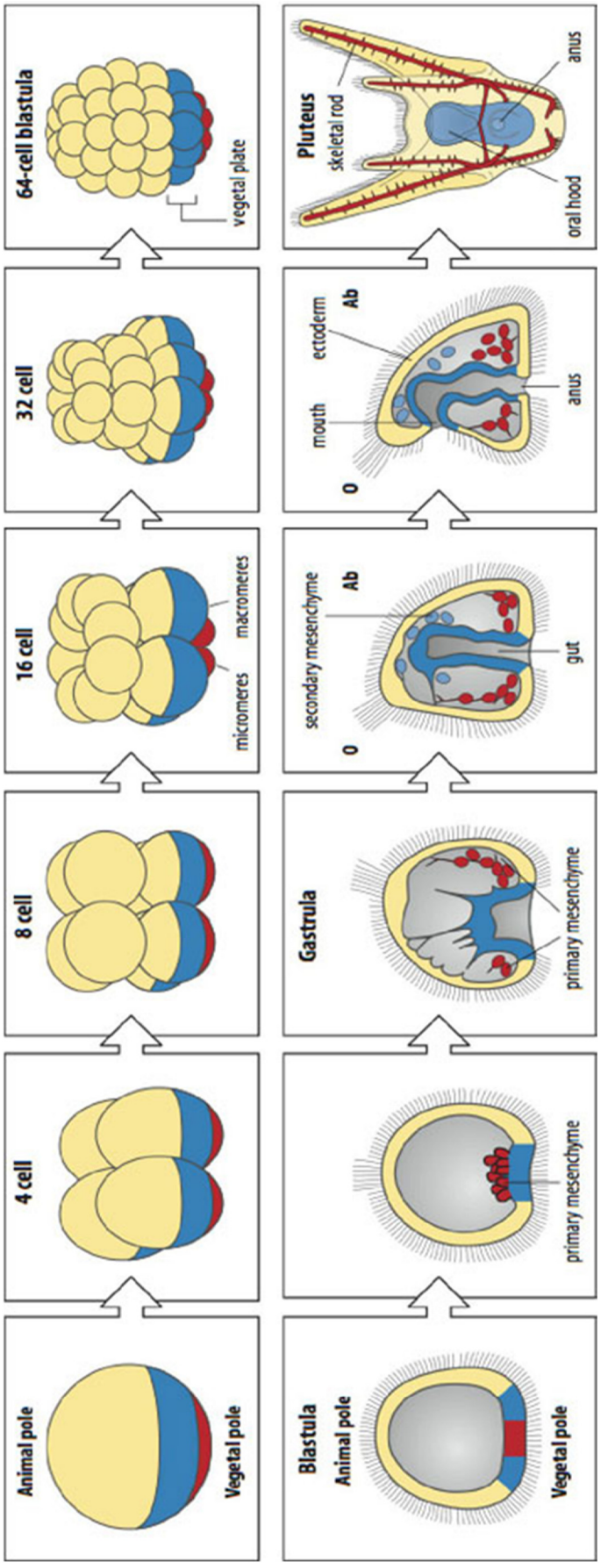


**Fig. 4.** Illustration of sea urchin development. See text for clarification.

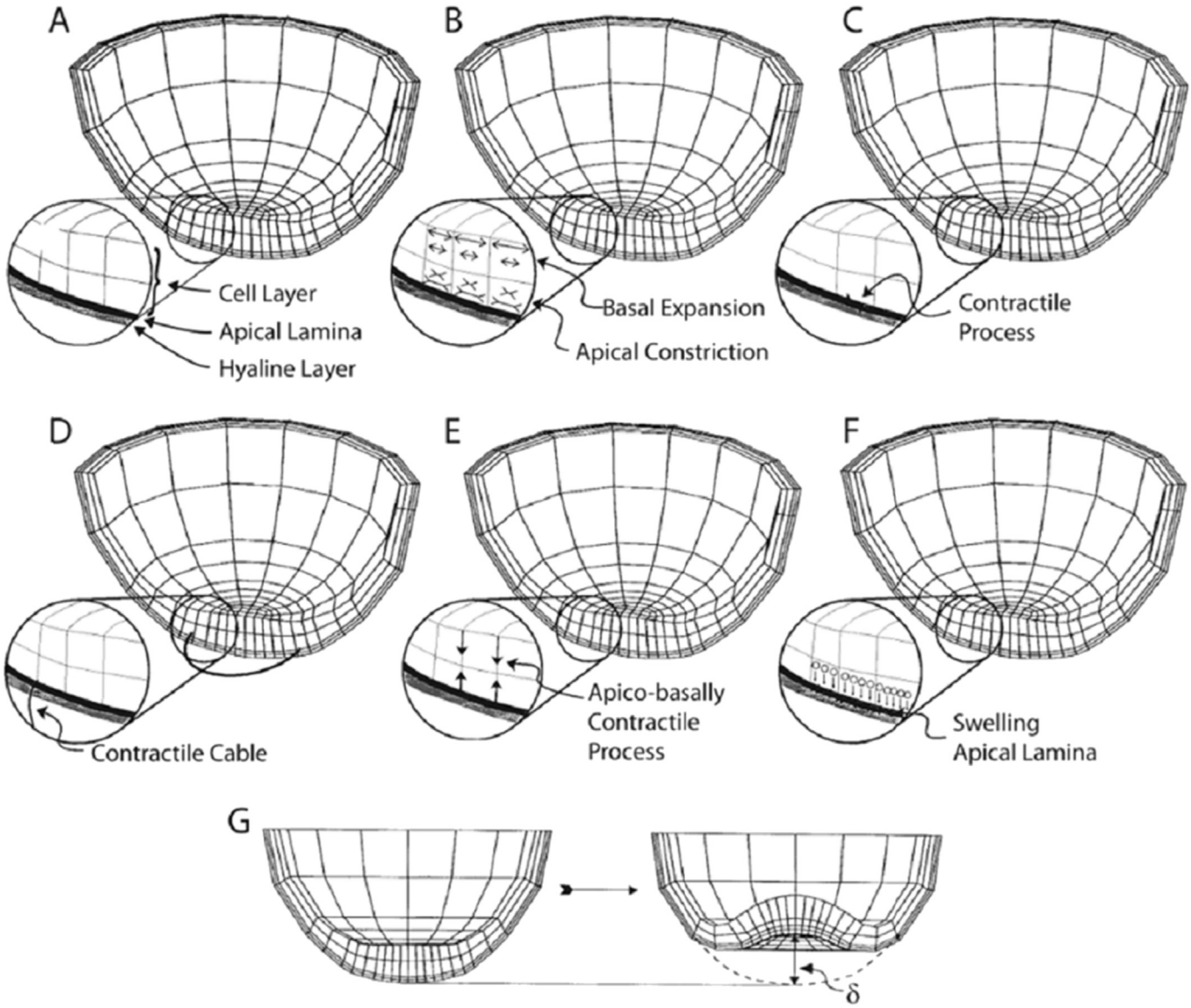


**Fig. 5.** Simulation of the five mechanisms for primary invagination via the finite element method (source: Davidson et al., 1995, reprinted with permission). A) Illustration of the inner cell layer, the apical lamina and hyaline layer. B) *The apical contraction hypothesis*: Invagination results from apical constriction of cells in the vegetal plate. C) *The cell tractor hypothesis*: Invagination follows a directed movement of cells towards the center of the plate while tractoring on contracting protrusions on the outer layer of the blastula. D) *The apical contractile ring hypothesis*: Circumferential contraction of an apical protein cable encircling the vegetal plate, causing inward bending of the vegetal plate. D) *The apicobasal contraction hypothesis*: Contraction in the cytoskeleton generates a compressive force that causes the vegetal epithelium cells to buckle inward. E) *The gel swelling hypothesis*: A glycosylated protein is secreted regionally into the apical lamina, leading the cells to swell. The swelling of the apical lamina cells, but not the hyaline layer, creates inward bending. G) Representation of the deformation of the embryo.

relative elastic moduli of the cells in the late mesenchyme blastula (cell layer, apical lamina and hyaline layer). This allows the model to make testable predictions for mechanical properties associated with the cell shape changes for each mechanism. Fig. 6 shows the material parameter space allowed for efficient invagination (greater than 12 μm) by the five mechanisms. For instance, the *apical constriction mechanism* (5B) can only work if the relative stiffness between the apical lamina and cell layer is more than 13 to 1, *and* the hyaline layer less than 5 times as stiff as the cell layer (20 Pa). In contrast, the *gel swelling mechanism* (5E) entails that the hyaline layer is more than 60 times as stiff as the cell layer.

The mechanical model was subsequently used to design experimental procedures to measure relevant mechanical properties of the elastic modulus of the cellular and extracellular matrix. Davidson, Oster, Keller, and Koehl (1999) conducted a compression test of the blastula wall of sea urchin embryos to measure the stiffness of the wall over time. The test showed that the apical constriction mechanism and the apical ring contraction mechanism are physically implausible because the stiffness of the blastula wall is much lower than the apical ECM.

Physical modeling can thus reveal important insights to the possible parameter space for robust geometric pattern production as well as insights concerning the sensitivity of developmental processes to biomechanical factors. As mentioned in Section 2.1, because the physical forces act on the integrated effects of cell populations, biomechanical modeling of the viscous materials requires macro-scale measurements and models (Davidson, 2012; Davidson et al., 2009). One possible objection is that the examined macroscale properties are merely background conditions for biological explanations. The following section responds to this

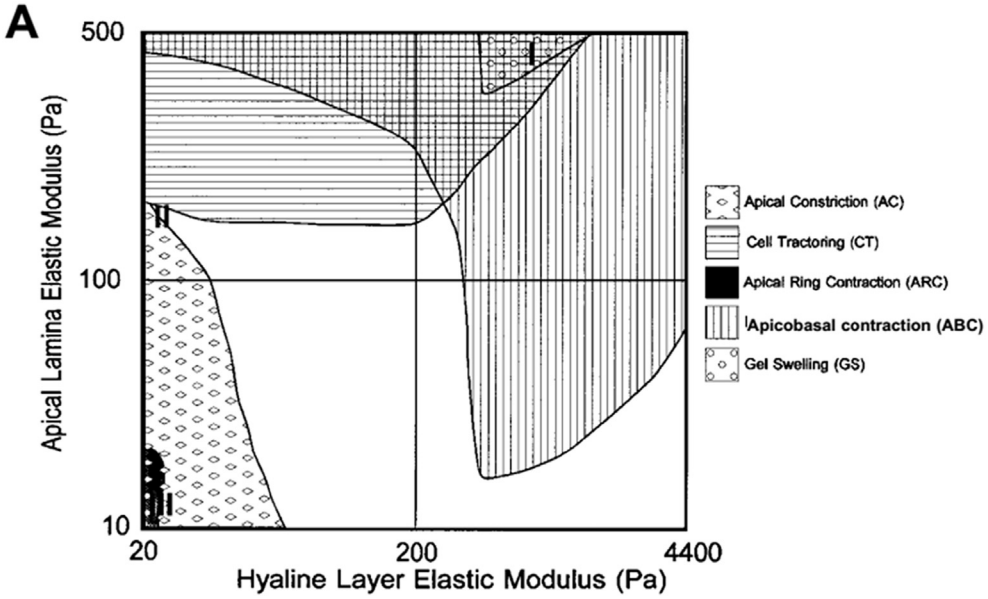


**Fig. 6.** Material parameter space. See text for details. Source: Davidson et al. (1995), reprinted with permission.

objection by drawing on recent investigations of how mechanical feedback from the cellular environment influences gene expression, cell differentiation and cell movement.

### 3.2. The explanatory power of macroscale biomechanics

The example just described shows that developmental processes are sensitive to physical properties of cell layers and extracellular matrices. We consider the above example to be a search for an explanatory model that includes both force generating mechanisms and mechanical constraints. In the words of Davidson et al. (1995, p. 2005): "Any explanation of how primary invagination works must incorporate both the passive mechanical properties of the embryo as well as the force-generating mechanisms within the epithelial sheet driving invagination". Note here that the different aspects are distinguished by the conceptualization of "passive" (external) physical properties and "active" (cell-driven) molecular mechanisms. While these terms nicely capture the difference between cell-autonomous "programmed" mechanisms and trajectories towards physically constrained states, the terminology may lead to an underestimation of the explanatory relevance of the latter.[15] Many developmental biologists, including Davidson himself (personal communication), are therefore concerned with the way that this terminology may downplay the explanatory power of physical aspects of development (see also Davidson et al., 2009; Love, 2015). The concern is that physical aspects are mainly taken to *describe* but not *explain*.

This issue is relevant also to debates about explanatory reduction. The reductionist may claim that the higher-level phenomena picked out by mechanical studies, although useful for the analysis, do no genuine explanatory work since the difference-making factors captured by biological explanations are all encoded in the gene regulatory network.[16] Such arguments are dependent on further assumptions about what Hüttemann and Love (2011) call *intrinsicality,* i.e., about the way in which a phenomenon or system is individuated and other aspects regarded as background conditions. For instance, attempts to explain cell functions in terms of molecular mechanisms rely on the cell membrane as a boundary between internal and external (background) conditions. Since background conditions are often taken to play a minor explanatory role (if any), an important question is whether such individuation criteria are justified, or rather reflect local and perhaps idiosyncratic explanatory norms and methodologies.

The explanatory priority of molecular models and explanations may be partly justified through the successful empirical demonstration of genetic difference-making. Appealing to genetic difference-making is, however, insufficient in the context of developmental biology because manipulations of genetic causes typically treat environmental and biomechanical factors as fixed (Brigandt & Love, 2012; Robert, 2004). In Section 2 we argued that measurement and modeling of biophysical parameters at different scales require that different aspects of the system are fixed as boundary conditions. Similarly, Davidson argues that it is not possible to simultaneously measure and manipulate genetic or molecular pathways and physical forces in a similar way to determine their relative influence on the bending behavior and movement of cells:

"The model systems where molecular and cellular manipulations are simplest are some of the most challenging to measure absolute forces or material properties. By contrast, the model systems where tissue-scale forces and material properties can be directly measured are the most challenging to manipulate genetically", (Davidson, 2012, p. 85, p. 85).

[15] Davidson et al. (2009) ponder about whether one can distinguish "active" from "passive" properties in practice as these may not be easily defined or distinguished. The terms do, however, point to an interesting difference in how biological analysis draws on functional concepts that are not as apparent in physics. Discussions about the implications of functional language in biology are, however, beyond the scope of this paper.

[16] For example, Rosenberg (1997) questions the explanatory autonomy of macroscale models and explanations in molecular developmental biology. While acknowledging that factors such as the maternal cellular structure play a causal role in embryogenesis, he highlights that explanations provided by developmental molecular biologists do not include all conditions that would be causally sufficient for the development of the embryo. For further discussion of this issue, see (Kaiser, 2015, Chapter 6).

In other words, because of the tyranny of scales and the complexity of biological systems, modeling of morphogenesis requires either developing models for the tissue scale processes, where many details on genetic and molecular force factors must be ignored, *or* on molecular scales where tissue level parameters must be held constant as boundary conditions. Accordingly, appealing to genetic difference-making is insufficient to dismiss the explanatory relevance of macroscale features.

What would it take to settle this issue in practice? Miller and Davidson (2013) describe the greatest challenge for studies of biomechanics as the difficulty of studying physical forces in the same way as genetic difference-making is studied. In the latter case, individual genes or regulatory circuits can sometimes be 'knocked out' to study their effects, but it is typically not possible to knock out a physical force. However, new experimental techniques afford an examination of physical difference making in biology (Wyczalkowski et al., 2012; Love, forthcoming). Imaging tools such as video or traction force microscopy, confocal time-lapse microscopy, and fluorescent techniques now allow for quantitative measurement of geometry changes and gradient velocities of moving cells (Brodland, Conte, Cranston, & Miodownik, 2010; Davidson, 2012). Isolation and detection of force-generating effects in tissues can also be conducted via laser microdissection or microsurgery (Miller & Davidson, 2013). For instance, a portion of the epithelial layer can be cut with a laser to isolate effects of force transformation on other cells from this layer.[17] Additionally, physical models are increasingly supplemented with advanced computer simulations for studying of trajectories of changes in cell and tissue-shapes (Wyczalkowski et al., 2012). Importantly, experimental designs utilizing these new technologies have revealed that treating cell and tissue mechanics as non-explanatory background conditions is misleading because mechanical cues can directly influence cell differentiation and gene expression through force transmission (Hutson & Ma, 2008; Levayer & Lecuit, 2012; Vogel & Sheetz, 2006; Wozniak & Chen, 2009). We mention just a few examples below.

Many cells respond to mechanical signals from the microenvironment where strain for instance can be transmitted via ECM fibers or sensed by stretch-sensitive channels (Miller & Davidson, 2013). By culturing human mesenchymal stem cells on elastic substrates with controllable stiffness, Engler, Sen, Sweeney, and Discher (2006) were able to direct cells towards osteogenic, neurogenic and myogenic fates with decreasing stiffness. Similarly, *in vivo* studies of fruit fly and amphibian development show how high levels of mechanical strain can trigger cell differentiation (Beloussov & Grabovsky, 2006; Beloussov, Luchinskaya, Ermakov, & Glagoleva, 2006; Brodland et al., 1994). In the developing *Drosophila* embryo, gene expression is initiated by mechanical deformation (Farge, 2003), and *in vitro* studies of human mesenchymal stem cells suggest that certain transcription factors are sensitive to changes in mechanical forces such as stiffness of the growth substrate (Fu et al., 2010). These experiments suggest that biomechanical properties of macroscale structures, such as mechano-transduction through stretching, contraction and compression of tissues, can serve as effective inputs on lower-scale processes and produce measurable effects on gene expression and signaling pathways (see also Desprat, Supatto, Pouille, Beaurepaire, & Farge, 2008; Fernandez-Gonzalez, de Matos Simoes, Röper, Eaton, & Zallen, 2009; Pouille, Ahmadi, Brunet, & Farge, 2009). The dependency of the fate and identity of cells on the macro- and microenvironment, e.g., on the tissue boundaries and stiffness of the substrate, confounds the assumption that initial and boundary conditions are merely background conditions in models and explanations.

Boundary conditions identified through biomechanical modeling, as described in the work of Davidson et al. (1995), can also help address questions about how physical constraints influence biological variation, in extant organisms as well as in future evolutionary transitions. Fig. 6 shows how the different mechanisms of gastrulation are constrained by physical factors defining a space of possible parameter values for the relative stiffness of the cell layer, apical lamina, and hyaline layers. Importantly, the material parameter space illustrated in Fig. 6 does not allow for gradual transitions between these mechanisms; "For example, gradual steps along a trajectory through elastic property space do not allow the gel swelling mechanism to change to the apical constriction mechanism because neither mechanism can generate a sufficiently invaginated gastrula with intermediate elastic moduli" (Davidson et al., 1995, p. 2016). Thus, defining the boundaries of causal possibilities through biomechanical modeling can provide insights to how the mechanical design of embryos constrains possible evolutionary transitions between different developmental mechanisms.

The examples presented also allow us to specify what we mean by indispensability concerning macro-scale features and boundary conditions. Modeling of and experimentation on physical factors, as exemplified in this section, show how biomechanics make a necessary difference to developmental outcomes. Biomechanical modeling in this context stresses that macro-scale parameters such as tissue stiffness are causally and explanatorily indispensable. Attention to boundary conditions helps specify the aspects that make macro-scale features requisite. As we have shown in Sections 2 and 3, imposing limits on the domain of a model or an experiment by holding some properties fixed is often required for solving equations or for intervening on a complex system. The reliance on such strategies reveals interesting aspects of the complexity and scale-dependency of multi-scale systems. Boundary conditions inform about what features of the system are ignored or fixed when investigating processes at characteristic scales, and the requirement to combine models with different boundary conditions reveals the contexts for which such assumptions are no longer feasible. Moreover, boundary conditions are used to represent the constraints imposed by meso- and macroscale structures on the behavior of processes at lower scales. In the following, we elaborate further on how attention to the role of boundary conditions in biomechanical modeling can help specify the functional role of higher-level constraints as top-down effects.

## 4. Understanding living systems across scales

So far, we have argued that the appeal to an ideal physics or molecular biology to provide reductive explanations of upper-scale properties to features at lower scales must confront several challenges. Even when modeling relatively simple physical systems across scales, bottom-up modeling is not feasible (Section 2). Moreover, work on biomechanics in developmental biology underscores the importance of macroscale structures and constraints, rather than appealing to explanations at lower levels (Section 3).

The increasing application of physical science approaches in developmental biology neither supports explanatory reductionism nor diminishes the importance of biological models. Rather, multi-scale modeling projects reveal the requirement to combine different types of models. In the context of multi-scale modeling in nano-science, Bursten (2015) has described the relation between models synthesized in an explanation as a *non-reductive model interaction*. This notion captures how the models intersect at the

[17] An alternative strategy is to fix the tissue on silicone membranes or other deformable substrates for mechanical manipulation (Miller & Davidson, 2013).

point where the boundary conditions employed by one model can no longer be ignored (see also Wilson, 2012). How information from various sources is more specifically integrated in multiscale modeling in biology, and whether an explanatory synthesis can be reached like in examples from physics and nanoscience, will be important topics for future research. As a first step, our aim in this paper is primarily to demonstrate the need for multi-scale modeling strategies and to argue that the relations between the models are non-reductive.

As mentioned in Section 2, continuum models used in developmental biology abstract from and distort many aspects of the microscale properties of discrete cells. However, as illustrated through studies of cell adhesion, it is typically not possible to account for the behavior that dominates at higher scales through measurements and models at the molecular scales (Davidson, 2012). As in the examples with steel and the violin, modeling tissue biomechanics requires treating some microscale details as boundary conditions. Attention to boundary conditions can help clarify which details are considered explanatorily irrelevant for specific modeling tasks. To reach a full explanation the researcher must employ models that capture different aspects. Recall, for instance, the importance of the details at the boundaries for hearing the violin or how the production of specific morphogens can impact cell differentiation (Section 2). In the latter case, the modeler must incorporate results from microscale models that can account for kinetic details of specific molecular processes (Newman et al., 2011). Such models, in contrast, treat many biomechanical properties as fixed and do not account for how the system as a whole changes over time. Accordingly, modelers must find ways to employ different models so as to bridge between different processes at different temporal and spatial scales. Moreover, in the context of multi-scale modeling in biology, gaps must also be bridged between physical and biological models.

How is this gap bridged in practice? As mentioned in Section 2, cell adhesion plays an important role in force-transmission and mechano-transduction. Accordingly, Wyczalkowski et al. (2012, p. 132) stress that: "Biomechanical forces are the bridge that connects genetic and molecular-level to tissue-level deformations that shape the developing embryo". The connection is established via models of meso-scale structures. Noble (2012) refers to the modeling approach as a 'middle-out' approach. In the context of modeling of steel, a middle-out approach involves treating crack, voids, and pearlitic inclusions as structures that dominate at scales intermediate between the continuum and the atomic. Some biological modelers have similarly stressed the importance of studying structures at levels intermediate between genes/molecules and tissue. The most important intermediate structure, quite naturally, is that of the cell (or a block of cells with similar properties) and their relationships to the extra cellular matrix. Cell-centered modeling thus establishes a connection to tissue-scale deformations while allowing for a kind of "black-boxing" of lower scale genetic and chemical features by treating these as boundary conditions or constraints on cell behaviors. Merks and Glazier argue that:

"[T]he cell provides a natural level of abstraction for mathematical and computational modeling of development. Treating cells phenomenologically immediately reduces the interactions of roughly $10^5$-$10^6$ gene products to 10 or so behaviors: cells can move, divide, die, differentiate, change shape, exert forces, secrete and absorb chemicals and electrical charges, and change their distribution of surface properties", (Merks & Glazier, 2005, p. 117, p. 117).

Furthermore, as noted above, it would be a mistake to focus only tissue behaviors:

"Ignoring cells is dangerous. Macroscopic models, which treat tissues as continuous substances with bulk mechanical properties … reproduce many biological phenomena but fail when structure develops or functions at the cell scale. Although continuum models are computationally efficient for describing non-cellular materials like bone, extracellular matrix (ECM), fluids and diffusing chemicals, many cell-centered models reproduce experimental observations missing from continuum models", (Merks & Glazier, 2005, p. 118, p. 118).

One must therefore ask what is the appropriate scale or level to begin to address developmental modeling. In practice, modelers often start at the cellular scale and "feed forward" (or "up") to the level of tissue and ultimately to the level of the organ or organism. To bridge between stochastic models at the molecular scale and continuum models at tissue scale researchers for instance rely on lattice-gas automaton models that mimic cell movement via simple migration and interaction rules between cells (see Fig. 7; Hatzikirou & Deutsch, 2011). Another example of cell-based modeling is the Subcellular Element Model (Newman et al., 2011). These models include dynamics at the smaller scale in a coarse-grained manner, but they are also dependent on parameters that set boundary conditions for environmental influences. Depending on the stability of the environment and tissue deformations, cell models can efface environmental dynamics as a static boundary condition or incorporate a dynamic environment (e.g. by drawing on a matrix describing a vector field).

As mentioned above, connecting the models does not involve a process of reduction of one model to a more fundamental one. The models combined in multi-scale modeling projects like the ones examined are *explanatorily independent* but *epistemologically interdependent* (Potochnik, 2009, p. 19). They are explanatorily independent in the sense that processes at different scales must be modeled by different, and often conflicting, theoretical frameworks. The combination of models forms a pluralistic mosaic of different strategies rather than a reductive explanation (see Fig. 7). However, when models are combined in complex multi-scale simulations, they are not completely autonomous. They are epistemologically interdependent in the sense that the success of the patchwork of explanatory models, and also the solution of specific mathematical models, often depend on sources of information that are represented by another model. Feedback between models in terms of model inputs, e.g., as boundary conditions and inputs to meso-scale models, imposes a self-consistent scheme on modeling across scales.

### 4.1. Boundary conditions and top-down effects

The identification of boundary conditions at higher scales can also feed into models at lower scales and give a concrete interpretation of top-down effects. To make sense of the feedback across different scales, modelers typically divide up the system into complex processes at different scales as pictured on Fig. 8 (see also Noble, 2012). Gene expression patterns influence the availability and frequency of proteins (such as cell adhesion molecules) that again influence cell differentiation, cell-cell interactions, intermolecular force production, surface tension variation, and resulting phase separations that give rise to tissue deformations. Similarly, molecular signaling affects the intra- and extracellular force production that can alter cell-shape changes and lead to cell movement, cross-scale signaling, and ultimately to morphological changes. However, the view that the molecular scale has explanatory priority is challenged by how force propagation in tissues and cells is continuously fed back into the molecular level, and how biomechanical factors regulate levels of gene expression through

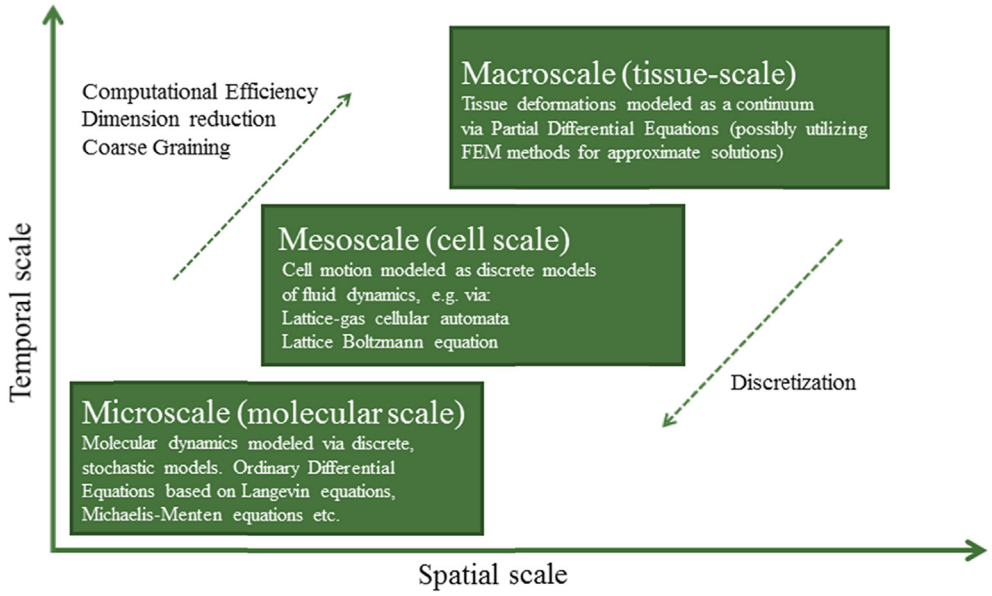


**Fig. 7.** Models used at different temporal and spatial scales in developmental biology. Figure drawn with inspiration from (Hatzikirou & Deutsch, 2011; Lesne, 2013; Newman et al., 2011).

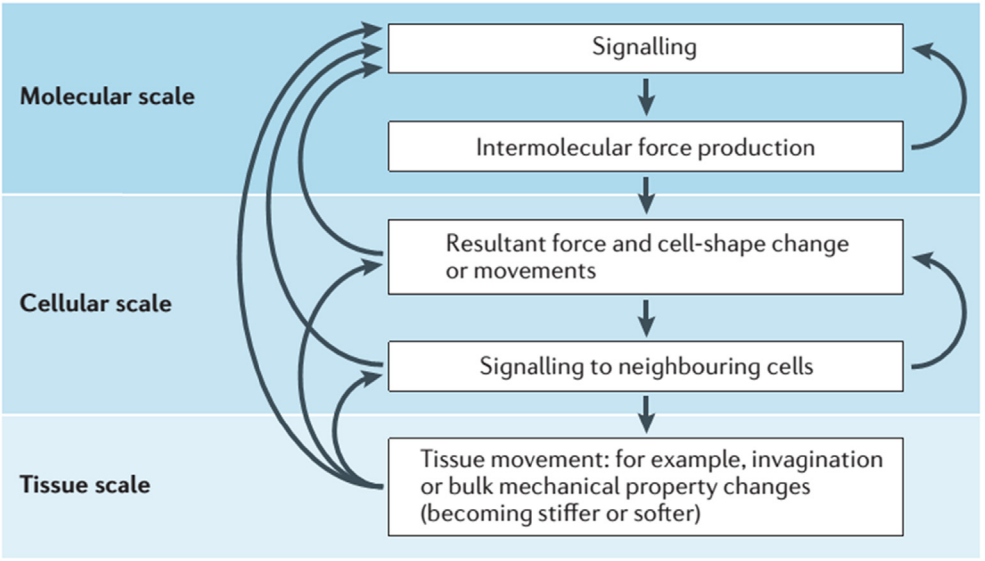


**Fig. 8.** Interplay and feedback between processes at different scales, e.g. between signaling at the cellular scale and mechanical property changes at tissue scale. Reprinted by permission from Macmillan Publishers Ltd: Nature Review Genetics (Miller and Davidson, 2013), copyright (2013).

biomechanical constraining relations and initial conditions defined by the microenvironment (Section 3.2).

Influences going from the scale of tissues to the molecular level are often described through metaphysically challenging concepts such as 'top-down' or 'downward' causation, and the feedback relations are sometimes described as 'reciprocal', 'circular' or 'mutual' causation (Lesne, 2013). It is not our aim in this paper to engage in a philosophical discussion about whether such top-down effects are causal or constitutive (see e.g., Craver & Bechtel, 2007; Kaiser, 2015). Rather, we aim to clarify how attention to the role of boundary conditions in multi-scale modeling allows for a more concrete interpretation of top-down effects.

Section 3 highlighted how cell differentiation and cell motility not only depend upon gene regulation and cell signaling but also on the material properties of the cells and tissues. Boundary conditions in this context represent the constraints on the bending behavior and deformations given by the physical properties of the biological structures and the environmental context of the cell. An additional example is how developmental mechanisms in fish and frog embryos are dependent on the ability of the notochord 'backbone' (that later provides attachment for skeletal muscles) to spatially straighten the embryo. This phenomenon cannot be explained solely at the molecular level. Davidson et al. (2009, p. 2017) stress that "[t]he capacity of the notochord to resist bending as it extends the embryo comes from the structure of the whole notochord. Measurements at the level of the individual collagen fiber or fluid-filled cell that make up the structure would not reveal the mechanical properties of the whole notochord". Biomechanical

approaches to development thus stress the importance of spatial organization and the influence of system-level constraints on the behavior of processes at lower scales. These are often represented as boundary conditions for lower-scale models.

Macroscale parameters are often needed to account for higher-level constraints that are ignored in studies of micro-scales processes. With the example of studies of adhesion across scales (Section 2), we highlighted that measurements at the molecular or cell scale cannot account for mechanical properties at the tissue scale because the analysis removes molecules and cells from the context of the system as a whole. Tissue-scale parameters are therefore needed as boundary conditions that constrain the dynamical modeling of microscale processes. Boundary conditions play a similar role in multi-scale cardiac modeling. Noble highlights that phenomena like cardiac rhythms and fibrillations cannot be understood or modeled at the level of proteins and DNA alone because these phenomena only exist because of the (productive and limiting) constraints of cell and tissue structures. He expresses the notion of downward causation mathematically as "the influences of initial and boundary conditions on the solutions of the differential equations used to represent the lower level processes" (Noble, 2012, p. 55). The heart rhythm cannot be modeled bottom-up from ionic current models because solving the equations requires boundary conditions (e.g., cell voltage) provided by action potential models.[18] Similarly, if the attempt is to develop simulations of the whole heart, molecular and cell-scale models need boundary conditions from propagation models (partial differential equations) that incorporate data from optimal mapping of tissue structures (Carusi, Burrage, & Rodríguez, 2012).

Multi-scale modeling of biological systems thus highlights how researchers need to take into account the ways in which tissue structures productively constrain the degrees of freedom of lower-scale processes such as gene expression, cell differentiation, cell proliferation, and cell movement. This aspect has recently also been emphasized as relevant for cancer research, where an increasing group of scholars argues that tumor development should be understood as an abnormal developmental process enabled by alteration of tissue constraints (Nelson & Bissel, 2006; Shawky & Davidson, 2015). Reporting on simulation results from modeling of chick embryos, Newman et al. (2011, p. 162) for instance observe that: "Repeated growth and division will lead to the formation of either epithelial sheets or spherical tumor-like cell clusters, depending on the boundary conditions". Some cancer researchers have recently argued that since stiffness of the tissue and of the ECM has been shown to influence tumor development, macroscale biomechanical factors are relevant not only for understanding cancer but also for designing treatments (see also Paszek et al., 2005; Bizzarri & Cucina, 2014). Biological research is still at an early stage of investigating the details and significance of such interactions, but it is a research area in rapid development with important philosophical and scientific implications.

## 5. Concluding remarks

It seems intuitive that if a biological system is composed of nothing but physical components it should be possible to derive macroscale properties "bottom-up" from molecular descriptions. From this picture, macroscale properties seem to follow from molecular descriptions, in an explanatory as well as an ontological sense. The simple steel example (Section 2) offers resistance to this view because the dynamics of the system at macroscale that are relevant to engineers *cannot* be derived from the structural models of the atomic lattice structures. This conclusion may be surprising given that physics is often associated with reductionism in the context of biology. Yet, examples from multi-scale modeling in both physics and biology show that modelers in both domains must confront the tyranny of scales problem. There is no single approach that can account for all relevant aspects of multi-scale systems.

We have illustrated how the scale-dependency of physical and biological behaviors forces modelers to combine different mathematical models relying on different boundary conditions. By examining the role of biomechanics in multi-scale modeling of morphogenesis in developmental biology, we have described the relation between models at different scales and from different disciplines as explanatorily independent and epistemologically interdependent (Potochnik, 2009, p. 19). The models are explanatorily independent in the sense that they describe different processes at characteristic scales while drawing on different (and often conflicting) theoretical frameworks. At the same time, the models are interdependent in the sense that the success of one model depends on sources of information that are not explicitly represented but covered via other models or sources of information. Multi-scale modeling can therefore shed light on how processes at different scales are interconnected.

The examples we have discussed also shed new light on the role of physical science approaches in biology. The role of biomechanical modeling in developmental is very different from the expectation that physical science approaches target a lower and more fundamental level than biological models and explanations. The examples show that morphological changes in the developing embryo are not only regulated genetically (bottom-up) but also by the capacity of the tissue as a whole to generate force and maintain tissue stiffness. In this context, the input from physics does not support explanatory reduction of higher to lower scales. Rather, the examples provide insight to how macroscale parameters influence or set boundaries for causal processes at lower scales.

Inputs from physics may also help identify conditions that make some system-level processes relatively independent of molecular details. Intuitively, the explanatory power of models would increase as the number of molecular details increase. But this intuition is often misguided. In many instances, macroscale behaviors are not dependent on specific atomic or molecular details. Accounts arguing for the explanatory priority of lower-scale descriptions fail to explain why macro-scale explanation have *explanatory autonomy*, i.e., why models and explanations are successful despite ignoring many microscale details (cf., Batterman, 2016). Moreover, the examples outlined in this paper suggest that many developmental processes, including gene regulation and cell differentiation, are directly influenced by meso-and macro-scale biomechanical parameters. Failing to account for environmental and systemic constraints on lower-scale processes often result in a failure to understand the functionality of the system (Noble, 2012). The requirement of boundary conditions to represent such top-down influences may thus provide a concrete interpretation of top-down effects. Taken together, these aspects provide resistance to the view that macroscale properties are dispensable for explaining multi-scale biological systems.

## Acknowledgements

We would like to thank William Bechtel, Fridolin Gross, Maria Serban, Nicholaos Jones, and two anonymous reviewers for very useful comments to an earlier version of this paper. We also thank

[18] The parameters required for identifying such constraints are not possible to measure and model at the molecular scale. Isolation procedures that use the voltage-clamp technique to study individual ion channels neglect the context of the cell as a whole, and micro-scale models must be calibrated via microelectrode measurements of the cell voltage (a cell parameter).

Alan Love for bringing the work of Lance Davidson to our attention. R. B.'s work was supported by a grant from the John Templeton Foundation.

## References


Bassingthwaighte, J., Hunter, P., & Noble, D. (2009). The cardiac Physiome: Perspectives for the future. *Experimental Physiology, 94*(5), 597–605.

Batterman, R. W. (2012). The tyranny of scales. In R. W. Batterman (Ed.), *Oxford handbook of philosophy of physics* (pp. 255–286). Oxford: Oxford University Press.

Batterman, R. W. (2016). Autonomy of theories. An exlanatory problem. *Nous*. http://dx.doi.org/10.1111/nous.12191.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Beloussov, L. V., & Grabovsky, V. I. (2006). Morphomechanics: Goals, basic experiments and models. *International Journal of Developmental Biology, 50*, 81–92.

Beloussov, L. V., Luchinskaya, N. N., Ermakov, A. S., & Glagoleva, N. S. (2006). Gastrulation in amphibian embryos, regarded as a succession of biomechanical feedback events. *International Journal of Developmental Biology, 50*(2/3), 113–122.

Bertalanffy, L. v (1969). *General system theory. Foundations, development, applications*. New York: George: Braziller.

Bizzarri, M., & Cucina, A. (2014). Tumor and the microenvironment: A chance to reframe the paradigm of carcinogenesis? *BioMed Research International, 2014*(3). Article 934038.

Boogerd, F. C., Bruggeman, F., Richardson, R., Achim, S., & Westerhoff, H. V. (2005). Emergence and its place in nature: A case study of biochemical networks. *Synthese, 145*, 131–164.

Brigandt, I., & Love, A. (2012). *Reductionism in biology. Stanford encyclopaedia of philosophy*. http://stanford.library.sydney.edu.au/entries/reduction-biology/.

Brodland, G. W., Conte, V., Cranston, P. G., & …Miodownik, M. (2010). Video force microscopy reveals the mechanics of ventral furrow invagination in drosophila. *Proceedings of the National Academy of Sciences of the United States of America, 107*(51), 22111–22116.

Brodland, G. W., Gordon, R., Scott, M. J., Björklund, N. K., Luchka, K. B., Martin, C. C., & …Shu, D. (1994). Furrowing surface contraction wave coincident with primary neural induction in amphibian embryos. *Journal of Morphology, 219*(2), 131–142.

Burian, R., Richardson, R. C., & Van der Steen, W. J. (1996). Against Generality: Meaning in genetics and philosophy. *Studies in History and Philosophy of Science, 27*(1), 1–29.

Bursten, J. (2015). *Surfaces, scales and Synthesis: Reasoning at the nanoscale*. Dissertation. University of Pittsburgh.

Carusi, A., Burrage, K., & Rodríguez, B. (2012). Bridging experiments, models and simulations: An integrative approach to validation in computational cardiac electrophysiology. *American Journal of Physiology - Heart and Circulatory Physiology, 303*(2), H144–H155.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology & Philosophy, 22*(4), 547–563.

Davidson, L. A. (2012). Epithelial machines that shape the embryo. *Trends in Cell Biology, 22*(2), 82–87.

Davidson, L. A., Koehl, M. A., Keller, R., & Oster, G. F. (1995). How do sea urchins invaginate? Using biomechanics to distinguish between mechanisms of primary invagination. *Development, 121*(7), 2005–2018.

Davidson, L. A., Oster, G. F., Keller, R. E., & Koehl, M. A. R. (1999). Measurements of mechanical properties of the blastula wall reveal which hypothesized mechanisms of primary invagination are physically plausible in the sea urchin strongylocentrotus purpuratus. *Developmental Biology, 209*(2), 221–238.

Davidson, L. A., von Dassow, M., & Zhou, J. (2009). Multi-scale mechanics from molecules to morphogenesis. *The International Journal of Biochemistry & Cell Biology, 41*(11), 2147–2162.

Desprat, N., Supatto, W., Pouille, P., Beaurepaire, E., & Farge, E. (2008). Tissue deformation modulates twist expression to determine anterior midgut differentiation in drosophila embryos. *Developmental Cell, 15*(3), 470–477.

Dunn, A., & Price, A. (2015). Energetics and forces in living cells. *Physics Today, 68*(2), 27–32.

Dupré, J. (1993). *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, Mass: Harvard University Press.

Engler, A. J., Sen, S., Sweeney, H. L., & Discher, D. E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell, 126*(4), 677–689.

Farge, E. (2003). Mechanical induction of twist in the drosophila foregut/stomodeal primordium. *Current Biology, 13*(16), 1365–1377.

Fernandez-Gonzalez, R., de Matos Simoes, S., Röper, J., Eaton, S., & Zallen, J. A. (2009). Myosin II dynamics are regulated by tension in intercalating cells. *Developmental Cell, 17*(5), 736–743.

Fu, J., Wang, Y., Yang, M. T., Desai, R. A., Yu, X., Liu, Z., et al. (2010). Mechanical regulation of cell function with geometrically modulated elastomeric substrates. *Nature Methods, 7*(9), 733–736.

Green, S. (2015). Can biological complexity be reverse engineered? *Studies in History and Philosophy of Biological and Biomedical Sciences, 53*, 73–83.

Hatzikirou, H., & Deutsch, A. (2011). Cellular automata as microscopic models of cell migration in heterogeneous environmens. In G. P. Schatten, S. Schnell, P. Maini, S. A. Newman, & T. Newman (Eds.), *Multiscale modeling of developmental systems* (pp. 311–340). San Diego: Academic Press.

Hooker, C. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science, 18*(4), 757–780.

Hutson, M. S., & Ma, X. (2008). Mechanical aspects of developmental biology: Perspectives on growth and form in the (post)-genomic age. *Physical Biology, 5*(1), 015001.

Hüttemann, A., & Love, A. C. (2011). Aspects of reductive explanation in biological science: Intrinsicality, fundamentality, and temporality. *The British Journal for the Philosophy of Science, 62*, 519–549.

Hüttemann, A., & Love, A. C. (2016). Reduction. In P. Humphreys (Ed.), *The Oxford handbook of philosophy of science* (pp. 460–483). New York, NY: Oxford University Press.

Kaiser, M. I. (2015). *Reductive explanation in the biological sciences*. Springer International Publishing Switzerland.

Lesne, A. (2013). Multiscale analysis of biological systems. *Acta Biotheoretica, 61*(1), 3–19.

Levayer, R., & Lecuit, T. (2012). Biomechanical regulation of contractility: Spatial control and dynamics. *Trends in Cell Biology, 22*(2), 61–81.

Love, A. Combining genetic and physical causation in developmental explanations. In C. K. Waters, & J. Woodward (Eds.), *Causal reasoning in biology*. Minneapolis: University of Minnesota Press (forthcoming).

Love, A. (2015). *Developmental biology. Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/biology-developmental/.

Love, A. C., & Hüttemann, A. (2011). Comparing part-whole explanations in biology and physics. In D. Dieks, W. J. Gonzalez, S. Hartmann, T. Uebel, & M. Weber (Eds.), *Explanation, prediction, and confirmation* (pp. 183–202). Berlin: Springer.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Mayr, E. (1988). *Toward a new philosophy of biology: Observations of an evolutionist*. Cambridge, MA: Harvard University Press.

Mayr, E. (2004). *What makes biology unique? Considerations on the autonomy of a scientifi c discipline*. New York: Cambridge University Press.

Merks, R. M., & Glazier, J. A. (2005). A cell-centered approach to developmental biology. *Physica A: Statistical Mechanics and its Applications, 352*(1), 113–130.

Miller, C. J., & Davidson, L. A. (2013). The interplay between cell signalling and mechanics in developmental processes. *Nature Reviews Genetics, 14*(10), 733–744.

Morange, M. (2011). Recent opportunities for an increasing role for physical explanations in biology. *Studies in History and Philosophy of Biological and Biomedical Research, 42*, 139–144.

Nelson, C. M., & Bissell, M. J. (2006). Of extracellular matrix, scaffolds, and signaling: Tissue architecture regulates development, homeostasis, and cancer. *Annual review of cell and developmental biology, 22*, 287.

Newman, S. A., Christley, S., Glimm, T., Hentschell, H. G. E., Kazmierczak, B., Zhang, Y.-T., et al. (2011). Multiscale models for vertebrate limb development. In G. P. Schatten, S. Schnell, P. Maini, S. A. Newman, & T. Newman (Eds.), *Multiscale modeling of developmental systems* (pp. 311–340). San Diego: Academic Press.

Noble, D. (2012). A theory of biological relativity: No privileged level of causation. *Interface Focus, 2*(1), 55–64.

Oden, J. T. (2006). *Finite elements of nonlinear continua*. New York: Dover Publications.

Oppenheim, P., & Putnam, H. (1958). The unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind-body problem* (pp. 3–36). Minneapolis: University of Minnesota Press.

Paszek, M. J., Zahir, N., Johnson, K. R., Lakins, J. N., Rozenberg, G. I., Gefen, A., & …Hammer, D. A. (2005). Tensional homeostasis and the malignant phenotype. *Cancer Cell, 8*(3), 241–254.

Peter, I. S., & Davidson, E. H. (2015). *Genomic control process, development and evolution*. Oxford: Academic Press, Elsevier.

Potochnik, A. (2009). Explanatory independence and epistemic interdependence: A case study of the optimality approach. *British Society for the Philosophy of Science, 61*, 213–233.

Pouille, P. A., Ahmadi, P., Brunet, A. C., & Farge, E. (2009). Mechanical signals trigger myosin II redistribution and mesoderm invagination in drosophila embryos. *Science Signaling, 2*(66), 16.

Purcell, E. M. (1977). Life at low Reynolds number. *American Journal of Physics, 45*(1), 3–11.

Qu, Z., Garfinkel, A., Weiss, J. N., & Nivala, M. (2011). Multi-scale modeling in biology: How to bridge the gaps between scales? *Progress in Biophysics and Molecular Biology, 107*(1), 21–31.

Rast, J. P., Smith, L. C., Loza-Coll, M., Hibino, T., & Litman, G. W. (2006). Genomic insights into the immune system of the sea urchin. *Science (New York, N.Y.), 314*(5801), 952–956.

Robert, J. S. (2004). *Embryology, epigenesis, and evolution: Taking development seriously*. New York: Cambridge University Press.

Rosenberg, A. (1997). Reductionism redux: Computing the embryo. *Biology and Philosophy, 12*(4), 445–470.

Rosenberg, A., & Arp, R. (2010). *Philosophy of biology: An anthology*. Chichester, U.K: Wiley-Blackwell.

Sarkar, S. (1998). *Genetics and reductionism*. Cambridge: Cambridge University Press.

Schaffner, K. (1993). Theory structure, reduction, and disciplinary integration in biology. *Biology and Philosophy, 8*(3), 319–347.
Shawky, J. H., & Davidson, L. A. (2015). Tissue mechanics and adhesion during embryo development. *Developmental Biology, 401*(1), 152–164.
Skillings, D. J. (2015). Mechanistic explanation of biological processes. *Philosophy of Science, 82*(5), 1139–1151.
Vogel, S. (2009). *Glimpses of creatures in their physical worlds*. Princeton, NJ: Princeton University Press.
Vogel, V., & Sheetz, M. (2006). Local force and geometry sensing regulate cell functions. *Nature Reviews Molecular Cell Biology, 7*(4), 265–275.
Weber, M. (2008). Causes without mechanisms: Experimental regularities, physical laws, and neuroscientific explanation. *Philosophy of Science, 75*, 995–1007.
Wilson, M. (2012). What is classical mechanics anyway? In R. Batterman (Ed.), *Oxford handbook of philsophy of physics* (pp. 43–106) Oxford: Oxford University Press.
Winther, R. G. (2009). Schaffner's model of theory reduction: Critique and reconstruction. *Philosophy of Science, 76*(2), 119–142.
Wozniak, M. A., & Chen, C. S. (2009). Mechanotransduction in development: A growing role for contractility. *Nature Reviews Molecular Cell Biology, 10*(1), 34–43.
Wyczalkowski, M. A., Chen, Z., Filas, B. A., Varner, V. D., & Taber, L. A. (2012). Computational models for mechanics of morphogenesis. *Birth Defects Research Part C: Embryo Today: Reviews, 96*(2), 132–152.

frontiers in Physiology

HYPOTHESIS AND THEORY
published: 25 January 2019
doi: 10.3389/fphys.2018.01966

# The Dynamical Emergence of Biology From Physics: Branching Causation via Biomolecules

*George F. R. Ellis*[1]* *and Jonathan Kopel*[2]

[1] Mathematics Department, University of Cape Town, Cape Town, South Africa, [2] Texas Tech University Health Sciences Center (TTUHSC), Lubbock, TX, United States

Biology differs fundamentally from the physics that underlies it. This paper[1] proposes that the essential difference is that while physics at its fundamental level is Hamiltonian, in biology, once life has come into existence, causation of a contextual branching nature occurs at every level of the hierarchy of emergence at each time. The key feature allowing this to happen is the way biomolecules such as voltage-gated ion channels can act to enable branching logic to arise from the underlying physics, despite that physics *per se* being of a deterministic nature. Much randomness occurs at the molecular level, which enables higher level functions to select lower level outcomes according to higher level needs. Intelligent causation occurs when organisms engage in deduction, enabling prediction and planning. This is possible because ion channels enable action potentials to propagate in axons. The further key feature is that such branching biological behavior acts down to cause the underlying physical interactions to also exhibit a contextual branching behavior.

**Keywords: hierarchy of emergence, bio-molecules, top-down causation, branching logic, natural selection, voltage-gated ion channels**

OPEN ACCESS

***Edited by:***
*Matteo Mossio, UMR8590 Institut d'Histoire et de Philosophie des Sciences et des Techniques (IHPST), France*

***Reviewed by:***
*Kepa Ruiz-Mirazo, University of the Basque Country, Spain*
*Giuseppe Longo, Center for the National Scientific Research (CNRS), France*

***Correspondence:**
*George F. R. Ellis*
*gfrellis@gmail.com*

***Specialty section:***
*This article was submitted to Integrative Physiology, a section of the journal Frontiers in Physiology*

***Received:*** *31 August 2018*
***Accepted:*** *31 December 2018*
***Published:*** *25 January 2019*

***Citation:***
*Ellis GFR and Kopel J (2019) The Dynamical Emergence of Biology From Physics: Branching Causation via Biomolecules. Front. Physiol. 9:1966. doi: 10.3389/fphys.2018.01966*

## 1. BIOLOGY VS. PHYSICS

Biology arises out of the underlying physics, but living systems have an essentially different nature than natural systems because *inter alia* they involve purpose or function (Hartwell et al., 1999), information (Nurse, 2008), organization (Mossio et al., 2016), and variation (Montévil et al., 2016). How do they arise from the underlying physics, which has none of these characteristics? Physics and biology are essentially different, even though physics underlies biology. We will identify the physics-biology difference, once life has come into existence, as being due to the fact that biological causation is based at the cellular level in logical branching shaped by context, enabled in physical terms by the nature of particular proteins. Because this branching is controlled in a top down way by physiological conditions (Noble, 2008, 2012, 2016) this leads to contextual emergence (Atmanspacher and beim Graben, 2009), which is a form of strong emergence, enabling branching behavior to also emerge at the higher levels.

[1] An abbreviated version of the proposal made here appears in a book chapter Ellis and Kopel (2018). This version is extensively revised and extended to consider further areas in integrative physiology.

## 1.1. The Nature of Physics

Physics deals with laws expressing the inevitable interactions of matter and fields according to boundary and initial conditions, and their consequences for emergent physical systems such as gases, liquids, crystals, rocks, planets, stars, and galaxies.

Classical physics proceeds in a deterministic fashion, described by Hamiltonian dynamics (section 2.1). The interactions proceed in a remorseless impersonal way as described by these laws, with no hint of function or purpose. They can exhibit branching behavior in phase changes, as discussed in section 2.2 below, but there is again no trace of purpose or choice in that behavior. Quantum physics has a branching behavior, but that again is nothing to do with choice or function: it is to do with irreducible randomness of quantum outcomes (section 2.3).

When applied to large collections of particles, statistical physics emerges from these interactions and describes how ensembles of particles behave (Penrose, 1979; Blundell and Blundell, 2008). This gives constraints on biology (England, 2013; Perunov et al., 2016) which are necessary, but are not sufficient in themselves to explain function or purpose as in section 1.2.

## 1.2. The Nature of Biology

Many characterizations of life have been given. They include,

- All life exhibits function or purpose (Hartwell et al., 1999), as discussed in the next section.
- In order that this can arise, there must be organization (Solms and Friston, 2018) in the form of adaptive modular hierarchical structures (Ellis, 2016).
- As well as bottom up emergence of higher level structures and function in that hierarchy, there must be top-down realization of higher level processes (Noble, 2012, 2016; Ellis, 2016; Flack, 2017), enabling same level causation at each level (Noble, 2012) and closure of constraints (Mossio and Moreno, 2010; Montévil and Mossio, 2015), with processes thereby generating their own constraints with a mutual dependence such that they both depend on and contribute to maintaining each other.
- This is all enabled by information flows (Nurse, 2008) and associated cell signaling (Berridge, 2014).
- Adaptation to context is taking place all the time at all levels of the hierarchy through variation and selection (Ellis, 2016; Solms and Friston, 2018)
- In particular it is through evo-devo processes (Carroll, 2005; Müller, 2007; Gilbert and Epel, 2009) that all levels of physiological systems come into being, once life has begun[2].
- These processes have a very noisy and contingent nature at the lower levels (Montévil et al., 2016), despite which reliable physiological functioning emerges at higher levels (Rhoades and Pflanzer, 1989; Randall et al., 2002).

[2]We do not attempt to deal in this article with the vexed issue of how life started in the first place. Thus we do not discuss how compartmentalization, metabolism, or adaptive selection and associated genetic information came into being. We assume that they are already in place, and propose that our discussion is then a valid representation of the difference between physics and biology in that context.

As summarized by Hartwell et al. (1999):

> "*Although living systems obey the laws of physics and chemistry, the notion of function or purpose differentiates biology from other natural sciences. Organisms exist to reproduce, whereas, outside religious belief, rocks and stars have no purpose. Selection for function has produced the living cell, with a unique set of properties that distinguish it from inanimate systems of interacting molecules. Cells exist far from thermal equilibrium by harvesting energy from their environment. They are composed of thousands of different types of molecule. They contain information for their survival and reproduction, in the form of their DNA*."

To make this happens involves *inter alia* multiple interactions and non-linearities, the coupling of self-assembly and self-organization processes with chemical/metabolic reactions, existence of cyclic networks, modular/hierarchical substructures, compartmentalization, and cellular individualization.

Finally, what is life? Our view will be (cf. Hartwell et al., 1999) that a living system is a material system that exhibits all the characteristics just listed. From now on we will take that for granted.

## 1.3. The Concept of Function

Functional talk is a contested area in the philosophy of biology (Millikan, 1989; Neander, 1991; Amundson and Lauder, 1994; Godfrey-Smith, 1994).It is discussed in depth by Mossio et al. (2009). One cannot sensibly talk about physiology of living systems without talking about function or purpose (Hartwell et al., 1999): the heart exists in order to circulate blood (Randall et al., 2002, p. 476–510), pacemaking cells exist in order to determine the rhythm of the heart, blood exists in order to transport oxygen, mitochondria in eukaryotes provide energy for cell processes by converting sugars to ATP (Randall et al., 2002, p. 74), and so on Rhoades and Pflanzer (1989). This crucial role of many functions is taken for granted by working biologists, as in the Hartwell et al. quote above. We will have in mind below functions that are indeed crucial in enabling survival (e.g., the pumping of blood by the heart), and not just incidental byproducts (e.g., the sound the heart makes while pumping).

This amounts to a physiological definition however, another tradition exists that relates function to its evolutionary origin. Mossio et al. Mossio et al. (2009) state

> "*A first tradition, usually labeled 'etiological', has tried to justify and naturalize the teleological dimension of functions by appealing to a scientifically acceptable causal explanation. In the mainstream formulation, etiological approaches appeal to a historical selective causal process, through which the existence of current functional traits is the consequence of the selection exerted on the effects of previous occurrences of the trait. A second tradition, called 'systemic' or 'dispositional', discards the teleological dimension of functional attributions as a relevant explanandum by interpreting functions as causal means-end relations at work in a system. From this second perspective, functions do not explain the existence of the bearer; they refer to current contributions of functional traits to some capacity of the system to which they belong.*"

In our view it is crucial to define function in terms of physiological concepts (the dispositional view) rather than evolutionary ones (the etiological view ), because if one goes the latter route it is not easily possible to discuss the issue of drift raised in Kimura (1983) and discussed in depth in Nei (2005). We return to this in section 5.2.

After discussing the options in depth, Mossio et al. (2009) in effect go this route. They propose an organizational account (OA) of functions, as follows:

> *"According to the OA, a trait type T has a function if, and only if, it is submitted to organizational closure C in a differentiated self-maintaining system S. This definition implies the fulfillment of three different conditions. Accordingly, a trait T has a function if and only if:*
>
> *C1. T contributes to the maintenance of the organization O of S;*
>
> *C2. T is produced and maintained under some constraints exerted by O;*
>
> *C3. S is organizationally differentiated."*

If such a trait exists, its function will tend to lead to evolutionary success and hence to selection for this trait, which will explain its existence (up to the issue of drift).

We will adopt this account of functions in what follows. Three further points arise: First, it is crucial that function exists at each level of the hierarchy in interrelated ways, as discussed by Farnsworth et al. (2017). They consider a *function* to describe a process (an action) and a *trait* to be a property of a biological system at one level which may enable a function to be performed in relation to another level. This is consistent with the above. Second, the organizational closure mentioned is conditional on top-down constraint or realization occurring as well as bottom-up emergence in the modular hierarchy (Noble, 2012, 2016; Ellis, 2016)[3]This is again implicit in the above.

Finally, the above does not necessarily imply consciousness or intention. However, intention does indeed come into play in the case of conscious animals, when purposive behavior (Mayr, 2004, p. 57), perhaps including deductive causation (section 6), occurs. Its emergence is based on the reliable functioning of the underlying physiological systems in the brain (Randall et al., 2002; beim Graben, 2016). We discuss this in section 6.

## 1.4. The Key Problem

The issue we address in this paper is thus, how does purpose or function emerge from purposeless physics on developmental and functional timescales? How does deterministic physics lead to logical branching enabling function?

At the macro level, this occurs through plastic neural networks (Kandel et al., 2013) and physiological systems (Rhoades and Pflanzer, 1989). At the micro level, it occurs through epigenetic effects (Pigliucci and Müller, 2000; Gilbert and Epel, 2009) mediated by gene regulatory networks (Gilbert and Epel, 2009) and signal transduction networks (Janes and Yaffe, 2006; Berridge, 2014), and synaptic interactions (Kandel et al., 2013). But at the underlying physical level, dynamics is Hamiltonian and does not allow a branching evolution depending on context. How are these compatible with each other? The theme of this paper is that biomolecules are the key enabling these branching processes to happen. They enable turning molecular processes ON or OFF depending on cell signals, (Berridge, 2014), which is determined by the context in which they exist (Noble, 2008, 2011, 2016). As described by Berridge in the Introduction to *Cell Signaling Biology* Berridge (2014):

> *"The basic principle of cell signaling pathways are that stimuli (e.g., hormones, neurotransmitters or growth factors) acting on cell-surface receptors relay information through intracellular signaling pathways that can have a number of components. They usually begin with the activation of transducers that use amplifiers to generate internal messengers that either act locally or can diffuse throughout the cell. These messengers then engage sensors that are coupled to the effectors that are responsible for activating cellular responses. .. cell signaling is a dynamic process consisting of ON mechanisms during which information flows down the pathway, opposed by the OFF mechanisms that switch off the different steps of the signaling pathway"* (See Module 1: Figure cell signaling mechanism.)

This is an example of the kind of contextual branching that takes place in biology (section 3.1) and distinguishes it from physics.

Note that this is not the same as saying that biological processes can be considered as computational processes, because it is not implying there is a computation or program of some kind determining the branching choices that are made[4]. It is saying that the branching processes which take place at the lower levels, controlled by a large number of cell signaling processes discussed in depth in Berridge's magisterial text (Berridge, 2014), can be regarded to a very good approximation as Boolean (digital) choice processes governed in a top-down contextual way according to functional need. Thus the core of his discussion is how signals turn a large variety of processes ON and OFF. This is a digital logic, emerging from the underlying physics, that underlies all the higher level processes discussed above (section 1.2); they could not be contextually branching processes (which they are) unless there was the possibility of such branching processes at the underlying molecular and cellular levels. To be sure in practice they are not precisely digital processes, for example ion channels do not precisely behave as ON/OFF channels but rather have are a sigmoidal approximation to such behavior [5]. Nevertheless that description gives an excellent encapsulation of what occurs, as Berridge discusses, and is used for example by Davies and Walker (2016) and Walker et al. (2016) in Boolean network models of gene regulation in yeast.

However, there is also a major random element at the molecular level introducing statistical variation in happenings at

[3]It will in some cases be appropriate to call this top-down causation. That usage is controversial: it will be defended in a forthcoming paper (Ellis and Gabriel, 2019).

[4]Although this is effectively true in some specific contexts in developmental biology, where pre-determined developmental stages occur at specific times in an organisms developmental history Gilbert (2006); Wolpert (2002) via specific mechanisms whereby "*groups of cells are progressively apportioned distinct fates through a process of cell specification*" (Berridge, 2014: Module 8). For example, gastrulation occurs at a specific stage of development (Wolpert, 2002).

[5]There may also be intermediate states of channel opening. Then the logic is (10) rather than (9).

that level. It is then remarkable that these lower level processes produce reliable physiological outcomes at higher levels, such as regular heartbeats and breathing (Rhoades and Pflanzer, 1989), as well as evolutionary convergence to produce physiological function (Natarajan et al., 2016). The view here will be, in accordance with Noble and Noble (2018) that it is precisely this variation at the lower level that allows higher level processes to determine what occurs at the lower levels in order to adapt them to higher level needs (section 5.4). Thus despite this variation one can usefully analyse gene regulation via the above mentioned Boolean network models (Davies and Walker, 2016; Walker et al., 2016), which rely on the kind of branching logic discussed in this paper. Indeed the key point is that

**The lower level basis of higher level contextual functioning:** *None of the complex higher level biological features mentioned in section 1.2 would be possible if there was not a possibility of contextual branching function at the molecular level, which can often be well described by digital (Boolean) logic, despite the statistical nature of molecular processes.*

How that happens is the concern of this paper.

This paper focuses initially on the voltage gated ion channels that underlie neuronal functioning, although the same applies for example to the active sites of enzyme molecules which are complementary to the shape of the substrate. We first consider the difference between the logic of physics (section 2) and the logic of biology (section 3), then the biomolecules that make this difference possible (section 4), and finally how such molecules have come into being (section 5). The processes of deductive causation are discussed in section (6). The conclusion (section 7) clarifies first the three general kinds of causation that occur in biology, and second how contextual biological dynamics causes branching behavior at the underlying physical level. Overall, this is a view of how physics underlies integrative physiology (where everything occurs in a contextual way Noble, 2012, 2016; Ellis, 2016). We take it for granted that living systems are open non-equilibrium systems (Friston and Stephan, 2007). However, that by itself does not suffice to characterize life: a burning candle satisfies those criteria. More is required (section 1.2).

# 2. LOGIC OF PHYSICS

Basic physics evolution is Hamiltonian (section 2.1), and so does not display any branching behavior. However, two aspects of physical laws do exhibit branching: phase changes (section 2.2) and quantum wave function collapse (section 2.3); but neither of these relate to function as characterized above, enabled by branching dynamics. How then does physics enable such branching to emerge? Through symmetry breaking (section 2.4), which is how quite different behavior can emerge from the underlying physics. In a biological context where higher level branching dynamics occurs, that leads to branching physical behavior at the electron level, as discussed in section 7.2.

## 2.1. Classical Dynamics

Classical physics determines the evolution of a physical system by energy and momentum conservation equations (Arnold, 1989, p. 15–27), a force law (Arnold, 1989, p. 28–50), a Lagrangian (Arnold, 1989, p. 55–61), or a Hamiltonian (Arnold, 1989, p. 65–70,165–266). The context $C$ consists of boundary and constraint conditions. The dynamical law uniquely determines later states of the relevant variable $\mathbf{X}$ from suitable initial conditions $\mathbf{X}(t_1)$ (Arnold, 1989):

$$\begin{aligned}\texttt{IF at time } t_1, \ \mathbf{X} &= \mathbf{X}(t_1), \texttt{ THEN at time } t_2,\\ \mathbf{X} &= H(C, \mathbf{X}(t_1), t_2).\end{aligned} \tag{1}$$

Here the context $C$ is expressed via constraint equations

$$C(c, \mathbf{X}) = C_0, \ dC_0/dt = 0 \tag{2}$$

on the possible values of the variables, with control parameters $c$ affecting the form of those constraints. Examples are the dynamics of a classical pendulum (Arnold, 1989), and the gravitational dynamics of celestial objects (Binney and Tremaine, 2008). The dynamic equations have unique solutions, as shown by Arnold (Arnold, 1989, p. 8) (this is a result of $dC/dt = dC_0/dt = 0$). Thus there is a specific unique outcome: no branching takes place as in (9).

### 2.1.1. Invariance of Physics

The basic point is that we cannot alter the physical laws that govern what happens. We can however shape outcomes by determining what they act on, for example a pendulum or a digital computer; mathematically this is expressed through the constraints $C$. The physical laws relevant to daily life on Earth are Newton's laws of motion together with Galileo's equations for a falling body and Maxwell's equations for electromagnetism:

$$\nabla \cdot \mathbf{E} = 4\pi\rho, \qquad \nabla \times \mathbf{E} = -\tfrac{1}{c}\tfrac{\partial \mathbf{B}}{\partial t}, \tag{3}$$

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{B} = \frac{1}{c}\left(4\pi\mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}\right) \tag{4}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ the magnetic field, $\rho$ the charge, and $\mathbf{J}$ the current. Nothing can change those interactions. The motion of a particle with charge $e$, mass $m$, and velocity $\mathbf{v}$ is determined by

$$\mathbf{F} = m\frac{d\mathbf{v}}{dt} = e\{\mathbf{E} + \mathbf{v} \times \mathbf{B}\} + m\mathbf{g}. \tag{5}$$

where $\mathbf{g}$ is the gravitational field. Equation (1) represents the solutions that necessarily follow from (3–5), proceeding purposelessly on the basis of the context $C$. These equations are time symmetric and imply energy conservation. Bifurcations can occur in some cases when a small change in a contextual parameter or initial data occurs, but the outcomes are still determined uniquely by the dynamical equations (Arnold, 1989), even though the outcomes may be unpredictable in practical terms in the case of chaotic dynamics.

Statistical physics laws for aggregates of particles follow from the fundamental physics laws (Penrose, 1979; Blundell and

Blundell, 2008), which emergent laws by their nature determine probabilistic outcomes $P(q)$ for states $q$. They may also have stochastic elements due to random environmental effects, leading to stochastical dynamics represented by coupling deterministic equations of motion to "noise" that mimics the effect of many unknown variables. Then a stochastic term $\eta(t)$ must be added to (5) (see Longtin, 2010). The outcome will then not be determinate, but it will not relate in any way to function or purpose.

## 2.2. Phase Changes

One might suggest that bifurcations as proposed below (Equation 9) happen in physics when phase changes takes place, for example solid/liquid/gas transitions for a substance $S$ (Blundell and Blundell, 2008). These generically have a form like

$$\begin{aligned}&\texttt{GIVEN pressure } P \texttt{ and temperature T},\\ &\qquad \texttt{IF } \{P,T\} \in S_{P,V} \texttt{ THEN S is solid},\\ &\texttt{ELSE IF } \{P,T\} \in L_{P,V} \texttt{ THEN S is liquid},\\ &\qquad \texttt{ELSE S is gaseous.}\end{aligned} \tag{6}$$

Here the context is represented by the pressure $P$ and temperature $T$, and $S_{P,V}$, $L_{P,V}$ and $G_{P,V}$ are the subsets of the $(P, V)$ plane for solids, liquids, and gases respectively. At first glance this looks like it has the biological branching form (9). However, the regions $S_{P,V}$, $L_{P,V}$, and $G_{P,V}$ are fixed by the physics of the substance. Thus this is physical logic, determined purely by the laws of physics; no historical or evolutionary factor enters. Note for example the contrast with the homeostatic process governing core body temperature, where the setpoint of 98.4°F is not determined by physical laws; it was determined through evolutionary processes related to physiological optimization.

## 2.3. Quantum Physics

The Schrödinger evolution is Hamiltonian, but wave function collapse, as occurs when a measurement takes place, is a branching operation. However, such wave function collapse of a wave function $|\Psi(t_1)\rangle$ (an "event") is not deterministic. It has the logic

$$\begin{aligned}&\texttt{IF } |\Psi(t_1)\rangle = c_1|u_1\rangle + c_2|u_2\rangle + .... + c_n|u_n\rangle,\\ &\texttt{THEN } |\Psi(t_2)\rangle = \texttt{EITHER } a_1|u_1\rangle \texttt{ OR } a_2|u_2\rangle \texttt{ ....OR } a_N|u_N\rangle\\ &\texttt{with probabilities } |c_1|^2, |c_2|^2, ..., |c_N|^2 \texttt{ respectively.}\end{aligned} \tag{7}$$

where $a_i$ is the eigenvalue associated with the basis vector $|u_i\rangle$. Thus branching takes place, but the outcome that occurs is not fixed by the initial state, although the statistics of such outcomes is. It is a contextual process (Drossel and Ellis, 2018), but the logic (7) is not directly related to function. In the end all the processes we discuss in this paper are underlain by such contextual quantum-to-classical transitions.

## 2.4. Symmetry Breaking

The key physical effect enabling the existence of the biomolecules discussed here, with their functional properties arising out of complex molecular structures, is the existence of *broken symmetries* (Longo et al., 2012). These are what allow quite different kinds of behavior to emerge at higher levels out of the underlying physical laws, with all their symmetry properties, as explained by Anderson in his foundational paper "More is Different" (Anderson, 1972). Thus the underlying standard model of particle physics is Lorentz invariant, but the emergent biomolecules (such as shown in **Figures 3**, **4**) are not. Contextless physics is Hamiltonian, but physics in a biomolecular context is not (section 7.2). Hence in the end this is what enables the difference between physics and biology.

Again the underlying physics relevant to biological functioning is time symmetric, but biological effects such as cell signaling (Berridge, 2014) and adaptive selection (18) are not. The contextual process of wave function collapse in quantum physics (7) breaks the time symmetric of the Hamiltonian evolution of the wave function, and this underlies the way the cosmological arrow of time leads to the arrows of time in quantum physics and thermodynamics (Drossel and Ellis, 2018), and so underlies the crucial feature of the emergence of the arrow of time in biology. We will not comment further on this issue here.

# 3. LOGIC OF LIFE

Life of course obeys the laws of physics, so at each level whatever constraints are implied by physics are obeyed (Cockell, 2018). However, additionally living systems behave according to biological logic, leading to what Mayr characterizes as goal directed behavior (Mayr, 2004, p. 52) furthering function (section 1.3). Living systems collect and analyse information (Nurse, 2008), using it to predict probabilities and thereby use it to execute functional actions in the light of both genetic heritage and acquired information (Hartwell et al., 1999; Campbell and Reece, 2005). This involves a branching logic where outcomes are selected on the basis of context, as revealed by incoming information.

## 3.1. Dynamical Branching

The dynamics followed at each level of biological hierarchies is based on contextually informed dynamical branching $L$ that support the functions $\alpha$ of a trait $T$. Thus biological dynamics can be functionally-directed rather than driven by inevitability or chance:

$$\begin{aligned}&\texttt{Biological dynamics tends to further}\\ &\texttt{the function } \alpha \texttt{ of a trait } T\\ &\texttt{through contextually informed branching}\\ &\texttt{dynamics } L\end{aligned} \tag{8}$$

where function is defined as in section 1.3, and in its simplest form $L$ is branching logic of the form

$$\begin{aligned}&\texttt{L: given context } C, \quad \texttt{IF } T(\mathbf{X}) \texttt{ THEN } F1(\mathbf{Y}),\\ &\qquad\qquad\qquad\qquad \texttt{ELSE } F2(\mathbf{Z}).\end{aligned} \tag{9}$$

Here $\mathbf{X}$ is a contextual variable which can have many dimensions, and $\mathbf{Y}$ and $\mathbf{Z}$ are also variables that may have many dimensions;

they may be the same variables or not. $T(\mathbf{X})$ is the truth value of arbitrary evaluative statements depending on $\mathbf{X}$. It can arise from any combination of Boolean logical operations (NOT, AND, OR, NOR, etc.), perhaps combined with mathematical operations, while $F1(\mathbf{Y})$ and $F2(\mathbf{Z})$ are outcomes tending to further the function $\alpha$. Thus they might be the homeostatic response"If blood sugar levels are too high, release insulin," or the conscious "If the calculated range of the aircraft as presently fueled is $<$500 km, add more fuel" (a default unstated "ELSE" is always to leave the status quo).

Together with (8), the crucial point is

**Independence of physics**: *The evaluative function $T(\mathbf{X})$ and the outcome options $F1(\mathbf{Y})$ and $F2(\mathbf{Z})$ are not determined by the underlying physical laws, despite being enabled by them.*

Thus these branching processes are not determined by Newton's laws of motion, Maxwell's equations, Newton's or Einstein's theory of gravity, the fundamental theory of particle physics, or statistical physics. Rather they are shaped by evolutionary or developmental processes (Gilbert, 2006; Gilbert and Epel, 2009) to give highly complex outcomes (Rhoades and Pflanzer, 1989; Campbell and Reece, 2005) resulting from plant or animal phsyiology or animal behavior, or can be conceived by human thought so as to result in planned outcomes (Bronowski, 1973; Harford, 2017). In many cases at the molecular level this branching logic is to a very good approximation of a discrete (digital) nature: this is clear for example in Berridge's discussion (Berridge, 2014) of cell signaling systems. There will in practice be noise and time lags in real situations, leading to more complex contextual dynamics. However, a discrete description such as given by Berridge will adequately capture the causal essence of what is going on at a molecular level from a biological viewpoint (if that were not the case, his magisterial book would not make sense).

In more complex cases, there will be multidimensional spaces of options and responses:

$$\texttt{L: given context } C, \quad \texttt{IF } B_N(\mathbf{X}) \texttt{ THEN } F_N(\mathbf{Y}) \tag{10}$$

where $B_N$ is the $N$th truth function and $F_N$ is the $N$th response function. The key point is the same: there is an evaluation function $B_N$ independent of the underlying physics, and a branching dynamics $F_N$ that is followed depending on that function. In principle one can take a limit where evaluation outcome is continuous but in practice that is unrealistic: there will always be sensitivity limits to detection or response processes, so that in fact responses will be discrete responses to discrete ranges of input variables. In any case we will give a number of key cases below where the biological dynamics is well represented by (9) and it is the higher level dynamics emerging out of combinations of such operations that need description as in (10). In particular (9) is true for the cell signaling networks described by Berridge Berridge (2014), which are at the heart of much molecular biology.

One can suggest that trivially any dynamics of a physical system can be programmed in terms of branching logic equivalent to (10), so (10) is really not different from (1), but as discussed in detail in Binder and Ellis (2016), physical laws are not the same as programs: a physical law is not an algorithm (it is Newton's *Law of Gravity*, not Newton's *Algorithm for Gravity*). Furthermore, there is no Hamiltonian or Lagrangian that leads to (10), and in the physics case there is no function $\alpha$ associated with the dynamics, as in (8). Physics *per se* is not teleonomic and does not show branching behavior related to function (section 2). That is the import of the plethora of existence and uniqueness theorems for fundamental physics (for the gravitational case, see Hawking and Ellis, 1973) whereby initial data determines a unique outcome in a specific spacetime domain (therefore the dynamics does not have a branching nature). Unlike the case of physical laws, where the relevant interactions cannot be changed or chosen because they are given by Nature and are invariable, the branching interactions (10) can fulfill widely varying biological or social or mental functions or purposes and can be selected for those purposes. Once one has this basic logical branching enabled at the molecular level, it is possible for complex emergence to take place where branching dynamics is possible at higher levels, and information can be causally effective (Nurse, 2008; Walker et al., 2017)[6].

It is of course not intended here to imply that this kind of causation is deterministic: that is why the word "tends" is used in (8); probabilities may be the best description of the branching logic at play. In particular, chance plays a key role in evolutionary theory (Glymour, 2001; Mayr, 2002) and molecular interactions. Nevertheless such causation is often reliable (Rhoades and Pflanzer, 1989; Randall et al., 2002), for example in the case of the developmental programs which underlie developmental biology (Wolpert, 2002; Gilbert, 2006; Berridge, 2014: Module 8), in the case of molecular machines (Hoffmann, 2012), the systems underlying heart function described by Noble (Fink and Noble, 2008), and the metabolic networks and gene regulatory networks described by Wagner (Wagner, 2017). We take that issue up in section 5.4. In the next sections, we look at various forms the branching logic (9) can take, always taking (8) for granted. Key cases are homeostasis (11) and adaptive selection (18).

## 3.2. Homeostasis

A crucial form of branching logic in biology is implemented in feedback control circuits that are the foundations of *homeostasis* (Ashby, 1956; Rhoades and Pflanzer, 1989; Randall et al., 2002; Campbell and Reece, 2005, p. 8–10). These are basically of the form (Randall et al., 2002, p. 11, Modell et al., 2015)

$$\texttt{IF } X < X_{MIN}(C) \texttt{ THEN } X_{INC}(\mathbf{Y}), \quad \texttt{ELSE IF } X > X_{MAX}(C) \texttt{ THEN } X_{DEC}(\mathbf{Z}) \tag{11}$$

where $X_{INC}(\mathbf{Y})$ is some operation that increases the value of the target variable $X$ through changing the value of the control

[6]The concept of information is contentious in biology (Godfrey-Smith and Sterelny, 2016; Koonin, 2016). However, signaling is not (Berridge, 2014). We will take the pragmatic view that signals convey information.

variable **Y**, and $X_{DEC}(\mathbf{Z})$ is some operation that decreases the value of $X$ through changing the value of **Z** (which may or may not be the same as **Y**). The default is to leave the situation as is. Note that this is not a simple ON/OFF effect (Modell et al., 2015): it is a mechanism which will tend to correct the value of $X$ over time to lie between $X_{MIN}(C)$ and $X_{MAX}(C)$, with dynamics described by the equations of feedback control systems (Di Steffano et al., 1967; Sauro, 2017), using Laplace transforms to model the system and signals, in contrast to the physics Equations (3–5). The triggering values $X_{MIN}(C)$ and $X_{MAX}(C)$ are in general dependent on the context (e.g., if the organism is sleeping as against running).

This is a particular case of (9). Note that this is just one part of the complex interacting processes generating their own constraints, immersed in many dimensional interactions. However, (11) undoubtedly occurs at both macro and micro levels as part of this larger set of interactions. Thus such processes control blood pressure and core body temperature at the macro level, and potassium and sodium levels in axons and glucose concentration in extracellular fluid at the micro level [7]. Because biological homeostatic systems have been tuned through evolutionary processes, they are less subject to instabilities that afflict feedback control systems in general.

## 3.3. The Physical Hierarchy

The structural hierarchy of life (Ellis, 2016) is indicated in **Figure 1**. Networks of interactions between lower level modules lead to emergence of higher levels, which in turn act down on the lower levels to shape their interactions (Noble, 2008, 2016; Ellis, 2016). This leads to adaptive same level causation at each level of the hierarchy Noble (2012).

## 3.4. Building the Hierarchy: Black Boxing

Branching dynamics occurs at the molecular and cellular level (Berridge, 2014). When built into cell signaling networks, gene regulatory networks, metabolic networks, and neural networks, this bifurcating dynamics at the lower levels enable emergence of higher order operations such as occur in physiology and the brain, with branching logic (9) or (10) occurring at each level. However, the function of the lower levels is in turn contextually controlled by higher level elements (Noble, 2012), resulting in contextual emergence (Atmanspacher and beim Graben, 2009) where lower level logical choices are set so as to fulfill higher level purpose or function (Noble, 2008, 2012; Ellis, 2016). The combination of bottom-up and top-down effects enables the closure of constraints (Montévil and Mossio, 2015).

**Figure 2** from Goelzer et al. (2008) illustrates how branching operations at molecular level in a metabolic pathway can be regulated by higher order circuits through transcription factors that control the transcription of genes. They may be ON (that is, able to bind to DNA) or OFF (Berridge, 2014), in this way controlling transcription of DNA to messenger RNA.

[7] See for example "*Regulation of $Ca^{2+}$ homoeostasis by multiple hormonal and organ effector systems*" in Berridge (2014): Module 7, p. 76, and "*Hormonal regulation of blood $Na^{+}$ levels*" in Berridge (2014): Module 7 , p. 105.

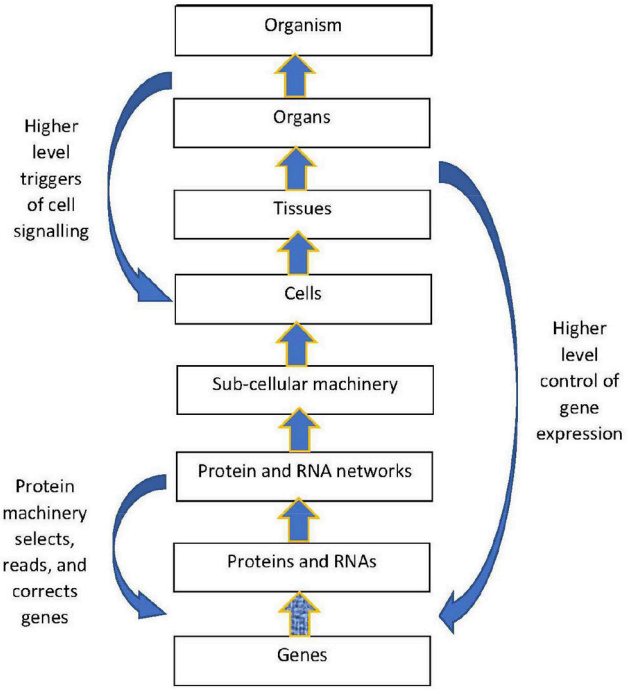


**FIGURE 1 |** Contextual control in the hierarchy. There is epigenetic control of lower level biological processes by higher level physiological states. These higher states determine what branching will take place at the cellular level by switching genes ON and OFF on the basis of higher level needs. Adapted from Noble (2012), with permission.

The transcription factor $TF_2$ is a local variable that is responsive to an intermediate metabolite $X_n$. It modulates synthesis of enzymes in the pathway, embodying branching logic of the form

$$IF\ TF_2\ on,\ THEN\ X_2 \rightarrow X_3,\ ELSE\ NOT \tag{12}$$

which is of the form (9). This is local branching within the module. However, the higher level regulator $TF_1$, sensitive to variables such as blood pressure or heart rate, modulates the synthesis of both intermediate enzymes and the local transcription factor $TF_2$. Thus the internal branching of the module results in a "black box" whereby conversion of metabolite $X_1$ to $X_n$ is determined by the higher level variable $TF_1$:

$$IF\ TF_1\ on,\ THEN\ X_1 \rightarrow X_n,\ ELSE\ NOT \tag{13}$$

The outcome is again of the branching form (9), but occurring at a higher level (because $TF_1$ is a higher level variable). The function is production of $X_n$ when and only when it is needed. Thus lower level branching circuits such as (12) can be used to build up higher level branching logic such as (13). This is how *abstraction* occurs in a modular hierarchy (Booch, 1994), so that internal workings of a module are hidden [in this case $TF_2$, $E_2$, $X_2$, and $X_3$ are internal variables that do not occur in the higher level relation (13)]. From the system view, what

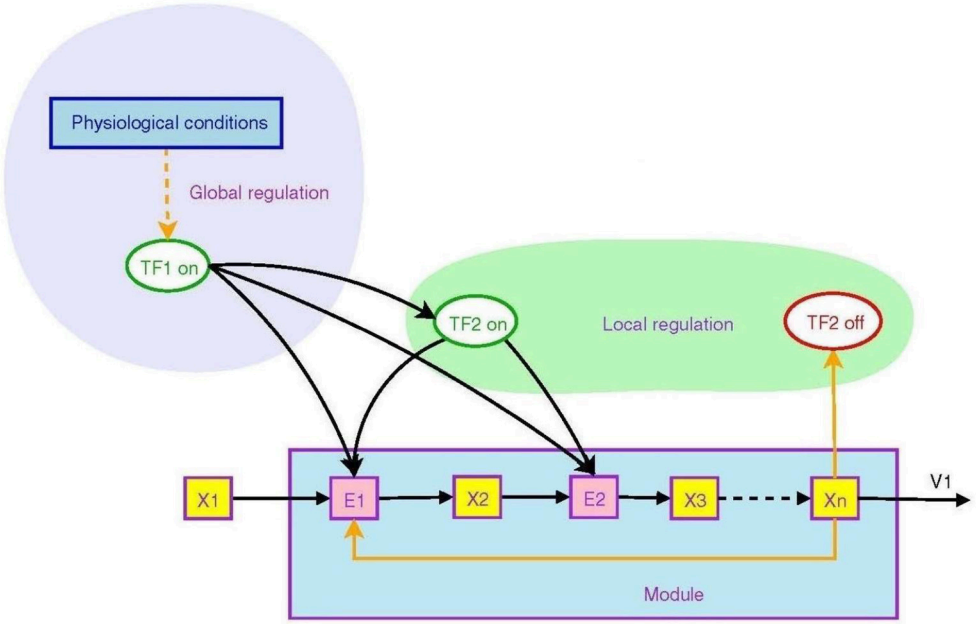


**FIGURE 2 |** Non-local and local regulation in the context of a metabolic pathway. Depicted is the overall logical structure of a metabolic pathway module that converts food into energy and proteins, with binding of various factors serving as ON/OFF switches. This enables the top-down effect of the physiological environment. Transcription factors are TF's, enzyme pools are E's, metabolic pools are X's. From Goelzer et al. (2008) (open access) .

matters is the emerging logic (13) where transcription factor $TF1$ controls conversion of metabolite $X_1$ to $X_n$. Regulation of lower levels through higher level conditions is possible between any adjacent levels in the hierarchy. Through it, metabolic regulation can control gene expression in a top-down way (Alam, 2016), as in Figure 1. The underlying assumption is that there is a suitable cellular context for this to happen (Hofmeyr, 2017).

### 3.4.1. Black Boxing

As just demonstrated, in the case of a complex logical system, you do not get the higher level behavior by coarse graining, as in the case of determining density and pressure from statistical physics (Penrose, 1979). Instead, you get it by *black boxing* and *logical combination*, involving information hiding and abstraction to characterize the exterior behavior of a module (Ashby, 1960; Oizumi et al., 2014). This is particularly clear in the case of digital computer systems, with their explicit apparatus of abstraction, information hiding, and carefully specified module interfaces, see Grady Booch's book *Object Oriented Analysis* (Booch, 1994). Even though biological systems are not running logical programs, they use the same basic principles of modularity and abstraction in cell signaling systems.

## 3.5. Multiple Realization

A key feature in the emergence of higher level structure and functions is the multiple realization of higher level structures and functions at lower levels. This is central to the way modularity and black boxing works: the function of a module can be realized by many different internal variables and causal networks. Thus in **Figure 2**, it does not matter what the internal dynamics of the module is provided it leads to the emergent result (13). This degeneracy occurs in all biology in relation to the underlying microbiology and physics: many different lower level realizations of the needed higher level functions can occur. Such multiple realization occurs *inter alia* in the metabolic networks in a cell, gene regulatory networks,and neural networks.

The key underlying analytic concept is existence of *functional equivalence classes* of lower level structures and functions (Auletta et al., 2008; Ellis, 2016) corresponding to a specific emergent structure or function. Equivalence classes at a lower level collect elements whose differences are irrelevant for the emergent target feature at the higher level; it does not matter which one is used to realize the higher level feature. Existence of such functional equivalence classes is an indication of top-down causation (Auletta et al., 2008). An important example is the relation of developmental systems to the genome: a huge number of different genotypes (a *genotype network*) can result in the same phenotype (Wagner, 2017). Any one of these genotypes can be selected for through evolutionary processes in order to lead to a particular emergent function that promotes survival. As far as the higher level function is concerned, it is irrelevant which specific genotype is selected, so it is membership of the equivalence class at the lower level that is the key to what genotype gets selected when adaptation takes place. The huge size of these equivalence classes is what enables adaptive selection to find the needed biomolecules and interaction networks on geological timescales (Wagner, 2017).

## 4. LINKING PHYSICS AND BIOLOGY: THE PHYSICAL BASIS

All these branching operations emerge from the underlying physics, but are of a quite different nature than the deterministic function of physical laws *per se* (section 2). So how is it possible that they can be realized through the functioning of the underlying physical levels? We will now focus on the brain to give the discussion a specific biological context.

### 4.1. The Nervous System

The operations of brains is based in the functioning of neurons that are linked together by synapses, thereby being structured as neural networks (Kandel et al., 2013) enabling neoronal signaling (Berridge, 2014): (Module 10). Spike trains proceed via dendrites to the neuron soma where a summation operation is performed. Spike trains then proceed from the cell body down axons to synapses, where another summation process occurs; signals are passed on to other neurons if the sum is above an activation threshold (Kandel et al., 2013). The function is to underlie the processes of the nervous system that enable an animal to anticipate and counter threats to its existence, thus enhancing its chances of survival.

The flow of currents in the dendrites and axons is determined by the underlying physics, described by equations (3–5) plus statistical relations and diffusion equations. In a neuronal context, these lead to the Hodgkin-Huxley equations (Hodgkin and Huxley, 1952) which characterize how ion flows underlie the existence of action potential spike trains (Randall et al., 2002, p. 132–1139). These equations result from the physical structure of ion channels (Catterall, 2000; Randall et al., 2002, p. 141–150) which control flow of ions in and out of the cell membranes. The constants occurring in these equations are not universal physical constants, but rather are constants that characterize the membrane structure. It is not possible to deduce them from the laws of physics *per se* (Scott, 1995).

### 4.2. Linking Physics to Logic: The Molecular Basis

The branching logical function (10) that emerges is enabled by particular proteins: namely voltage gated ion channels imbedded in axon and dendrite membranes (Catterall, 2000; Randall et al., 2002; Magleby, 2017, p. 146–151) (see **Figures 3**, **4**). They control the flow of potassium, sodium, and chloride ions, leading to action potential spike chain propagation along the axons and dendrites. Their molecular structure and function is discussed in (Randall et al., 2002, p. 139–147).

The ion channels result in branching dynamics with the following logical structure:[8]

$$\texttt{IF voltage difference } V > V_0 \texttt{ THEN allow ion flow, ELSE not} \quad (14)$$

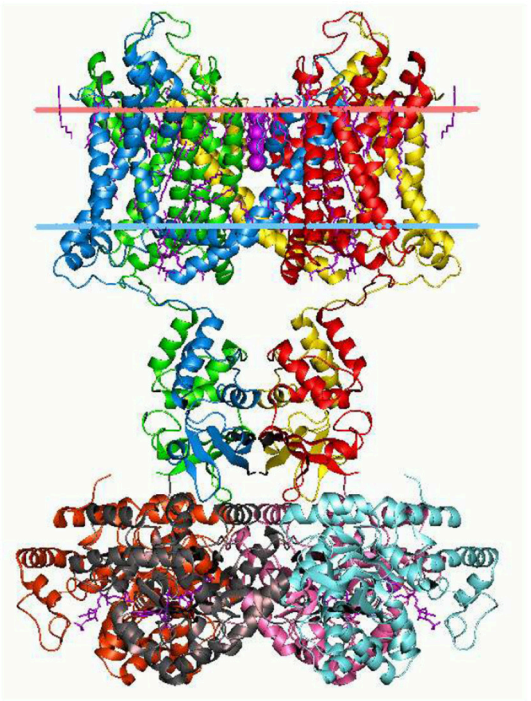

**FIGURE 3 |** Potassium ion channel structure in a membrane-like environment. This 3-dimensional structure alters according to the voltage difference across the membrane, hence allowing or impeding ion passage. Diagram by Andrei Lomize. From the Open Membranes (OPM) database, with permission.

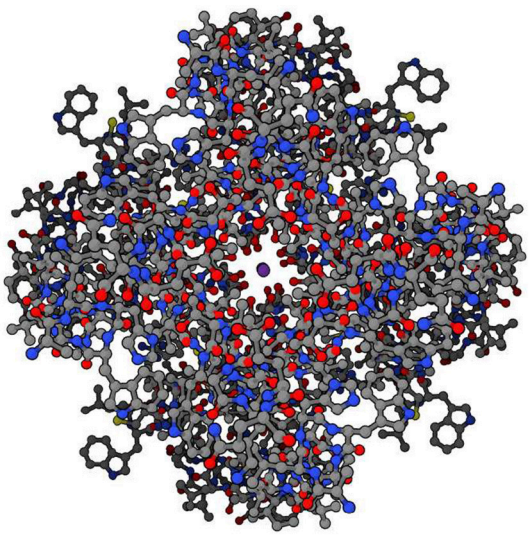

**FIGURE 4 |** Potassium ion channel functioning. Top view of potassium ion (purple, at center) moving through potassium channel when channel is open. (Protein Data Bank:1BL8, open access).

which is a specific case of the form (9). The function is to facilitate the propagation of action potentials in axons, and so enable functioning of the nervous system (Randall et al., 2002). It is

[8]In practice, the response function is not discontinuous as in this representation, but is a a logistic curve linking 'ON' and 'OFF' states. The principle remains the same: but one now uses a more complex response function. Equation (14) is a good first approximation (cf. Berridge, 2014.)

the detailed 3-dimensional structural form of the ion channels, specifically its tertiary and quaternary structure (see **Figures 3**, **4**), that enables conformational changes in response to local conditions that controls the flow of ions in and out of the cell. This is what enables branching dynamics to emerge from the underlying physics (Farnsworth et al., 2017, p. 313; Kandel et al., 2013; beim Graben, 2016). Similar issues arise via synapses (Kandel et al., 2013; Berridge, 2014: Module 10, p. 28–41), where a branching logic.

$$\texttt{IF summed input voltage } V > V_0 \texttt{ THEN fire action potential, ELSE not} \quad (15)$$

holds, enabled by voltage-gated $Ca^{++}$ channels in conjunction with pre- and post-synaptic neurotransmitter transporters and post-synaptic receptors.

Once physical implementation of logical processes have been achieved at the lower levels, this provides the building blocks for implementing logical processes at higher levels, enabling emergence of branching function in cortical networks. ON/OFF logical units can be used to give the basic operations AND, OR, NOT, and can then be combined in neural networks with thousands of synaptic connections per neuron, and with both upward and downward connections. This enables the coordinated neural dynamics involved in higher level cognitive functioning. Thus the relevant low level physical structure enabling lower level branching function that then enables emergence of higher level branching function is that of proteins (Petsko and Ringe, 2009) imbedded in the cell membrane.

In summary, Given the right cellular context (Hofmeyr, 2017), biomolecules such as ion channels (Catterall, 2000; Magleby, 2017) can act as logic gates, underlying the emergence of complex life processes where branching logic occurs at the higher levels of physiological systems (Rhoades and Pflanzer, 1989; Campbell and Reece, 2005; Goelzer et al., 2008; Kandel et al., 2013).

## 4.3. More General Biological Contexts

The basic branching logic discussed here occurs also in the metabolic processes, cell signaling networks, and gene expression (controlled by gene regulatory networks) which underlie the functioning of all cells (Berridge, 2014; Hofmeyr, 2017, 2018; Wagner, 2017).

### 4.3.1. Metabolism

The purpose of metabolism (Krebs, 1993; Berridge, 2014: Module 7; Hofmeyr, 2017) is to produce molecules and free energy needed by the cell in usable form, which are crucial for its function and survival. Enzymes and ribosomes catalyse metaboliism, providing the building blocks of life. This is only possible because of the presence of extremely efficient catalysts, particularly enzymes, that are highly specific with respect to the substrates they recognize and so the reactions they catalyze. The branching logic is (cf. section 3.4),

$$\texttt{IF catalyst for reaction R1 present THEN R1 proceeds, ELSE not.} \quad (16)$$

Its molecular basis is the relevant lock and key recognition mechanism (Lehn, 1995, 2007; Alberts et al., 2007).

### 4.3.2. Cell Signaling Networks

These are discussed in depth in Berridge (2014). They are again based in the lock and key recognition mechanism, which at a functional level can be well-described in terms of digital logic as an ON/OFF mechanism (Berridge, 2014). At the molecular level, it is based in complementary molecular shapes (Alberts et al., 2007; Watson, 2013).

### 4.3.3. Gene Expression and Gene Regulatory Networks

The purpose of the genetic code is to specify the sequence of amino acids that will lead to existence of proteins with crucial cellular functions (Alberts et al., 2007; Watson, 2013). Given the cellular context $\mathcal{C}$ (without which no reading of the genetic code would take place Hofmeyr, 2017), the branching logic is

$$\texttt{IF triplet GGU THEN Gly ELSEIF triplet GGC THEN Gly ELSEIF ...,} \quad (17)$$

with a unique mapping specified for each of the 64 codon triplets. Again it is based in complementary molecular shapes that lead to molecular recognition (Watson, 2013). This particular highly degenerate mapping (Watson, 2013; Wagner, 2017) implemented by cellular processes (Alberts et al., 2007) has been determined by the specific historical events of the evolutionary history of life on Earth (Campbell and Reece, 2005; Godfrey-Smith, 2017): many other mappings are chemically possible. Physics by itself does not determine the specific mapping that in fact has occurred (Watson, 2013), represented by the logic (17).

Which sections of DNA are read where and when is under epigenetic control (Carroll, 2005; Gilbert and Epel, 2009), enabled by cell signaling networks (Berridge, 2014) and gene regulatory networks (Wagner, 2017). A key feature of DNA expression is alternative splicing, whereby a single gene codes for multiple proteins, and overlapping genes, where an expressable nucleotide sequence for one gene is also an expressable nucleotide sequence for another. Given epigenetic control that determines these aspects, readout from nucleotide sequences to amino acids takes place as in (17). Furthermore, the epigenetic systems are themselves made up of interacting molecules that arise through the kind of branching logic we discuss here through gene regulatory networks that can be described in a Boolean way to a good approximation (e.g., Wagner, 2011; Davies and Walker, 2016).

# 5. EXISTENCE OF THE RELEVANT PROTEINS

Two issues arise here: the possibility of existence of the biomolecules needed, for example those that comprise ion channels, and how they come into being.

## 5.1. The Possibility of Their Existence

Given the nature of physics as we know it (with particular values for the fundamental constants of nature such as the fine structure constant Uzan, 2003), the nature of possible physical structures at the molecular level is controlled by electromagnetism together with quantum physics. Thus the possibility of the existence of biomolecules, and specifically the proteins controlling biological activity (Petsko and Ringe, 2009), is a result of covalent bonds, hydrogen bonds, and van der Waals forces (Watson, 2013).

The result is a space of possible proteins (Petsko and Ringe, 2009) of vast dimensions: an unchanging space of all possible molecular structures (Wagner, 2017). However, their possible existence is not by itself enough: there must be viable mechanisms that can bring them into being.

## 5.2. Their Coming Into Being: Development and Evolution

Given this vast possibility space, how have the specific proteins that actually exist come into existence? This question has developmental and evolutionary aspects.

### 5.2.1. Developmental and Epigenetic Aspects

The relevant proteins come into being through molecular processes transcribing genetic information coded in DNA (Alberts et al., 2007; Watson, 2013) into amino acid chains, which then fold to create biologically active proteins. This reading of the genotype occurs in a contextual way (Wolpert, 2002; Gilbert, 2006; Noble, 2012) because epigenetic processes (Pigliucci and Müller, 2000; Gilbert and Epel, 2009), controlled by gene regulatory networks, determine which gene segment gets read at a specific time and place, thereby shaping developmental processes according to the local environment (Oyama et al., 2001; Gilbert and Epel, 2009). Epigenetic effects even allow genetic rewriting (Lee et al., 2018) so that "genes are more followers than promoters of evolution" (West-Eberhard, 2003). As stated by Noble and Noble (2017),

> "*Organisms and their interacting populations have evolved mechanisms by which they can harness blind stochasticity and so generate rapid functional responses to environmental challenges. They can achieve this by re-organizing their genomes and/or their regulatory networks. Epigenetic as well as DNA changes are involved. Evolution may have no foresight, but it is at least partially directed by organisms themselves and by the populations of which they form part.*"

Nevertheless the reading of the DNA still takes place as above (section 4.3), once epigenetic processes have selected which specific DNA segments will be read in what order.

### 5.2.2. Evolutionary Aspects

The question then is, how did that genetic information get written? As stated before, we do not enter here into the discussion of how life started: we assume here that somehow cells came into existence, allowing metabolism and the existence and reading of genetic information. In that context, how was it that the genotype for the specific proteins that actually occur (Petsko and Ringe, 2009) come to be written, given that there is a vast space of possible proteins that might have existed (Wagner, 2017)? What about the origin of the gene regulatory networks controlling body plan development (Peter and Davidson, 2011)?

The relevant proteins are extraordinary complex biomolecules (Petsko and Ringe, 2009) with specific functions that are essential for survival, where function is as characterized in section 1.3. For example, hemoglobin transports oxygen in our blood stream; chlorophyll enables plants to harvest solar energy, and so on. Thus they will have been strongly subject to selection pressure because of these vital functions, and so arguably cannot have come into being through genetic mutation, drift, or recombination alone (Morris and Lundberg, 2011, p. 21) without selection playing a decisive role (Farnsworth et al., 2017, p. 313). The natural hypothesis is that they were selected through the process of Darwinian adaptive selection (Darwin, 1872; Mayr, 2002; Campbell and Reece, 2005; Morris and Lundberg, 2011) occurring at the organism level, with these selective outcomes chaining down to the genotype level within a functional cellular context (Hofmeyr, 2017). The genotype-phenotype map has massive degeneracy that would have played a crucial role in enabling new phenotypes and hence associated genotypes to have come into being in the available time (Wagner, 2011), and doing so in such a way that the organism remains viable at each step. The process is contextually driven and hence is an example of top-down causation (Campbell, 1974; Ellis, 2016).

However caution is warranted. Genetic drift, leading to neutral selection (Kimura, 1983; Lynch and Hill, 1986; Nei, 2005) can explain some aspects of human physiology (Ackermann and Cheverud, 2004; Schroeder and Ackermann, 2017). How do we prove it was selection rather than drift that lead to existence of specific proteins? In the case of phenotypes, one can sometimes determine which features are due to selection pressure and which due to drift, thus a detailed study shows "*during the early evolution of the genus Homo, [...] genetic drift was probably the primary force responsible for facial diversification*" (Ackermann and Cheverud, 2004).[9].

How to determine this for proteins or gene regulatory networks is a fascinating challenge. They obviously play a key physiological role but, particularly given the existence of vast equivalence classes of genotypes that can produce acceptable phenotypes (Wagner, 2011), it is far from clear how to determine what aspects of the proteins are selectively determined and what are due to drift. We simply comment, in agreement with (Wagner, 2017) that there has to have been a major selective aspect underlying their evolutionary development, as otherwise they would not exist able to play the functional roles they do.

## 5.3. The Generic Selection Process

Darwinian adaptive selection is a special case of the generic selection process that is ubiquitous in biology. The basic nature of this process is that there is a random input ensemble of entities which is is filtered so as to produce an output ensemble that fulfills some environmentally dependent selection criteria (**Figure 5**), and so is more ordered than the input ensemble. The

[9]We note here that positive sexual selection may also have played a role

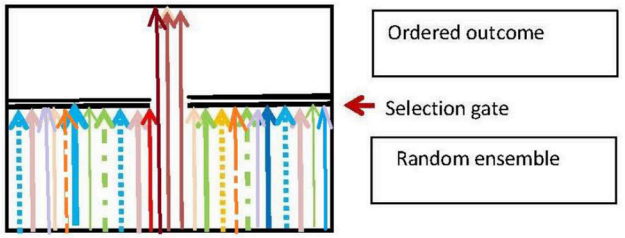


**FIGURE 5 |** The generic selection process. Adapted from Ellis and Kopel (2018).

branching logic of the process is:

$$\Pi_S(X):\{\texttt{IF } X \notin S(C,\mathcal{E}) \texttt{ THEN delete } X\} \tag{18}$$

Here $S$ is the subset of elements that is selected to survive on the basis of the selection criterion $C$, and the environmental context is $\mathcal{E}$. The resulting effect on the input ensemble $\{E(X)\}$ is a projection operation $\Pi_S$ that gives the output ensemble $\{\hat{E}(X)$:

$$\Pi_S:\{E(X)\} \rightarrow \{\hat{E}(X) : X \in S(C,\mathcal{E})\}. \tag{19}$$

The function of the process is to produce a population of entities that fulfill the selection criterion $C$. The basic physics case is Maxwell's Demon (Von Baeyer, 1998), where the criterion $C$ for allowing a molecule to pass the trapdoor is $|\mathbf{v}| > v_0$ where $|\mathbf{v}|$ is molecular speed. A biological case is the immune system, deleting invading pathogens (Rhoades and Pflanzer, 1989; Randall et al., 2002). A logical case is the deletion of emails or files on a computer, in accord with some relevance criterion $C$.

Darwinian selection (Godfrey-Smith, 2001; Mayr, 2002; Campbell and Reece, 2005) has the overall structure (18) where $C$ is a measure of inclusive fitness (West and Gardner, 2013) in the context of the environment, and the input ensemble at each time $t_2$ is a randomized variant of the output of the previous process at time $t_1$:

$$\{E(X)\}(t_2) = R\{\hat{E}(X)(t_1)\}. \tag{20}$$

At the genotype level, $R$ is randomization based in recombination, mutations, and horizontal gene transfer. This results in variation at the phenotype level, which is where the selection (survival of an animal or plant until reproduction can take place) actually occurs[10]; that selection then chains down to the genotype level. Thus the process is a continually repeated multilevel 2-step process (Mayr, 2002: p. 130–133): *reproduction with variation* (20), which is where directed sexual selection and differential reproductive success enters, followed by *elimination* (18), which is where differential survival rates matter (this only requires selection of individuals who are "good enough" (Mayr, 2002, p. 130–131); they don't have to be the fittest, which is partly why drift is possible). It is the elimination phase (18) that leads on average, in suitable circumstances, to selection of individuals with traits that are better fit to the environment. The combination of these two processes leads to inclusive fitness (West and Gardner, 2013). Thus this adaptive selection process (Morris and Lundberg, 2011) functions to produce individuals fit to survive in a specific environmental context through their physiology and functioning even though the process has no intentional "purpose" (Mayr, 2004, p. 58). It thereby leads *inter alia* to existence of the molecules we discuss in this paper (Wagner, 2017).

[10]Or perhaps at the group level; this is the contentious issue of levels of selection (Okasha, 2010)

## 5.4. What Role Does Chance Play?

Biological processes display a great deal of randomness, particularly at the molecular level where there occurs a "molecular storm" (Hoffmann, 2012). The occurrence of this noise does not mean the outcome is random: reliable physiological function emerges at higher levels (Rhoades and Pflanzer, 1989; Randall et al., 2002). In fact microbiology thrives on randomness (Hoffmann, 2012; Noble, 2017), and this is also the case for brain function (Glimcher, 2005; Rolls and Deco, 2010). Furthermore, randomness plays a key role in evolution (see Glymour, 2001; Mayr, 2002, p. 252–254, Kampourakis, 2014, p. 184–191), underlying that vast variety of life on Earth by providing a very varied set of genotypes on which selection can operate, for example leading to predictable convergence in hemoglobin function (Natarajan et al., 2016).

We propose, in agreement with Noble and Noble (2018), that randomness plays a key role at the molecular level by providing an ensemble of variants from which higher level selection processes can choose what happens through selection of outcomes according to higher level selection criteria (18), thus creating order out of disorder in a reliable way (Noble and Noble, 2018), as represented by (19) . As stated by Noble and Noble Noble and Noble (2018),

> *"Choice in the behavior of organisms involves novelty, which may be unpredictable. Yet in retrospect, we can usually provide a rationale for the choice. A deterministic view of life cannot explain this. The solution to this paradox is that organisms can harness stochasticity through which they can generate many possible solutions to environmental challenges. They must then employ a comparator to find the solution that fits the challenge. What therefore is unpredictable in prospect can become comprehensible in retrospect. Harnessing stochastic and/or chaotic processes is essential to the ability of organisms to have agency and to make choices"*

For example, molecular binding processes depend on random presence of the appropriate substrate for a binding site, and the adaptive immune system depends on random generation of antibodies to find the one that works against a particular pathogen. This is also the essential feature of Edelman's Neural Group Selection (Edelman, 1987), which envisages initial random neuronal connections (Wolpert, 2002) being pruned and strengthened according to selection criteria provided by an innate 'value system' in the brain (which in psychological terms can be associated with innate primary emotional systems; Toronchuk and Ellis, 2013; Ellis and Solms, 2017). Furthermore,

this underlies the possibiity of real mental emergence, as proposed by Mitchell Mitchell (2018):

> "*I argue here that physical indeterminacy provides room for the information entailed in patterns of neuronal firing—the mental content of beliefs, goals, and intentions—to have real causal power in decision-making.*"

## 6. DEDUCTIVE CAUSATION

Deductive causation takes place when effects are the outcome of explicit logical processes, as contrasted to the biological cases discussed so far, where they are processes that are indeed carrying out what can be characterized as logical operations, but these are implicit in the biology rather than explicit.

Deductive causation requires mental processes that explicitly consider alternative logical inevitabilities or probabilities and decide outcomes on this basis, for example, "If I wait till 10am I will miss the bus, so I'd better leave now". This requires conscious intelligence[11], and certainly occurs in the case of humans. It may also occur to some degree in animals, but we will not enter that debate here: the essential point is that it does indeed occur in the real world, as evidenced by the existence of books, aircraft, digital computers, and all the other products of conscious design (Harford, 2017). It is made possible by the existence of brains (at the macro scale) (Kandel et al., 2013) and their underlying biomolecules such as voltage gated ion channels (at the micro scale) (Scott, 1995; Kandel et al., 2013), as discussed in section 4, enabling information to be causally effective (Walker et al., 2017).

We look in section 6.1 at deductive argumentation $\mathcal{D}$, whose truth is valid independent of contingent facts, in section 6.2 at evidence based deduction $\mathcal{DE}$, where the addition of empirical data $E$ leads to conclusions that follow from that evidence via logical deduction $\mathcal{D}$, and in section 6.3 at deductively based predictions of outcomes $\mathcal{DEO}$, which are used to decide on best choices of actions $\mathcal{DEOC}$ on the basis of logical predictions of outcomes $O$ following from the data $E$ together with choice criteria $C$.

### 6.1. Deductive Argumentation

Deductive argumentation can be definite or probabilistic. *Definite deductive arguments* deal with inevitable outcomes of abstract relationships between variables:[12] thus[13]

$$\mathcal{D}: \ \texttt{IF}\ T1(\mathbf{X})\ \texttt{THEN necessarily}\ T2(\mathbf{Z}), \tag{21}$$

where $T1(\mathbf{X})$ may involve logical operations AND, OR, NOT, and their combinations, or mathematical equalities or inequalities, or both logical and mathematical relations in any combination. Thus one might have a conjunction of conditions

$$\mathcal{D}2: \ \texttt{IF}\ T1(\mathbf{X})\ \texttt{AND}\ T2(\mathbf{Y}),\ \texttt{THEN necessarily}\ T3(\mathbf{Z}), \tag{22}$$

where **X**, **Y** and **Z** may or may not be the same variables. These are of the same logical form as (9), but the key difference is that in that case, the context was the logic implicitly embodied in biological processes, whereas here the relations refer to explicit logical thought patterns. They may be realized at some moment in a brain, or written down on paper, or recorded in some other way (such as on a black board or a computer screen), but the patterns themselves are abstract relations with their own internal logic that is independent of whatever specific realization may occur.

Mathematical examples are the relations

$$\texttt{IF}\ \{\texttt{X=2}\}\ \texttt{THEN}\ \{\sqrt{X}\ \texttt{is irrational}\} \tag{23}$$

which is proved by algebraic argumentation, and the partial differential equation result

$$\begin{gathered}\texttt{IF}\ \{\texttt{Eqns.(3),(4) hold with}\ \mathbf{J}=\rho=0\},\\ \texttt{THEN}\ \{\texttt{wave solutions}\ u(x,t)=F(x-ct)+G(x+ct)\ \texttt{exist}\}\end{gathered} \tag{24}$$

(which mathematical fact underlies the existence of radios, TV, cellphones, etc).

Logical examples are the relations

$$IF\ \{A \Rightarrow B\}\ AND\ \{B \Rightarrow C\}\ THEN\ \{A \Rightarrow C\} \tag{25}$$

and the combinatorial rules of Boolean logic involving AND, OR, NOT, and so on.

*Probabilistic logical arguments* deal with likely outcomes on the basis of statistical evidence, for example:

$$\texttt{IF}\ \ T1(\mathbf{X,P1})\ \texttt{AND}\ T2(\mathbf{Y,P2}),\ \texttt{THEN probably}\ T3(\mathbf{Z,P3}), \tag{26}$$

where $T1(\mathbf{X,P1})$ means $T1$ is valid with probability $P1$, and so on. A key example is Bayes' Theorem (Stone, 2015):

$$\texttt{IF}\ \ \{P(A)\ \texttt{AND}\ P(B|A)\ \texttt{AND}\ P(B)\}\ \ \texttt{THEN}\ \ P(A|B)=\frac{P(B|A)\,P(A)}{P(B)}, \tag{27}$$

where $P(A)$ and $P(B)$ are the probabilities of observing events $A$ and $B$ independent of each other, $P(A|B)$ is the conditional probability of observing event $A$ given that $B$ is true, and $P(B|A)$ is the conditional probability of observing event $B$ given that $A$ is true. This relation, which is of the form (26), underlies the learning processes of the predictive brain (Huang and Rao, 2011; Clark, 2013; Hohwy, 2013), enabled by suitable neural structures (Hawkins, 2004; Bogacz, 2017, section 2.3–2.5) built from biomolecules (Scott, 1995). This topic is developed further in section 6.5.

[11] We note here that these processes can become automated after much practice so that they are intuitive rather than the result of directed mental effort. Nevertheless the nature of the causation is the same.

[12] We are not giving a formal definition of logic here, but rather a sketch of how it works. It can be any form of logic that has been discovered by the human mind.

[13] Strictly speaking, the word "necessarily" is superfluous. We add it for emphasis here and below. A similar remark applies to "probably" in (26).

## 6.2. The Link to Data: Evidence Based Deduction

It may well be that we know that the antecedents in some of these arguments are either true, or are highly probable, in which case we can move to evidence based deduction: (21) becomes

$$\mathcal{DE}: \texttt{SINCE T1(X) THEN necessarily T2(Z)}, \tag{28}$$

where $T2(Z)$ necessarily follows from $T1(X)$, and we know $T1(X)$ to be true either because we have seen it to be true (there is a dog in the room), or it is common knowledge (England is near France), or it is an established scientific fact (DNA is a key molecule underlying genetic inheritance), or at least it is a best explanation (established by abduction, i.e., inference to best explanation from observations). For example

$$\texttt{SINCE } E = mc^2 \texttt{ THEN binding energy can be made available via nuclear fission of heavy atoms}, \tag{29}$$

In other words, because we know special relativity is true, we know we can in principle make nuclear power stations and nuclear bombs. Thus reliable data (the experimental verification of the logically deduced relation $E = mc^2$) relates deductive argumentation to real world possibilities. Similarly an extension of a simple case of (26) becomes

$$\texttt{SINCE } T1(\mathbf{X,P1}) \texttt{ THEN probably } T2(\mathbf{Z,P3}), \tag{30}$$

in the probabilistic case, for example

SINCE there are dark clouds in the sky
THEN it will probably rain today.

The deduction leads to the conclusion that a specific outcome is likely to actually occur.

## 6.3. Deductively Based Action

Following on (28) and (30), we can deductively determine that specific actions will inevitably or probably have specific outcomes:

$$\mathcal{DEO}: \texttt{ SINCE } T1 \texttt{ is true THEN action A will lead to outcome O .} \tag{31}$$

This leads to the basis of deductive choice of best actions:

$$\mathcal{DEOC}: \texttt{ WHEN } T1 \texttt{ is true THEN DO A(}\mathbf{V}\texttt{) TO } \mathcal{C}\texttt{-optimize O} \tag{32}$$

where $\mathcal{C}$ is a selection criterion for the best outcome $O_*$, and $A(\mathbf{V})$ is some action chosen to alter $O$ via a control variable $V$. The purpose is to produce an optimal outcome $O$ on the basis of a representation of the situation founded on the best available evidence (Papineau, 2016). An example is

$$WHEN\,\{T > T_0\} \texttt{ THEN \{set V ON\} SO THAT } \mathcal{C}:\{T_1 < T_0\} \tag{33}$$

which might be part of a computer program implementing feedback control (14) to ensure that temperature $T$ is kept below a critical level $T_0$ via the cooling control variable $V$. In the probabilistic case it might be

SINCE {there is 60% chance of rain}
THEN {take an umbrella} TO {keep dry}.

When we carry out such deductive argumentation, the abstract logic of the argument $\mathcal{D}$ [see (21)] is the causal element determining the nature of the resulting outcomes. The aircraft flies well because we have used explicit deductive mathematical logic $\mathcal{D}$, together with our knowledge of the laws of fluid dynamics $T1$, to optimize its design O by running computer aided design packages $A(V)$ representing the aircraft design via variables $V$. We call $\mathcal{D}$ a "causal element" because of the counter-factual argument (Menzies, 2014) that if this abstract logic were different, the outcome would be different. The same applies to $\mathcal{C}$: if the decision criteria are changed the outcome changes, for example the wing design will be different if the plane is a fighter or an Airbus. This kind of argument is a key part of planning (Epstude and Roese, 2008).

In practice (e.g., in economic planning) the argument is often probabilistic because we can never be absolutely certain of the outcome, due to uncertainty concerning the contextual effects $C$. Overall, the import of this section is that

> **Deductive causation**: *Logical deductions about scientific, engineering, and social issues can lead to action plans that are causally effective in terms of altering the world. In these cases it is explicit abstract logic $\mathcal{D}$ realized in brains and/or computers that guides and shapes what happens in highly productive ways (Harford, 2017) and hence may be said to be the essential cause of what happens.*

This is all possible because of the properties of brains as prediction machines that are also able to make choices between alternatives. The logical operations of deduction $\mathcal{D}$ and prediction $\mathcal{DEO}$ take place at the psychological level in the brain (Ellis, 2016), while being realized at the neural network level through spike chains, at the axon level through ion flows, and at the electronic level through electron movements (Scott, 1995). Each level does work appropriate to the logic at that level, but it is the high level deductive logic $\mathcal{D}$ that determines what happens in terms of specific outcomes through logically based choices $\mathcal{DEOC}$ (Ellis, 2016).

## 6.4. The Creative Element

Deductive causation depends on being able to choose between options, which is where imagination comes in. There must be a process in the brain that generates the options that are taken into account when a choice between various options is made:

$$IF\,\{The\ situation\ is\ S\} \texttt{ THEN \{options are } O_1, O_2, ..., O_n\} \tag{34}$$

Given this ensemble of choices, one can choose between them using selection criteria as above (section 6.3): a process of adaptive selection takes place whereby an option is chosen, whether it be physical (going to a bus stop, changing a light bulb) or mental (choosing between theories, making a plan). This generation of options to choose from takes place at the psychological level (Byrne, 2005), assisted by the PLAY primary emotional system (Toronchuk and Ellis, 2013; Ellis and Solms, 2017) which is a key source of creativity. There may be an element of randomness in the options available for consideration at

the psychological level due to the underlying stochasticity at the neural level (Glimcher, 2005; Rolls and Deco, 2010), in turn due to molecular randomness (section 5.4).

## 6.5. The Adaptive Bayesian Brain

The deductive processes of section 6.1 are determined as valid by the brain through adaptive learning processes leading to logical understanding (Churchland, 2013), enabled by underlying brain plasticity. How does the predictive brain (Hawkins, 2004; Clark, 2013) emerge, whereby the brain estimates prediction errors leading to the Bayesian processes of Equation (27) that then enable learning (Friston, 2018) and prediction (Hohwy, 2013)? This is developed in Friston and Stephan (2007), Buckley et al. (2017), and Bogacz (2017).

which show the mechanism whereby such processes can arise in the brain through neural circuits such as shown in Bogacz (2017). Overall, this all emerges from a network of neurons connected by synapses (Kandel et al., 2013), enabled at the microlevel by the branching operation of biomolecules (section 4.2).

# 7. BIOLOGICAL EMERGENCE AND PHYSICAL BRANCHING

How is it possible that goal-oriented systems and deductive logic arise out of the goal-free underlying physics? The context is the hierarchy of emergence and causation, where all the complexities of biology as outlined in section 1.2, occur. Each level of the hierarchy is equally real (Noble, 2012), and branching causation takes place at each level via complex networks of interactions which, through a combination of bottom-up and top down causation, allow organizational closure. Despite the stochasticity of what occurs, the essential core of interactions at the molecular level can be well represented as binary ON/OFF choices (Berridge, 2014). It is at the network level that these individual choices become immensely complex and able to generate the processes of life (section 1.2). How can such branching dynamics emerge from physics which by its nature does not show such branching properties (section 2)? Our main conclusion is,

> *Biomolecules, and specifically proteins (Petsko and Ringe, 2009), provide the physical link between physics and biological causation by allowing branching dynamics at the molecular level, which can then underlie emergence of macro-scale branching dynamics and even deductive causation when incorporated in adaptive modular hierarchical networks. Both the networks and the proteins must have been shaped through processes of adaptive selection; however some of their aspects (that do not hinder their proper function) may be due to drift.*

Ion channels have been our main example, because they enable functioning of the brain, but many other biomolecules in cell signaling networks also carry out logical operations (Berridge, 2014), as do excitatory or inhibitory receptors in neurons (Kandel et al., 2013) with their synaptic thresholds. These branching functions are based in the lock and key mechanism of supra-molecular biology which enables molecular recognition (Lehn, 1995, 2007).

**TABLE 1 |** The three major forms of causation: physical, biological, and deductive. Each relies on the previous one to enable its emergence.

| | Causation | Agency | Outcome | References |
|---|---|---|---|---|
| C1 | Physical | Physical laws | Determinist | Equation (1) |
| C2 | Biological | Goal-seeking, Selection | Adaptive | Equations (11,18) |
| C3 | Deductive | Logical argument | Planned outcomes | Equation (32) |

## 7.1. The Major Distinctions: Three Kinds of Causation

The major difference between physics and life has been characterized above as due to the difference between the immutable impersonal logic of physical causation (1) and the branching functional logic of biological causation (9), enabled by biomolecules in general and proteins in particular (section 4.2).

The progression of emergence is illustrated in **Table 1**. Inanimate systems are subject only to causation C1. In all life from cells to organisms to populations to ecosystems, as well as causation C1, causation C2 occurs, involving logically based branching (9) such as homeostasis (11) and adaptive selection (18). Thus causation C2 characterizes life in general (Hartwell et al., 1999) as opposed to inanimate systems. Hence there is a major difference between these two kinds of emergence out of the same basic physical elements (Ellis, 2016). What enabled causation C2 to emerge in historical terms was the origin of life out of a physical substratum, when both metabolic and adaptive evolutionary processes first came into being. We do not know how that occurred.

However, a higher form of causation C3 occurs in intelligent life, when deductively based action (32) occurs, enabling deductive logic *per se* to have causal powers. Emergence of this kind of causation is a major transition in evolution (Maynard Smith and Szathm, 1995); we also do not know how that occurred. Intelligent organisms are those that can engage in deductive causation C3, which enables transcending the physical limitations of bodies through the power of abstract thought, prediction, planning, and imagination, enabling technology to develop (so that for example they can fly through the sky or make computer systems). It is this kind of causation (made possible by symbolic systems such as language and mathematics) that underlies the rise of civilisation and the domination of humans over the planet (Bronowski, 1973; Harford, 2017): we are no longer limited by the strength of our bodies but by the limits of our imagination and understanding.

Note that we are able to say this without having to make any specific comments on the relation between the brain and consciousness. What is indisputable is that deductive causation does indeed take place in the real world, as demonstrated by many examples (such as the existence of aircraft and computers), and is crucially different than the kind of causation characteristic of physics (section 2), although it is enabled by that kind of causation (which allows the brain to function as it does; Scott, 1995; Kandel et al., 2013).

## 7.2. Branching Physical Causation in a Biological Context

There is however a key underlying question: it is clear that branching dynamics takes place at the biomolecular level, so how then does the underlying physics allow this branching to take place? The physics *per se* does not show such branching dynamics (section 1.1); but physics in a biological context must do so, in order to allow the biological branching processes discussed here to emerge.

The solution (**Figure 6**) is that top-down causation takes place (Ellis, 2016) whereby the local biomolecular context causes bifurcation of the underlying physical dynamics. Firstly, the structural constraint caused by biomolecular shape channels causation at the electron and ion level. Thus for example when a photon releases an electron in a chlorophyll molecule, that is a non-Hamiltonian process that took place because of the biological context of existence of a chlorophyll

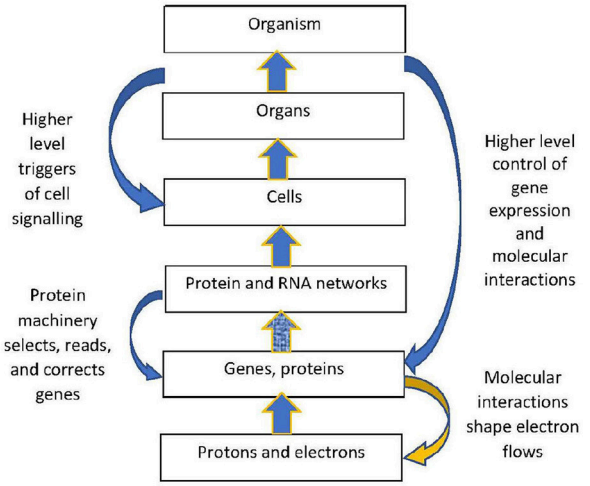


**FIGURE 6 |** Branching physics in a biological context. Branching biological logic chains down to cause the underlying physical dynamics to branch.

molecule in a leaf. This is underlain at the quantum level by contextual wavefunction collapse (Drossel and Ellis, 2018). Secondly, the cell signaling processes at the molecular and cellular level discussed by Berridge (Berridge, 2014) shape how electron flows take place at the underlying physical level, because when a messenger in a signaling pathway turns a process ON, that causes electrons in component molecules to flow in a structured way that would not otherwise have occurred.

In particular, such top down processes take place in the brain (Ellis, 2016; Ellis G., 2018), for example underlying the formation of memory. Eric Kandel states (Kandel, 2001), "*One of the most remarkable aspects of an animal's behavior is the ability to modify that behavior by learning*". He then identifies how this happens at the molecular level as what he calls "*A dialogue between genes and synapses*" (Kandel, 2001). A specific event, say seeing a car crash, results in gene expression that alters synaptic strengths, which is enabled by underlying flows of electrons as indicated in **Figure 6**. The physics acts in such a way as to instantiate the neural connections at the neuron level needed for that memory to be stored and then available for recall at the psychological level. Neural mechanisms such as those discussed by Kandel Kandel (2001); Kandel et al. (2013) and molecular mechanisms such as discussed by Berridge Berridge (2014) enable this to happen, so what happens at the electron level is determined (up to equivalence classes) by the overall social, psychological, and mental context in a top-down way (**Figure 6** omits those higher levels, but they are key parts of the causal context; Ellis, 2016).

In this way, branching physical dynamics at the bottom level emerges from the higher level branching biological dynamics (you might have seen the crash, or not; the outcomes at the electron level are affected by this contingent situation at the psychological level). Physical outcomes are determined by context, which break the symmetries of the underlying physical laws (Anderson, 1972). In the cases we consider, the relevant constraining context is the physical structure of bio-molecules in their cellular context (Hofmeyr, 2017, 2018).

**Biologically generated branching of physical outcomes:** *Biomolecules and cells shape electron flows at the physical level firstly by setting constraints on possible electron flows through their geometric shapes and dispositions (Gray and Winkler, 2009). Second, though signaling processes (Berridge, 2014) originating from higher levels (Noble, 2012) that shape (up to equivalence classes) what electron flows actually take place. This enables branching dynamics occurring in these signaling networks to cause branching outcomes at the electron level.*

This enables physiological processes such as those occurring in the heart (Fink and Noble, 2008) to influence electron flows at the micro-physical level through the top-down influences[14] in physiology described by Noble (2012). Mental processes such as learning (Kandel, 2001) and deductive causation (section 6.3) can do the same, enabled by the ON/OFF operations of cell signaling networks (Berridge, 2014). The way this works during deductive argumentation (section 6.1) is similar to the way algorithms control the flow of electrons in transistors in digital computers. The branching logic of an algorithm, realized in a digital computer program, controls branching electron dynamics (which transistors are ON, allowing electron flows, or OFF, at what time) at the physical level. In that case the physical structure enabling this branching logic at the electron level is the junctions between different layers in transistors[15].

**Biology-physics closure of constraints.** *Extension of the needed functional closure of constraints in biology (section 1.2, Mossio and Moreno, 2010; Montévil and Mossio, 2015) to the underlying physics level is provided by the fact that the branching biological logic at higher levels, including cellular (Rhoades and Pflanzer, 1989; Randall et al., 2002; Berridge, 2014), and mental (Kandel, 2001) functioning, induces congruent branching dynamics at the underlying physical level by changing constraints at that level.*

Equation (2) has to be replaced by

$$C(c(t), \mathbf{X}, t) = C(t) \tag{35}$$

where the time-dependent nature of the physics constraints derives from the time-dependent biological context, and means that the physics evolution is no longer subject to the uniqueness theorems mentioned in sections 2.1, 3.1. This has to be so in order that the biology-physics relation be consistent.

A physics analogy is a pendulum made of a bob of mass $m$ that is constrained to move on a circular arc by a string of length $L(t)$ that varies with time (this is the constraint $C(t)$ governing possible motions of the bob) (Feldman, 2007), see the **Appendix**. The evolution is determined by the macroscopic constraint $C(t)$, which controls outcomes at both macro and micro levels in a way that cannot be predicted from a knowledge of the initial data (starting position $\mathbf{X}_0$ and speed $\mathbf{v}_0$) alone. The dynamics can be controlled by an experimental protocol for $L(t)$ designed by a scientist (which is top down causation $\mathcal{DEOC}$ from the mental level as in section 6.3), or can be unpredictable even in principle, when $L(t)$ is controlled by a computer receiving signals from a detector of particles emitted by decay of a radioactive element (cf. section 2.3).

## AUTHOR CONTRIBUTIONS

GE provided the main idea and drafted the main text. JK contributed further ideas and helped develop the text.

[14] Philosophical objections to this possibility based in the idea of supervenience are countered in (Ellis, 2016) and will be fully refuted in a forthcoming paper (Ellis and Gabriel, 2019).

[15] See for example *Bipolar junction transistor* in Wikipedia.

## ACKNOWLEDGMENTS

We thank Jannie Hofmeyr, Vivienne Russell, Jonathan Birch, Tim Maudlin, Mark Solms, Karl Friston, Jeremy Butterfield, Markus Gabriel, and Jean-Philippe Uzan for very useful comments, and Denis Noble for a very helpful check of the text. We thank two referees and the Editor for comments that have resulted in a significantly improved paper. We thank the University of Cape Town Research Committee and the National Research Foundation (South Africa) for financial support.

## REFERENCES

Ackermann, R. R., and Cheverud, J. M. (2004). Detecting genetic drift versus selection in human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17946–17951. doi: 10.1073/pnas.0405919102

Alam, M. T., (2016). The metabolic background is a global player in Saccharomyces gene expression epistasis. *Nat. Microbiol.* 1:15030. doi: 10.1038/nmicrobiol.2015.30

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2007). *Molecular Biology of the Cell*. Boca Raton, FL: CRC Press.

Amundson, R., and Lauder, G. V. (1994). Function without purpose. *Biol. Philos.* 9, 443–469.

Anderson, P. W. (1972). More is different. *Science* 177, 393–396.

Arnol'd, V. I. (1989). *Mathematical Methods of Classical Mechanics*. Heidelberg: Springer.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman and Hall.

Ashby, W. R. (1960) *Design for a Brain: The Origin of Adaptive Behaviour*. Heidelberg: Springer.

Atmanspacher, H., and beim Graben, P. (2009). Contextual emergence. *Scholarpedia* 4:7997. doi: 10.4249/scholarpedia.7997

Auletta, G., Ellis, G., and Jaeger, L. (2008). Top-down causation: from a philosophical problem to a scientic research program. *J. R. Soc. Interface* 5, 1159–1172. doi: 10.1098/rsif.2008.0018

beim Graben, P. (2016). "Contextual emergence in neuroscience," in *Closed Loop Neuroscience*, ed E. Hady (Amsterdam: Elsevier), 171–184.

Berridge, M. (2014). *Cell Signalling Biology*. London: Portland Press.

Binder, P. M., and Ellis, G. F. R. (2016). Nature, computation and complexity. *Phys. Scripta* 91:064004. doi: 10.1088/0031-8949/91/6/064004

Binney, J., and Tremaine, S. (2008). *Galactic Dynamics*. Princeton, NJ: Princeton University Press.

Blundell, S. J., and Blundell, K. M. K. (2008). *Concepts in Thermal Physics*. Oxford: Oxford University Press.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *J. Math. Psychol.* 76, 198–211. doi: 10.1016/j.jmp.2015.11.003

Booch, G. (1994). *Object Oriented Analysis and Design With Applications*. New York, NY: Addison Wesley.

Bronowski, J. (1973). *The Ascent of Man*. London: BBC.

Buckley, C. L., Kim, C. S., McGregor, S., and Seth, A. K. (2017). The free energy principle for action and perception: a mathematical review. *J. Math. Psychol.* 81, 55–79. doi: 10.1016/j.jmp.2017.09.004

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.

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

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

Carroll, S. B. (2005). *The New Science of Evo Devo - Endless Forms Most Beautiful*. New York, NY: WW Norton and Company.

Catterall, W. A. (2000). From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron* 26, 13–25. doi: 10.1016/S0896-6273(00)81133-2

Churchland, P. M. (2013). *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. Cambridge, MA: MIT Press.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–253. doi: 10.1017/S0140525X12000477

Cockell, C. S. (2018). *The Equations of Life*. London: Atlantic Books.

Darwin, C. (1872). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, 6th Edn*. London: Oxford University Press.

Davies, P. C., and Walker, S. I. (2016). The hidden simplicity of biology. *Rep. Prog. Phys.* 79:102601. doi: 10.1088/0034-4885/79/10/102601

Di Steffano, J. J., Stubberud, A. R., and Williams, I. J. (1967). *Schaum's Outline of Theory and Problems of Feedback and Control Systems*. New York, NY: McGraw-Hill.

Drossel, B., and Ellis, G. F. (2018). Contextual wavefunction collapse: an integrated theory of quantum measurement. *N. J. Phys.* 20:18. doi: 10.1088/1367-2630/aaecec

Edelman, G. M. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. New York, NY: Basic Books, Harper Collins Publishers.

Ellis, G. (2018). Top-down effects in the brain. *Phys. Life Rev.* doi: 10.1016/j.plrev.2018.05.006. [Epub ahead of print].

Ellis, G., and Gabriel, M. (2019). Physical, logical, and mental top-down effects.

Ellis, G., and Solms, M. (2017). *Beyond Evolutionary Psychology*. Cambridge: Cambridge University Press.

Ellis, G. F. R. (2016). *How Can Physics Underlie the Mind? Top-Down Causation in the Human Context*. Heidelberg: Springer.

Ellis, G. F. R., and Kopel, J. (2018). "Wandering towards a goal: the key role of biomolecules," in *Wandering Towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?* eds A. Aguirre, B. Foster, and Z. Merali (Cham: Springer), p. 227–243.

England, J. L. (2013). Statistical physics of self-replication. *J. Chem. Phys.* 139:121923. doi: 10.1063/1.4818538

Epstude, K., and Roese, N. J. (2008). The functional theory of counterfactual thinking. *Pers. Soc. Psychol. Rev.* 12, 168–192. doi: 10.1177/1088868308316091

Farnsworth, K. D., Albantakis, L., and Caruso, T. (2017). Unifying concepts of biological function from molecules to ecosystems. *Oikos* 126, 1367–1376. doi: 10.1111/oik.04171

Farnsworth, K. D., Ellis, G., and Jaeger, L. (2017). "Living through downward causation," in *From Matter to Life: Information and Causality*, eds S. I. Walker, P. C. W. Davies, and G. F. R. Ellis (Cambridge: Cambridge University Press), 303–333.

Feldman, J. (2007). *The Variable Length Pendulum*. Available online at: https://www.math.ubc.ca/feldman/apps/vlPendulum.pdf

Fink, M., and Noble, D. (2008). Noble model. *Scholarpedia* 3:1803. doi: 10.4249/scholarpedia.1803

Flack, J. C. (2017). Coarse-graining as a downward causation mechanism. *Philos. Trans. R. Soc. A* 375:20160338. doi: 10.1098/rsta.2016.0338

Friston, K. (2018). Does predictive coding have a future? *Nat. Neurosci.* 21, 1019-1021. doi: 10.1038/s41593-018-0200-7

Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

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

Gilbert, S. F., and Epel, D. (2009). *Ecological Developmental Biology*. Sunderland, MA: Sinauer.

Glimcher, P. W. (2005). Indeterminacy in brain and behaviour. *Ann. Rev. Psychol.* 56, 25–56. doi: 10.1146/annurev.psych.55.090902.141429

Glymour, B. (2001). Selection, indeterminism, and evolutionary theory. *Philos. Sci.* 68, 518–535. doi: 10.1086/392940

Godfrey-Smith, P. (1994). A modern history theory of functions. *Nôus* 28, 344–362.

Godfrey-Smith, P. (2001). "Three kinds of adaptationism. in *Adaptationism and Optimality*, eds S. H. Orzack and E. Sober (Cambridge: Cambridge University Press), 335–357.

Godfrey-Smith, P. (2017). *Other Minds: The Octopus and the Evolution of Intelligent Life*. Glasgow: William Collins.

Godfrey-Smith, P., and Sterelny, K. (2016). "Biological information," in *The Stanford Encyclopedia of Philosophy (Summer 2016 Edition)*, ed E. N. Zalta (Stanford, CA: Stanford University). Available online at: https://plato.stanford.edu/archives/sum2016/entries/information-biological/

Goelzer, A., Bekkal Brikci, F., Martin-Verstraete, I., Noirot, P., Bessičres, P., Aymerich, S., et al. (2008) Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst. Biol.* 2:20. doi: 10.1186/1752-0509-2-20

Gray, H. B., and Winkler, J. R. (2009). Electron flow through proteins. *Chem. Phys. Lett.* 483, 1-9. doi: 10.1016/j.cplett.2009.10.051

Harford, T. (2017). *Fifty Things That Made the Modern Economy*. London: Little Brown.

Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999) From molecular to modular cell biology. *Nature* 402(Suppl.), C47–C52.

Hawking, S. W., and Ellis, G. F. R. (1973). *The Large Scale Structure of Space-Time.* Cambridge: Cambridge University Press.

Hawkins, J. (2004). *On Intelligence.* New York, NY: Holt Paperbacks.

Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.

Hoffmann, P. M. (2012). *Life's Ratchets: How Molecular Machines Extract Order from Chaos*. New York, NY: Basic Books.

Hofmeyr, J. H. S. (2017). "Basic biological anticipation," in *Handbook of Anticipation*, ed R. Poli (Heidelberg: Springer International Publishing AG).

Hofmeyr, J. S. (2018). Causation, constructors and codes. *Biosystems* 164, 121–127. doi: 10.1016/j.biosystems.2017.09.008

Hohwy, J. (2013). *The Predictive Mind.* Oxford: Oxford University Press.

Huang, Y., and Rao, R. P. N. (2011). Predictive coding. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 580–593. doi: 10.1002/wcs.142

Janes, K. A., and Yaffe, M. B. (2006). Data-driven modelling of signal-transduction networks. *Nat. Rev. Mol. Cell Biol.* 7, 820–828. doi: 10.1038/nrm2041

Kampourakis, K. (2014). *Understanding Evolution.* Cambridge: Cambridge University Press.

Kandel, E., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J. (2013). *Principles of Neural Science.* New York, NY: McGraw Hill Professional.

Kandel, E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science* 294, 1030–1038. doi: 10.1126/science.1067020

Kimura, M. (1983) *The Neutral Theory of Molecular Evolution,* Vol. 241. Cambridge University Press; Scientific American.

Koonin, E. V. (2016). The meaning of biological information. *Phil. Trans. R. Soc.* A374:20150065. doi: 10.1098/rsta.2015.0065

Krebs, E. G. (1993). Protein phosphorylation and cellular regulation I (Nobel lecture). *Angew. Chem. Int. Ed.* 32, 1122–1129.

Lee, M. H., Siddoway, B., Kaeser, G. E., Segota, I., Rivera, R., Romanow, W. J. C., et al. (2018). Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* 563, 639–645. doi: 10.1038/s41586-018-0718-6

Lehn, J. M. (1995). *Supramolecular Chemistry.* Hoboken, NJ: John Wiley

Lehn, J. M. (2007). From supramolecular chemistry towards constitutional dynamic chemistry and adaptive chemistry. *Chem. Soc. Rev.* 36, 151–160. doi: 10.1039/b616752g

Longo, G., Montévil, M. R., and Pocheville, A. (2012). From bottom-up approaches to levels of organization and extended critical transitions.*Front. Physiol.* 3:232. doi: 10.3389/fphys.2012.00232

Longtin, A. (2010). Stochastic dynamical systems. *Scholarpedia* 5:1619.

Lynch, M., and Hill, W. G. (1986). Phenotypic evolution by neutral mutation. *Evolution* 40, 915–935. doi: 10.4249/scholarpedia.1619

Magleby, K. L. (2017). Structural biology: ion-channel mechanisms revealed.*Nature* 541, 33–34. doi: 10.1038/nature21103

Maynard Smith, J., and Szathm, E (1995). *The Major Transitions in Evolution*. Oxford: Oxford University Press.

Mayr, E.(2002). *What Evolution Is.* Basic Books.

Mayr, E.(2004). *What Makes Biology Unique?* Cambridge: Cambridge University Press.

Menzies, P. (2014). "Counterfactual theories of causation," in *The Stanford Encyclopedia of Philosophy,* ed E. N. Zalta. Available online at: https://plato.stanford.edu/archives/spr2014/entries/causation-counterfactual

Millikan, R. G. (1989). In defense of proper functions. *Phil. Sci.* 56, 288–302.

Mitchell, K. J. (2018). Does neuroscience leave room for free will?" *Trends Neurosci.* 41, 573–576. doi: 10.1016/j.tins.2018.05.008

Modell, H., Cliff, W., Michael, J., McFarland, J., Wenderoth, M. P., and Wright, A. (2015). A physiologist's view of homeostasis. *Adv. Physiol. Educ.* 39, 259–266. doi: 10.1152/advan.00107.2015

Montévil, M., and Mossio, M. (2015). Biological organisation as closure of constraints. *J. Theor. Biol.* 372, 179–191. doi: 10.1016/j.jtbi.2015.02.029

Montévil, M., Mossio, M., Pocheville, A., and Longo, G. (2016). Theoretical principles for biology: variation. *Prog. Biophys. Mol. Biol.* 122, 36–50. doi: 10.1016/j.pbiomolbio.2016.08.005

Morris, D. W., and Lundberg, P. (2011). *Pillars of Evolution: Fundamental Processes of the Eco-Evolutionary Process.* Oxford University Press.

Mossio, M., Montévil, M., and Longo, G. (2016). Theoretical principles for biology: organization. *Prog. Biophys. Mol. Biol.* 122, 24–35. doi: 10.1016/j.pbiomolbio.2016.07.005

Mossio, M., and Moreno, A. (2010). Organisational closure in biological organisms. *Hist. Philos. Life Sci.* 32, 269–288.

Mossio, M., Saborido, C., and Moreno, A. (2009). An organizational account of biological functions. *Br. J. Philos. Sci.* 60, 813–841. doi: 10.1093/bjps/axp036

Müller, G. B. (2007). Evo-devo: extending the evolutionary synthesis. *Nat. Rev. Genet.* 8, 943–949. doi: 10.1038/nrg2219

Natarajan, C., Hoffmann, F. G., Weber, R. E., Fago, A., Witt, C. C., and Storz, J. F. (2016). Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* 354, 336–339. doi: 10.1126/science.aaf9070

Neander, K. (1991). Functions as selected effects: the conceptual analyst's defense. *Philos. Sci.* 58, 168–184.

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Mol. Biol. Evol.* 22, 2318–2342. doi: 10.1093/molbev/msi242

Noble, D. (2008). *The Music of Life: Biology Beyond Genes*. Oxford: Oxford University Press.

Noble, D. (2011). Neo-Darwinism, the modern synthesis and selfish genes: are they of use in physiology? *J. Physiol.* 589, 1007–1015. doi: 10.1113/jphysiol.2010.201384

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus* 2, 55–64. doi: 10.1098/rsfs.2011.0067

Noble, D. (2016). *Dance to the Tune of Life: Biological Relativity.* Cambridge: Cambridge University Press.

Noble, D. (2017). Evolution viewed from physics, physiology and medicine. *Interface Focus* 7:20160159. doi: 10.1098/rsfs.2016.0159

Noble, R., and Noble, D. (2017). Was the watchmaker blind? Or was she one-eyed? *Biology* 6:47. doi: 10.3390/biology6040047

Noble, R., and Noble, D. (2018). Harnessing stochasticity: how do organisms make choices? *Chaos* 28:106309. doi: 10.1063/1.5039668

Nurse, P. (2008). Life, logic and information. *Nature* 454, 424–426. doi: 10.1038/454424a

Oizumi, M., Albantakis, L., and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: integrated iNFORMATION theory 3.0. *PLoS Computat. Biol.* 10:e1003588. doi: 10.1371/journal.pcbi.1003588

Okasha, S. (2010). Quick guide: levels of selection. *Curr. Biol.* 20:R306–R307. doi: 10.1016/j.cub.2010.01.025

Oyama, S., Griffiths, P. E., and Gray, R. D. (2001). *Cycles of Contingency: Developmental Systems and Evolution.* Cambridge MA: MIT Press.

Papineau, D. (2016). "Teleosemantics," in *How Biology Shapes Philosophy,* ed D. L. Smith (Cambridge: Cambridge University Press), 95–120.

Penrose, O. (1979). Foundations of statistical mechanics. *Rep. Prog. Phys.* 42, 1937–2006.

Perunov, N., Marsland, R. A., and England, J. L. (2016). Statistical physics of adaptation. *Phys. Rev.* X 6:021036. doi: 10.1103/PhysRevX.6.021036

Peter, I. S., and Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* 144, 970–985. doi: 10.1016/j.cell.2011.02.017

Petsko, G. A., and Ringe, D. (2009). *Protein Structure and Function.* Oxford: Oxford University Press.

Pigliucci, M., and Müller, G. B. (2000). *Evolution - The Extended Synthesis.* Cambridge MA: MIT Press.

Randall, D., Burggren, W., and French, K. (2002). *Eckert Animal Physiology: Mechanisms and Adaptations* New York, NY: W. H. Freeman.

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

Rolls, E. T., and Deco, G. (2010). *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford: Oxford University Press.

Sauro, H. M. (2017). Control and regulation of pathways via negative feedback. *J. R. Soc. Interface* 14: 20160848. doi: 10.1098/rsif.2016.0848

Schroeder, L., and Ackermann, R. R. (2017). Evolutionary processes shaping diversity across the Homo lineage. *J. Hum. Evol.* 111, 1–17. doi: 10.1016/j.jhevol.2017.06.004

Scott, A. (1995). *Stairway to the Mind.* Heidelberg: Springer.

Solms, M., and Friston, K. (2018). How and why consciousness arises: some considerations from physics and physiology. *J. Conscious. Stud.* 25, 202–238. Available online at: https://www.ingentaconnect.com/content/imp/jcs/2018/00000025/f0020005/art00009

Stone, J. V. (2015). *Information Theory: A Tutorial Introduction*. Sebtel Press.

Toronchuk, J. A., and Ellis, G. F. (2013). Affective neuronal selection: the nature of the primordial emotion systems. *Front. Psychol.* 3:589. doi: 10.3389/fpsyg.2012.00589

Uzan, J-P. (2003). The fundamental constants and their variation: observational and theoretical status. *Rev. Mod. Phys.* 75:403. doi: 10.1103/RevModPhys.75.403

Von Baeyer, H. C. (1998). *Maxwell's Demon: Why Warmth Disperses and Time Passes.* New York, NY: Random House.

Wagner, A. (2011). *The Origins of Evolutionary Innovations.* Oxford: Oxford University Press.

Wagner, A. (2017). *Arrival of the Fittest.* New York, NY: Penguin Random House.

Walker, S. I., Kim, H., and Davies, P. C. (2016). The informational architecture of the cell. *Philos. Trans. R. Soc. A* 374:20150057. doi: 10.1098/rsta.2015.0057

Walker, S. W., Davies, P. C. W., and Ellis, G. F. R. (eds.). (2017). *From Matter to Life: Information and Causality*. Cambridge University Press.

Watson, J. D. (2013). *Molecular Biology of the Gene.* London: Pearson.

West, S. A., and Gardner, A. (2013). Adaptation and inclusive fitness. *Rev. Curr. Biol.* 23, R577–R584. doi: 10.1016/j.cub.2013.05.031

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

Wolpert, L. (2002) *Principles of Development.* Oxford: Oxford University Press.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# APPENDIX

## Pendulum With Varying Length

The dynamic equations for a pendulum of varying length are set out clearly in Feldman (2007). A bob on an idealised massless rod swings back and forth about a hinge; the rod angle from the vertical at the hinge is $\theta(t)$. The bob mass is $M$. The bob can slide up or down the rod, so the length $L(t)$ from the hinge to the bob in general varies: $\dot{L}(t) \neq 0$ for some time $t$. The bob position $(x(t), y(t))$ at time $t$ relative to the hinge is

$$x(t) = L(t)\sin\theta(t), \quad y(t) = -L(t)\cos\theta(t), \tag{A1}$$

giving the constraint equation (cf. Eqn.(35))

$$L(t) = \sqrt{x(t)^2 + y(t)^2}. \tag{A2}$$

The kinetic energy $T(t)$ and potential energy $U(t)$ are

$$T(t) = \frac{1}{2}M[\dot{L}(t)^2 + L(t)^2\dot{\theta}(t)^2], \quad V(t) = -MgL(t)\cos\theta(t) \tag{A3}$$

The Lagrangian is

$$\mathcal{L}(t) = T(t) - U(t) = M[\frac{1}{2}\{\dot{L}(t)^2 + L(t)^2\dot{\theta}^2(t)\} + gL(t)\cos\theta(t)] \tag{A4}$$

and the Lagrange equation of motion

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{\theta}}\right) - \frac{\partial\mathcal{L}}{\partial\theta} = 0 \tag{A5}$$

shows that

$$\frac{d^2\theta}{dt^2}(t) + 2\frac{\dot{L}(t)}{L(t)}\dot{\theta}(t) + \frac{g}{L(t)}\sin\theta(t) = 0, \tag{A6}$$

reducing to the standard pendulum equation when $L(t) = L_0 \Leftrightarrow \dot{L}(t) = 0$. The initial data $(\theta(t_0), \dot{\theta}(t_0))$ does not determine the solution $\theta(t)$ for $t > t_1$ if $\dot{L}(t) \neq 0$ at any time $t_1 > t_0$, because of this time-variation of the constraint $L(t)$.

-2020-

# Emergence in Solid State Physics and Biology

George F R Ellis
Mathematics Department, University of Cape Town

July 1, 2020

**Abstract**

There has been much controversy over weak and strong emergence in physics and biology. As pointed out by Phil Anderson in many papers, the existence of broken symmetries is the key to emergence of properties in much of solid state physics. By carefully distinguishing between different types of symmetry breaking and tracing the relation between broken symmetries at micro and macro scales, I demonstrate that the emergence of the properties of semiconductors is a case of strong emergence. This is due to the existence of quasiparticles such as phonons. Furthermore time dependent potentials enable downward causation as in the case of digital computers. Additionally I show that the processes of evolutionary emergence of living systems is also a case of strong emergence, as is the emergence of properties of life out of the underlying physics. A useful result emerges: standard physics theories and the emergent theories arising out of them are all effective theories that are equally valid.

## Contents

# 1 Emergence and Context

A major concern in the interplay between science and philosophy is whether emergence is strong or weak when higher level properties of a system emerge out of the properties of its constituent parts.

## 1.1 Weak and Strong emergence

David Chalmers [Chalmers 2000] defines weak and strong emergence as follows

- **Weak Emergence of phenomena** We can say that a high-level phenomenon is weakly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are unexpected given the principles governing the low-level domain ... It often happens that a high-level phenomenon is unexpected given principles of a low-level domain, but is nevertheless deducible in principle from truths concerning that domain.

- **Strong Emergence of phenomena** We can say that a high-level phenomenon is strongly emergent with respect to a low-level domain when the high-level phenomenon arises from the low-level domain, but truths concerning that phenomenon are not deducible even in principle from truths in the low-level domain.

I will accept those definitions; the latter is the concern of the body of paper. Note that in dealing with emergence from a base theory, one will always of necessity make approximations of various kinds, or take limits of the underlying theory. The key issue raised by Chalmers' definition of strong emergence is the following:

- **Strong and Weak Derivation of Phenomena** Given that approximations or limiting procedures will be used in determining emergent phenomena, (i) are the variables used in this process deducible in principle from the lower level properties alone, together with a generic procedure for deducing higher level phenomena? Or (ii) are they deducible from the lower level properties only when one adds concepts or variables that are derived from our knowledge of the nature of the higher level

emergent domain? The former case is strong derivation, indicating weak emergence, and the latter case is weak derivation, indicating strong emergence.

The point is that in case (ii), the lower level simply does not by itself imply the relevant higher level concepts or variables. This is key: proposed interlevel bridge laws do not determine the higher level variables, but rather depend on them, as when [Butterfield 2014] states one needs empirical input from the higher level theory to determine such bridge laws. Strong upward emergence should determine these higher variables strictly from lower level concepts together with appropriate limits, approximations, and 'upscaling' procedures (such as averaging) alone.

However 'phenomena' is (deliberately) a very broad concept: it could mean electrical conductivity or rigidity, galaxies or planets, life or amoebae or giraffes, human beings or consciousness. This discussion can be extended to a related but more specific concept:

- **Strong Emergence of Specific Outcomes** We can say that a high-level specific outcome is strongly emergent with respect to specific low-level data when the high-level outcome arises from the low-level data, but the details of that outcome is not deducible even in principle from data in the low-level domain alone.

Claims that this cannot occur are sometimes used in arguments against strong emergence; I will deal with this in Section 6.1.

**Higher Level Variables** As just indicated, a key point in the case of strong emergence is the need for "new concepts" at higher levels [Anderson 1972]. Whenever such occur, this is reason enough to think that the phenomena at the higher level are strongly emergent. But the issue then is, Why are these new concepts indispensable? Why would it not in principle be possible to translate them with the help of lower level concepts? If there is a "dictionary" relating different concepts at different levels, then emergence does not seem to introduce anything new. It rather seems to point to our limited ability to express complex emergent phenomena in terms of lower-level concepts.

I will show how such higher level variables do indeed occur as I consider strong emergence in various cases. In particular, this arises because of the issue of Multiple Realisation discussed in Section 4.7. I will consider the possibility of such a dictionary in Section 5.4 and give examples of irreducible higher level variables in Section 6.1.

## 1.2 Contexts

One may be interested in emergence in regard to broad areas of science such as physics, biology, or the mind/brain complex. Indeed a huge part of the literature on emergence is concerned with the latter case. However there are very interesting issues as regards emergence in the case of physics as well.

Given a choice of broad area, emergence may be rather different in different subfields in that area. If one is dealing with physics, then in the cases of particle physics or nuclear physics, use of Effective Field Theories (EFTs) [Hartmann 2001] [Castellani 2002] [Luu and Meißner 2019] is a powerful method. This often shades over into renormalisation group methods [Wilson 1982], as for example in [Burgess 2007]. However I will deal with somewhat different contexts of emergence: firstly solid state physics, giving details of how this works out for transistors in digital computers, and then broadly the case of biology.

**Solid state physics** In the case of solid state physics, an interesting situation arises as regards the coming into being of the relevant crystals. Crystallization is a case of weak

emergence, arising due to energy minimisation as the crystal forms. However when either metals or semiconductors occur, they only exist because of the very specific manufacturing processes required to bring them into being; they are the outcome of purposeful design and manufacture [Bronowski 2011] [Mellisinos 1990]. This is obvious from the fact that neither metals nor semiconductors occur naturally.

Once they are in existence, their properties may be strongly emergent for two quite different reasons. In cases of everyday emergence such as electrical and optical properties of metals and semiconductors, as well as in more exotic cases such as superconductivity, broken symmetries are the key, as forcefully pointed out by Phil Anderson [Anderson 1972] [Anderson 1981] [Anderson 1989] [Anderson 1994]. The emergent lattice structure in semiconductors leads to the existence of quasi-particles that inextricably link the lattice and electron levels via a quantum physics effect of interlevel wave-particle duality. This will be the focus of the first part of this paper. I will not consider topological effects [Kvorning 2018], such as occur in the Fractional Quantum Hall Effect, topological insulators, and some cases in Soft Matter Physics, even though they are excellent examples of strong emergence [McLeish *et al* 2019]. Neither will I deal with defect structures which lead to symmetry breaking with a complex topological nature ([Anderson 1984]:54-67).

**Biology** In the case of biology, the coming into being over long periods of time of species is a case of strong emergence because of the contextual nature of the selection process [Campbell 1974] [Mayr 2001]. Then functioning at the molecular biology level is a case of strong emergence of properties due to the breaking of symmetry by molecular structure, allowing time dependent constraints in the underlying Hamiltonian [Ellis and Kopel 2019].

## 1.3 Downward causation and equality of levels

There are two key points in the discussion. Firstly, I will claim strong emergence is the result of a combination of upward and downward causal effects. It is the downward influences that enable higher levels to shape lower level dynamic outcomes in accord with higher level emergent dynamics and needs. The existence of such downwards effects is denied by many physicists, but has been argued for strongly in the case of biology [Campbell 1974] [Noble 2008] and in science more generally [Ellis 2016], including e.g. the case of structure formation in cosmology on the one hand and the functioning of digital computers on the other. In this paper I give arguments for existence of downward causal effects in solid state physics (Section 4.4) as well as in biology (Section 5.3).

As a consequence of this confluence of effects, Effective Theories (ETs) (which may or may not be Effective Field Theories) emerge at each level of the hierarchy, characterised by a vocabulary and dynamics applicable at that level. The issue then is, Is there a fundamental level that is more basic and has more causal power than all of the others?

It seems to be assumed by most physicists that this is indeed the case - there is an underlying most basic level (a 'Theory of Everything', or TOE) from which all else follows. However the problem is that we do not have such a well-defined and tested fundamental theory to which all else could in principle be reduced. We do not even have a proof that such a level exists. That supposition is an unproven philosophical assumption, which may or may not be true. In any case alleged reduction to an ill-defined theory is problematic.

As a consequence, all the theories that physicists deal with in practice are Effective Theories [Castellani 2002], assumed to emerge from that unknown underlying TOE. None of them is *the* underlying theory. And in consequence physicists use whatever theory is convenient for a particular application, not worrying about the fact that it is not the alleged TOE. Various levels may be chosen as the 'lowest' level in any particular study.

Denis Noble formalized this feature in the case of biology by his 'Principle of Biological Relativity' [Noble 2012], stating that all emergent levels in biology should be regarded as equally causally efficient: none of them has supremacy over the others. In this paper (Section 6.4) I extend that statement to a generally applicable *Principle of Equal Causal Validity*: all emergent levels, whether physical or biological, are of equal causal validity in that they support an Effective Theory that may be considered to be valid at that level, and is not trumped by effective laws at any other level. And this is the case because of the downward causation that occurs equally with upward causation.

### 1.4 This paper

In the following, Section 2 introduces some key conceptual distinctions that underlie the analysis in the rest of the paper. Section 3 considers issues relating to emergence in solid state physics: the significance of broken symmetries, problems with purely bottom up derivations of properties, and the nature of quasiparticles such as phonons.

Section 4 demonstrate that strong emergence occurs in condensed matter physics (§4.3). Section 4.4 shows that standard microdynamics **m** is causally incomplete unless one includes downwards effects from higher levels. Section 4.6 shows how this all works in the case of transistors in digital computers. Section 4.7 comments on the multiple realisability of higher level states by lower level states, which is a key aspect of strong emergence.

Section 5 considers emergence and life. Life involves function and purpose that are emergent characteristics that cannot be expressed in terms of lower level variables (§5.1). The existence of life is a case of strong emergence due to natural selection (§5.2). Furthermore properties of all living systems are strongly emergent because of the symmetry breaking due to the existence of molecular structure (§5.3), allowing downward causation via time dependent constraints. Thus life is strongly emergent for multiple reasons (§5.5).

Section 6 discusses essentially higher level variables (§6.2) the important result that all emergent levels are equally causally valid in terms of supporting an effective theory (§6.4).

## 2 Conceptual Issues

This paper uses the idea of an Effective Theory (Section 2.1). The conceptual bases of the present paper are twofold: first, characterizing different aspects of emergence (2.2); second, characterizing different kinds of symmetry breaking (Section 2.3).

### 2.1 Effective Theories

Elena Castellani gives this definition [Castellani 2002]:

> "*An effective theory (ET) is a theory which 'effectively' captures what is physically relevant in a given domain, where 'theory' is a set of fundamental equations (or simply some Lagrangian) for describing some entities, their behaviour and interactions... More precisely, an ET is an appropriate description of the important (relevant) physics in a given region of the parameter space of the physical world. ... It is therefore an intrinsically approximate and context-dependent description.*"

A key result is that all emergent levels are equally causally effective in the sense that the dynamics of every level **L** we can deal with in an empirical way is described by an Effective Theory $\mathbf{EF_L}$ at that level (see (8)) that is as valid as the effective theories at all other levels (§6.4). There is no privileged level of causality [Noble 2012].

### 2.2 Different aspects of emergence

I distinguish two different aspects of the nature of emergence as follows.

- Emergence **E** of system from its components. For example, the emergence **E** of nuclei out of protons and neutrons, of water or a metal or hemoglobin molecules out of the underlying nuclei and electrons, or of a human body out of its constituent cells. The issue of phase transitions is important here. These occur when a major change in the emergent state takes place, such as the transition of water from a liquid to a gaseous state when boiling occurs.

- Emergence **P** of properties of the emergent system out of its underlying constituents once it has come into existence. How do properties of a nucleus arise out of the nature of its constituent neutrons and protons, and theirs out of the constituent quarks? How do rigidity or electrical conductivity or optical properties of a crystal, or chemical properties of a molecule, arise out of the underlying electrons, protons and neutrons? How do properties of a cell in a human body arise out of properties of its underlying biomolecules? How does behaviour arise out of those cells?

The issue of whether emergence is strong or weak arises in both cases.This paper is concerned with **P** rather than **E**, except in the case of biology where it matters what kinds of entities are present in the world today: **E** underlies **P** (Section 5).

### 2.3 Different kinds of symmetry breaking

Phil Anderson in his famous paper "More is Different" [Anderson 1972] and later writings [Anderson 1984] [Anderson 1989] [Anderson 1994] emphasized the crucial importance of symmetry breaking for emergence in condensed matter physics. For the purposes of this paper, it is useful to distinguish three different types of symmetry breaking occurring in the context of the relation of macro dynamics **M** to microdynamics **m**.

- **SSB**(**m**), Spontaneous Symmetry Breaking **SSB** occurring at the micro level **m**;

- **SSB**(**M**), Spontaneous Symmetry Breaking **SSB** occurring at the macro level **M** due to the process **E** creating that level from the micro level **m**, that is **E**: $\mathbf{m} \rightarrow \mathbf{M}$;

- **SB**(**NS**), Symmetry breaking **SB** in biology that has occurred due to Darwinian processes of natural selection **NS** leading to existence of specific kinds of molecules at the micro level **m** [Wagner 2014].

These differences will be key below.

## 3 Emergence and Solid State Physics

Solid state physics is a particular well-trodden area in which to investigate emergence of properties. This section sets the stage for that investigation in the next section.

It considers in turn, the key feature of Broken Symmetries in solid state physics (Section 3.1), problems arising in trying to determine properties of solid state systems in a purely bottom up way (Section 3.2), and the key role of phonons in their properties (Section 3.3). Section 3.4 views this in another way: in terms of Downward Emergence.

## 3.1 Broken Symmetries

Broken symmetries are key to emergence in condensed matter physics [Anderson 1972], [Anderson 1981]. Spontaneous symmetry breaking naturally leads to a class of emergent entities with broken symmetries.

The simplest example is a crystalline state ([Anderson 1989]:587):

> *"Why do we call the beautifully symmetric crystalline state a 'broken symmetry? Because, symmetrical as it is, the crystal has less symmetry than the atoms of the fluid from which it crystallized: these are in the ideal case featureless balls, while the crystal has no continuous symmetry or translation symmetry*".

Thus the symmetry of a crystal breaks the translational invariance of the theory that underlies it, and leads to the Bloch wave functions that are key to much of solid state physic. By Bloch's theorem, the electron wave functions in a crystal has a basis of Bloch wave energy eigenstate. This feature underlies the nature of electronic band structures, and thus for example determines electric resistivity and optical absorption, which are measurable phenomena. What this means is that physical properties of the emergent state (the specific crystal structure) reach down to influence the motions of electrons moving in the lattice. That's a downward effect underlying key physical properties of semiconductors [Ellis 2016]. I return to this in §3.3.

## 3.2 Problems with bottom up derivations

Anthony Leggett in his article "On the nature of research in condensed-state physics" ([Leggett 1992], quoted in [Drossel 2020]) writes as follows:[1]

> *"No significant advance in the theory of matter in bulk has ever come about through derivation from microscopic principles. (...) I would confidently argue further that it is in principle (my emphasis) and forever impossible to carry out such a derivation. (...) The so-called derivations of the results of solid state physics from microscopic principles alone are almost all bogus, if 'derivation' is meant to have anything like its usual sense.*"

Thus he is proclaiming strong emergence. He gives Ohm's Law as a specific example:

> *"Consider as elementary a principle as Ohm's law. As far as I know, no-one has ever come even remotely within reach of deriving Ohm's law from microscopic principles without a whole host of auxiliary assumptions ('physical approximations'), which one almost certainly would not have thought of making unless one knew in advance the result one wanted to get, (and some of which may be regarded as essentially begging the question)*" [Leggett 1992].

Thus one often needs some additional and distinct assumptions in order to derive an effective theory. These assumptions are usually called "auxiliary assumptions" and do not necessarily involve approximations. This is what happens when try we derive an effective theory in the bottom-up fashion, but need some "extra information" about the low energy degrees of freedom (e.g., some new symmetry principle) in order to do so.

Can one derive the results in a purely bottom up way? No. One can derive the micro-macro relation in an upward way provided one inserts some of the macro picture

[1] Basically the same is true in quantum chemistry, see [Hunger 2006].

either into the effective Lagrangian as a symmetry breaking term (see §4), or into the micro-macro relation, as Weinberg does in his effective field derivation of superconductivity [Weinberg 2009]: *"instead of counting powers of small momenta, one must count powers of the departures of momenta from the Fermi surface. Without that ingredient, the derivation fails"*. The point then is that you can only derive the Fermi surface position from the macro theory. Thus it is not a purely bottom up derivation. If you try one, you will fail, as stated strongly by [Laughlin 1999], so it is weak derivation.[2]

### 3.3 Phonons and Quasiparticles

The problem in the bottom up derivation of emergent properties in solid state physics is not just a question of not having enough computing power. It is a question of having the right concepts at hand. And you can't get those by studying the low level dynamics *per se*, in most cases, because they involve high level concepts.

The key feature is that Spontaneous Symmetry Breaking (**SSB**) takes place, so the symmetries of the equations are not shared by the solution [Anderson 1984]. This leads to the properties **P** of emergent highly ordered structures such as crystals that, through their ordered nature, lead to everyday properties such as stiffness and optical and electrical properties, but also can lead lead to complex behaviours such as superconductivity and superfluidity. This emergence is an essentially quantum phenomenon with two key aspects.

**Interlevel Wave-particle duality** The existence of quasiparticles such as phonons due to the broken symmetries of the emergent lattice structure is a situation where they come into being at the lower level because they are dynamically equivalent to collective vibrations of a higher level structure (the crystal lattice) ([Ziman 1979]:60, [Goodstein 1985]: 154-155, [Phillips 2012]:172-175, [Lancaster 2019]). While they are vibrational modes of the lattice as a whole, and hence emergent entities, they nevertheless have particle like properties. Steven Simon expresses it like this ([Simon 2013]:82-83):

> "*As is the case with the photon, we may think of the phonon as actually being a particle, or we can think of the phonon as being a quantized wave. If we think about the phonon as being a particle (as with the photon) then we see that we can put many phonons in the same state (i.e., the quantum number n can be increased to any value), thus we conclude that phonons, like photons, are bosons.*"

Thus the lattice vibrations at the macro scale are dynamically equivalent to a particle at the micro scale. Adding them to the micro scale description as quasi-particles such as phonons, one now has a symmetry breaking micro theory that can be the basis of for derivation of a symmetry breaking macro theory. The phonons are an essential element of the micro theory (see [Simon 2013], pp. 82 on). This is an essentially quantum phenomenon: a form of the standard wave-partical duality of quantum physics.

> **Wave-particle duality: There exists a *wave(macro)-particle(micro) duality* in crystal structures which provides the crucial interlevel link in emergence of properties.**

This is key to what happens physically. [Guay and Sartenaer 2018] discuss in depth the philosophical implications of the existence of quasi-particles.

[2]Reminder: I am using David Chalmer's definitions, see Section 1.1.

### 3.4 Downward emergence

One can view this in another way. A crystal structure causes existence of phonons at the micro level, a case of downward emergence [Franklin and Knox 2018].[3] These lead to the emergent properties that occur at the macro level in semiconductors [Ziman 1979], [Grundmann 2010], [Simon 2013]. Thus downward effects shape the lower level (electron) dynamics due to the higher level (crystal) structure, these effects not being directly implied by the lower level interactions by themselves.

> **Downward emergence: The higher level context alters the lower level dynamics by introducing into it quasi-particles such as phonons that play a crucial role in solid state physics**

But a lingering doubt remains: are they real, or fictitious? Stephen Blundell gives the answer ([Blundell 2019]:244),

> *"So now we come to the key question: Are these emergent particles real? From the perspective of quantum field theory, the answer is a resounding yes. Each of these particles emerges from a wave-like description in a manner that is entirely analogous to that of photons. These emergent particles behave like particles: you can scatter other particles off them. Electrons will scatter off phonons, an interaction that is involved in superconductivity. Neutrons can be used to study the dispersion relation of both phonons and magnons using inelastic scattering techniques. Yes, they are the result of a collective excitation of an underlying substrate. But so are 'ordinary' electrons and photons, which are excitations of quantum field modes."*

The Cooper pairs responsible for superconductivity are similarly downwardly emergent lower level effective variables, which can be regarded as produced either by electron-phonon interactions at the lower level, or by crystal distortions at the higher level. They would not exist were it not for specific kinds of crystal structure.

## 4 Strong Emergence in Solid State Physics

This section is the heart of the argument on emergence in relation to solid state physics. Section 4.1 clarifies what will be assumed to be the underlying microphysics. Section 4.2 shows that while Spontaneous Symmetry Breaking at the micro scale ($\mathbf{SSB}(\mathbf{m})$) can lead in principle to symmetry breaking at the macro scale, this is not a significant feature of the emergence of properties in solid state physics in practice.

Section 4.3 demonstrates that strong emergence takes place in solid state systems whose emergent properties are a result of Spontaneous Symmetry Breaking $\mathbf{SSB}(\mathbf{M})$ occurring via emergence processes $\mathbf{E} : \mathbf{m} \to \mathbf{M}$. Although the macro dynamics $\mathbf{M}$ do not follow from the micro dynamics $\mathbf{m}$ because they do not have the correct symmetries to allow this, one can nevertheless obtain an *Effective Theory* **m'** at the microscale by introducing quasiparticles such as phonons. This implies downward causation takes place, because their existence depends on the existence of the crystal structure. This leads to the statement that all the well established Laws of Physics are Effective Theories.

Section 4.4 examines this feature of downward causation and its relation to the alleged Causal Completeness of physics at the micro scale, and turns the usual argument on its

[3] Called a "Foundational Determinative Relation" (FDR) by Carl Gillett, see [Gillett 2019].

head: I claim that microphysics **m** cannot in fact be causally complete, because it is unable by itself to lead by any coarse graining process to the correct macro dynamics **M**. Section 4.5 on downward causation and effective laws confirms this result. To give flesh to this story, Section 4.6 discusses the case of transistors in a digital computer.

An underlying issue of importance is that multiple realisability of higher level structures and functions in terms of lower level structures and functions underlies downwards causation, and hence strong emergence. I discussion this in Section 4.7.

## 4.1 Setting the Foundations

The issue is whether strong emergence takes place in physics.

**Microphysics** To be clear about the context, what kind of micro theory **m** do people usually have in mind? Robert Bishop clarifies as follows, in relation to patching physics and chemistry together [Bishop 2005]:

> *"In quantum chemistry, one first specifies the fundamental physical interactions (electromagnetic, strong- and weak-nuclear, etc.), then enumerates the relevant particles and their properties (nucleon, electron, charge, mass, etc.). Next, one lists the pairwise interactions among the particles. Finally, one writes down the kinetic and potential energy operators and adds them to get the system Hamiltonian (an expression for the total energy of the system). With the Hamiltonian in hand, one then proceeds to derive the properties and behaviors of the chemical system in question."*

But what is the specific set of interactions one should take into account in **m**? For ordinary life, depending on the scale considered, it will be Newton's Laws of Motion, together with the appropriate force laws. These are given by Maxwell's equation of motion, together with Maxwell's equations for the electromagnetic field. One may need the Schrödinger equation, or maybe Quantum Field Theory (QFT) and the Dirac equation.

An interesting issue is gravitation. For ordinary laboratory scale physics we do not need Newton's Gravitational Laws, but if the scale is large enough, we may need to include Galilean gravity represented by an effective acceleration **g**. This is a symmetry breaking contextual term resulting from downward causation due to the existence of the Earth with a specific mass $M_E$ and radius $R_E$. This can have a significant effect on atomic scale phenomena, as in the case of atomic fountains underlying the existence of cesium fountain clocks. It will not play a role in the solid state physics considered here such as physics of transistors, although of course it does play an important role in biology and engineering.

The above is what I will have in mind in referring to microphysics **m** below.[4]

**Broken symmetries** As mentioned above, Anderson claims that the key to emergence in solid state physics is broken symmetries, which occurs via Spontaneous Symmetry Breaking (**SSB**). However there is a crucial issue here. **SSB** can take place because of the way the macrostructure **M** comes into existence. As discussed in Section 2, I call these cases **SSB**(**M**). But **SSB** can also take place at the micro level **m**. I call this **SSB**(**m**). This distinction plays a key role in the discussion below.

[4]This is made explicit by [Laughlin and Pines 2000] in their equations [1] and [2]; they say "Eqs. [1] and [2] are, for all practical purposes, the Theory of Everything for our everyday world."

## 4.2 The case of SSB(m)

Here the needed spontaneous symmetry breaking takes place at the micro scale, for example via the Higgs mechanism resulting from the Mexican Hat shape of the Higgs potential. Then macro symmetry breaking is in principle deducible in a bottom up way through coarse graining this symmetry-broken micro state. Weak emergence takes place. In notional terms, if the coarse graining operation **C** commutes with the symmetry **S**, then[5]

$$\{\mathbf{SSB(m)} : \mathbf{S(m)} \neq \mathbf{m},\ \mathbf{C(m)} = \mathbf{M},\ \mathbf{CS} = \mathbf{SC}\} \Rightarrow \mathbf{S(M)} \neq \mathbf{M}. \quad (1)$$

However the known symmetry breaking mechanisms producing their symmetry breaking effects at the microscale level have limited higher level effects. While their knock-on effects reach up to scales relevant to the Standard Model of Particle Physics, they do not, in a daily life context,[6] reach up to the scale of atomic physics or higher, and so do not affect phenomena such as solid state physics or chemistry or microbiology. The point is that the masses of quarks and gluons in the Standard Model (and hence the Higgs mechanism) are so much smaller than the mass of composite entities at larger distance scales (neutrons, protons, atoms, etc.) because the Electroweak scale is 246 GeV while the neutron and proton mass is 940 MeV, a difference of 3 orders of magnitude. Therefore "*energies of relevance to nuclei can never reveal their quark-gluon substructure, it is simply irrelevant at these energies*" [Luu and Meißner 2019]. Consequently we conclude

> **Conclusion** ***SSB(m) cases can produce the required symmetry breaking at higher levels M through weak emergence. However while of fundamental importance in the standard model of particle physics, this is not relevant to situations such as the emergence of properties of semiconductors and metals.***

## 4.3 The case of SSB(M)

Here the needed symmetry breaking does not take place at the micro scale, but rather through emergence of the macro scale, due to interactions that minimise energy of the emergent structure; think of atoms crystallizing to form a crystal, for instance. This is spontaneous emergence of the breaking of symmetry **S** at the macro scale **M**, hence is a case of **SSB**(**M**). In notional terms, if **E** is the process of emergence,

$$\{\mathbf{S(m)} = \mathbf{m},\ \mathbf{E(m)} = \mathbf{M}\} \Rightarrow \mathbf{M} :\ \mathbf{S(M)} \neq \mathbf{M}. \quad (2)$$

The emergence process **E** alters the macro symmetry by processes of energy minimisation based in the underlying microphysics **m**, and this is weak emergence. But that is not the concern here. The issue is **P**: how do we determine the physical properties of the emergent structure **M**, such as electrical conductivity, once it exists?

As discussed above (§3.3), existence of the macrostructure changes the context of the microphysics, resulting in effective interactions with quasiparticles such as phonons.[7] Thus the emergent broken symmetry at the macro scale reaches down to affect conditions at the

[5] Complexities can arise regarding the commutation here because of the "≠" relation. The implication "⇒" here and in (5) should be read "can imply" rather than "implies". This complication does not apply to (6) because that implication does not use commutation of an inequality.

[6] I am excluding discussion of the context of the very early universe where cosmic inflation took place. In that case, **SSB**(**m**) had a major effect at macro scales.

[7] See the Blundell quote in Section 3.4.

micro scale, thereby causing effective symmetry breaking at that scale. This produces from **m** an effective micro theory **m'** that breaks the symmetry **S**, for example by including phonons or other quasi-particles in the dynamics, and that therefore can produce the required symmetry-broken macro theory by coarse graining. In notional terms

$$\mathbf{SSB}(\mathbf{M}) \Rightarrow \mathbf{m} \rightarrow \mathbf{m'}: \ \mathbf{S}(\mathbf{m'}) \neq \mathbf{m'}. \tag{3}$$

Typically this is done by introducing at the micro level a variable $\mathbf{a}(\mathbf{M})$ that is determined by macro level conditions, and breaks the symmetry **S**:

$$\{\mathbf{m} \rightarrow \mathbf{m'} = \mathbf{m'}(\mathbf{a}), \ \partial \mathbf{m'}/\partial \mathbf{a} \neq 0, \ \mathbf{S}(\mathbf{a}) \neq \mathbf{a}\} \Rightarrow \mathbf{S}(\mathbf{m'}) \neq \mathbf{m'}. \tag{4}$$

This is the way that ETs for physics involving symmetry breaking can be derived: you apply a coarse graining process **C** to **m'**, not **m**, and thereby obtain the macro theory **M** that breaks the relevant symmetry and accords with experiment: that is,

$$\{\mathbf{C}(\mathbf{m'}) = \mathbf{M}\} \ \Rightarrow \ \mathbf{S}(\mathbf{M}) \neq \mathbf{M}. \tag{5}$$

How this works out in the case of functioning of transistors is explained in Section 4 of [Ellis and Drossel 2019], see Section 4.6 below. In some approaches to quantum chemistry, this is done via the Born-Oppenheimer approximation (cf.§4.5)). Again one replaces the basic micro theory **m**, as described in [Bishop 2005] (quoted above), with an effective theory **m'** that incorporates the approximations needed to make it work. In this case, one uses the fact that nuclei are far heavier than electrons to derive a new micro theory **m'** that breaks translational symmetry and gives the required results. It is not the same as the theory you started with: it is the lower level effective theory you need.

Why does one introduce the effective theory **m'** at the low level, rather than using the fundamental theory **m** based only in Newton's Laws, QFT, Maxwell's equations, and so on? The answer is that if we use **m**, unless the coarse graining operation **C** explicitly breaks the symmetry **S**, the correct emergent results cannot even in principle be deduced in a purely bottom up way. In notional terms,

$$\{\mathbf{S}(\mathbf{m}) = \mathbf{m}, \ \mathbf{C}(\mathbf{m}) = \mathbf{M}, \ \mathbf{CS} = \mathbf{SC}\} \ \Rightarrow \ \mathbf{S}(\mathbf{M}) = \mathbf{M} \tag{6}$$

contradicting the known macrolevel symmetry breaking of **M**. You cannot get **M** from **m** via any coarse graining **C** that does not explicitly break the symmetry **S**, so that $\mathbf{CS} \neq \mathbf{SC}$. This could happen if (cf. Eqn.(4)), $\{\mathbf{C} \rightarrow \mathbf{C'} = \mathbf{C'}(\mathbf{a})\} \Rightarrow \mathbf{SC'} \neq \mathbf{C'S}$, but then you are specifically introducing a macro-determined term into this averaging process.

> **Conclusion:** ***Strong emergence of properties P(d) takes place when SSB(M) occurs: the symmetry broken higher level dynamics cannot even in principle be obtained by coarse graining the fundamental theory m because S(m) = m. One has to coarse grain the effective micro theory m', of such a nature that S(m') ≠ m', to get the correct result M by coarse graining. This covers the way emergence of properties occurs in condensed matter physics and in quantum chemistry.***

In short: you have to add a symmetry breaking term into the micro theory in order to get the correct macro theory, because it's not there in the fundamental physics; but you only can work out what symmetry breaking term to add from your knowledge of the correct macro theory **M**. You have to use variables defined by that higher level theory to get the symmetry broken effective micro theory **m'** which gives the correct macro result **M**.

**Change of symmetry changes weak to strong emergence** When **SSB**(M) takes place, weak emergence **E** occurs via a *First Order Phase Transition.*

([Binney et al 1992]:1-2) describe it thus:

> "*Under ordinary circumstances the phase transitions of $H_2O$, or the solidification of a molten metal, are 'first-order' phase transitions ... that involve latent heat. When a material makes a first order phase transition from a high temperature phase to a low temperature phase, a non-zero quantity of heat, the latent heat, is given out as the material cools through an infinitesimally small temperature around the transition temperature $T_t$. This emission of heat at the transition tells us that the structure of the material is being radically altered reordered at $T_t$. ... Above the freezing point of water, there is no crystal lattice. Below the freezing point, the lattice is well defined even if not free of imperfections ('defects'). The transition from water to ordered ice is an all or nothing affair. Either the lattice is there and the vast majority of $H_2O$ molecules are comparatively tightly bound, or there is no lattice and the molecules are not optimally packed.*"

In the light of the above discussion, its meaning is as follows:

> **First order phase transitions:** ***As the temperature $T$ decreases through the transition temperature $T_t$, Spontaneous Symmetry Breaking (SSB) takes place. Weak emergence E occurs as the material goes from a more symmetric to a less symmetric state, the binding energy of the less symmetric state being equal to the latent heat given out. Consequent on this symmetry change, the properties P of the substance change from being weakly emergent to strongly emergent.***

This is similar to what happens when molecules form out of atoms A key comment is the following: when this takes place the relevant Hilbert space of the problem changes completely. This is pointed out in ([Anderson 1984]:126) in the case of molecules: "*A change of this sort puts the system, in the limit $N \to \infty$, into a wholly different, orthogonal Hilbert space, from which there is no easy continuous method of return*". This is how properties **P** become strongly emergent when a crystal or molecular structure arises through weak emergence **E** via **SSB**(**M**). [Anderson 1984] emphasizes that SSB leads to singular points and thus a breakdown in continuity.

**Giving more details** The above argument traces the key causal relations in what is going on. To give a more detailed proof, one would have *inter alia* to parse the micro dynamics **m** into a reliable and unchanging relation **L** ('the Laws of Physics') between initial conditions described by data $d$ and outcomes $o$, which is valid in some domain $\mathcal{D}$. That is,

$$\mathbf{L} : d \in \mathcal{D} \to \mathbf{L}[d] = o \in \mathcal{D} \tag{7}$$

in a reliable way, whether **L** is an exact or statistical law.[8] One would have to represent the emergent macro dynamics **M** in a similar way to (7), and then carry out an analysis analogous to that above.

I have not attempted this more complex project here because I do not believe it would throw much light on what is going on: rather it would probably obscure the key relations. I

[8]Note that the laws of physics are not algorithms - it's Newton's Laws of Motion, not Newton's algorithm - and they do not compute, see [Binder and Ellis 2016].

believe that the symbolism used above enables one to clearly understand the relevant causal links in a adequate way. However I will develop Equation (7) in important ways below: it will be modified to provide *Effective Laws* in various contexts. Specifically, Equation (9) applies in the case of dynamics with a potential shaped by a crystal structure, and Equation (19) when electron dynamics is controlled by molecular constraints.

It is important to relate (7) to emergent levels of complexity [Ellis 2020]. One can characterise an Effective Theory $\mathbf{ET_L}$ valid at some level $\mathbf{L}$ as follows:

> *An* **Effective Theory $\mathbf{ET_L}$** *at an emergent level $\boldsymbol{L}$ is a reliable relation between initial conditions described by effective variables $v_{\boldsymbol{L}} \in \boldsymbol{L}$ and outcomes $o_{\boldsymbol{L}} \in \boldsymbol{L}$:*
>
> $$\mathbf{ET_L} : v_{\mathbf{L}} \in \mathbf{L} \to \mathbf{ET_L}[v_{\mathbf{L}}] = o_{\mathbf{L}} \in \mathbf{L} \tag{8}$$
>
> *in a reliable way, whether $\boldsymbol{ET_L}$ is an exact or statistical law.*

**All well tested laws of physics are effective laws** The microphysics $\mathbf{m}$ that I have used as a base level is not in fact the fundamental level of physical laws. It itself arises out of lower levels - and we do not know what the bottom-most level is. It might be String Theory/M Theory, but again it might not. No such Theory of Everything (TOE) is well defined, let alone experimentally verified. Consequently,

> **The Laws of Physics are Effective Laws** ***Despite the fact that they are effective, eternal, and unchanging, all the well established and tested laws on which physics and engineering are based are effective laws***[9] ***of the form (8). This applies equally to Newton's Laws of motion, Newton's Law of Gravitation, Maxwell's equations, Einstein's gravitational field equations, the Schrödinger equation, Dirac's equation, and the Standard Model of Particle Physics. They all only hold in a restricted domain $\mathcal{D}$.***

This is discussed further in Section 6.4. What this also implies is that a key aspect of determining any Effective Theory is to determine its Domain of Applicability $\mathcal{D}$.

## 4.4 Downward Causation and Causal Completeness

Equation (4) is where the downward effect of the macro state on the micro state is explicitly represented, through the modification $\mathbf{m} \to \mathbf{m'}(\mathbf{a})$. Thus Downward Causation occurs. The point is that $\mathbf{a}=\mathbf{a(M)}$ is a variable that cannot even in principle be represented in terms of the microlevel variables occurring in $\mathbf{m}$. The reason is that

> ***All the variables occurring in microphysics m respect the symmetry S, whereas a(M) - and so m' - does not***

And as has been shown above, we must use $\mathbf{m'}$ as the micro level theory if we want to get the correct macro level results. The theory $\mathbf{m}$ is unable to do the job. The key effect underlying this physically in the case of solid state physics is the quantum theory wave(macro)-particle(micro) duality discussed in Section 3.3, which leads to crucial lower level effective variables (quasiparticles) existing, without which the theory would not work.

> **Effective Dynamics and Downward Causation** ***In order to derive the correct macrodynamics M, the microdynamics m and fundamental***

[9] They are effective theories in the sense of [Castellani 2002].

***laws L must be replaced by effective microdynamics m' and effective laws L' respectively. The maps $m \to m'$, $L \to L'$ represent the effects of context on the functioning of physics at the micro level. That is, they represent downward causation.***

**Causal completeness?** It is claimed by many that such a downward causal influence is not possible because of the causal completeness of physics at the microlevel together with supervenience of the macro level **M** on the micro level **m**, see for example [Gibb *et al* 2019] which gives links to Kim who originated the argument. There has to be something wrong with this claim, because of the remark just made: the allegedly causally complete dynamics **m** must be replaced by **m'** if you want your Effective Theory to accord with experiment. And **m'** can only be obtained by introducing variables not present in **m**. The fact is that **m** cannot do the job (see (6)), so it must be causally incomplete.

Robert Bishop refutes the causal completeness claim in the context of fluid convection [Bishop 2008]. In [Ellis 2019], I give a refutation of the argument from causal completeness based firstly in the issue of multiple realisability of a higher level state by lower level states (see Section 4.7), and secondly in the difference between synchronic and diachronic supervenience. Here, on the basis of the above results, I will reverse the argument.

**Causal completeness of the microdynamics m.** ***The microdynamics m arising purely out of the Laws of Nature L is incomplete in this sense: it cannot derive in a bottom-up way properties occurring in condensed matter physics, as just demonstrated. Causally complete dynamics can however be attained by including downward effects in m that result in an effective lower level theory m' that is able to give the correct emergent results.***

This theme of interlevel causal completeness is developed in depth in [Ellis 2020]. Here, I confirm this claim by investigating strong emergence **P** of properties in condensed matter physics, using the definitions given in Section 2.

## 4.5 Downward causation and Effective Laws

Consider properties of a crystal, such as its optical properties and electrical conductivity. An effective potential term at the lower level is provided by the symmetry-breaking lattice structure that is relatively immobile in comparison with electron motion, so it may be at first approximation be regarded as fixed (this is the Born-Oppenheimer approximation that breaks translational invariance). The resulting potential $V(\mathbf{q}, \mathbf{x})$ shapes the lower level dynamics. The symmetry breaking occurs via the fact that $\partial V(\mathbf{q}, \mathbf{x})\partial \mathbf{x} \neq 0$. In this case, equation (7) is modified to give the effective laws **L'** as follows: $\mathbf{L} \to \mathbf{L'}(V)$ such that

$$\mathbf{L'}(V) : d \in \mathcal{D} \to \mathbf{L'}(V)[d] = o \in \mathcal{D},\ V = V(\mathbf{q}, \mathbf{x}),\ \partial V(\mathbf{q}, \mathbf{x})/\partial \mathbf{x} \neq 0. \tag{9}$$

Now comes the essential point: because the inhomogeneity leads to lattice vibrations, the potential will be time dependent:[10]

$$\{\partial V/\partial x \neq 0 \Rightarrow \partial V/\partial t \neq 0\} \Rightarrow \mathbf{L'}(V) : d \in \mathcal{D} \to \mathbf{L'}(V(\mathbf{q},\mathbf{x}, t))[d] = o \in \mathcal{D}. \tag{10}$$

The lower level physics is not causally closed: outcomes depend on the higher level lattice vibrations, as indicated by the time dependence of the potential $V(\mathbf{q},x,t)$. They are

[10] "*The broken symmetry state is not an eigenstate of the system Hamiltonian H (it breaks one of the symmetries of H) and so it is not a stationary state.*" [Blundell 2019].

equivalent to the existence of quasi-particles such as phonons which shape causal outcomes at the lower level (Section 3.3).

> **Downward causation (solid state physics)**: *Symmetry breaking at the higher lead to a time dependent effective potential governing lower level dynamics. Equation (7) is replaced by (10) and outcomes are determined by the time dependence of the potential.*

This supports the analysis above and in [Ellis 2019] arguing for strong emergence when diachronic supervenience take place. However much more than that, (10) can be extended to represent downward action of higher level variables controlling lower level causal effects, as I now illustrate in the specific case of a transistor in a digital computer. Interlevel causal closure then involves many higher levels [Ellis 2020].

## 4.6 Time dependent potentials: Transistors in computers

A specific context where time dependent potentials control lower level dynamics is the case of transistors in a digital computer, where algorithms encoded in a high level language chain down via compilers to generate sequences of bits that control electron flows in transistors [Tanenbaum 2006]. This is discussed in depth in [Ellis and Drossel 2019]. I will reproduce here the key part of that argument, showing how time dependent potentials generated in that way control the electron dynamics.

**The Hamiltonian** description at the electron/ion level is ([**?**]:16):

$$\begin{aligned} H &= -\sum_i \frac{\hbar^2}{2M_i}\nabla^2_{\mathbf{R}_i} - \sum_i \frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_i} + \sum_i \sum_{j>i} \frac{Z_i Z_j e^2}{4\pi\epsilon_0 \left|\mathbf{R}_i - \mathbf{R}_j\right|} \\ &\quad - \sum_i \sum_j \frac{Z_i e^2}{4\pi\epsilon_0 \left|\mathbf{R}_i - \mathbf{r}_j\right|} + \sum_i \sum_{j>i} \frac{e^2}{4\pi\epsilon_0 \left|\mathbf{r}_i - \mathbf{r}_j\right|} \end{aligned} \tag{11}$$

To derive from this the effective Hamiltonian encoding the lattice structure and electron charge distribution in a transistor, we must make a series of approximations.

A: First, electrons are characterised as either conduction band electrons (essentially unbound and so free to move) or valence band electrons (closely bound to ions and so localised). This represents the Hamiltonian (11) in the form

$$H = T_i + T_e + V_{ii} + V_{ee} + V_{ei} + E_{core} \tag{12}$$

Note that this is where knowledge of the higher (crystal) level dynamics is injected downwards into the representation of the lower level (ion/electron) dynamics (one can't make that splitting without that knowledge).

B: Now one uses the Born-Oppenheimer (adiabatic) approximation, thereby altering the Hamiltonian used. Thus this represents the transition (4) discussed in Section 4.3. This approximation assumes that the electrons are in equilibrium with the locations of the ions ([Schwabl (2007)],Ch.15). The wave function is factorized into two parts: the electron part $\Psi_e(\mathbf{r}, \mathbf{R})$ for given positions of the ions, and the ion part $\Phi(\mathbf{R})$,

$$\Psi(r, \mathbf{R}) = \Phi(\mathbf{R})\Psi_e(\mathbf{r}, \mathbf{R})\,. \tag{13}$$

This results in an electron equation

$$(T_e + V_{ee} + V_{ei})\Psi_e(\mathbf{r}, \mathbf{R}) = E_e(\mathbf{R})\Psi_e(\mathbf{r},\mathbf{R}) \tag{14}$$

and an ion equation

$$(T_i + V_{ii} + E_{core} + E_e(\mathbf{R}))\Phi(\mathbf{R}) = E\Phi(\mathbf{R})\,. \tag{15}$$

C: Using the electron equation (14), one obtains the electronic band structure. This is based on a picture of non-interacting electrons, so the interaction term $V_{ee}$ must be dropped and the electron Hamiltonian becomes a sum of one-particle Hamiltonians. The symmetry breaking contextual term determining electron outcomes is $V_{ei}$. In the context of a specific periodic lattice, the solutions are Bloch waves. These give the energy bands $E_n(\mathbf{k})$, which are the lattice analog of free particle motion ([Phillips 2012]:22).

D: Electron-lattice interactions occur via phonons ([Simon 2013]:82-84, 90-95). Using the ion equation (15) and further approximations (ions are localized, harmonic expansion of energy around equilibrium ion positions), one can derive the phonon modes of the ions. The symmetry breaking term here is $V_{ii}$. The dispersion relations of phonons are determined by the normal modes of the resulting harmonic model.

E: In order to explicitly model electron-phonon interactions, a quantum field theoretical formalism is needed based on creation and annihilation operators for electrons and phonons ([Solyom (2009)], Ch.23). The dispersion relations of electrons and phonons obtained in this way determine the electron and phonon propagators, however evaluating the interaction term needs a separate determination of the cross section for scattering of electrons via emission or absorption of a phonon.

F: Doping with impurities adds electrons to the conduction band ('donor' n-doping) or holes to the valence band ('receptor' p-doping) ([Simon 2013]:187-194). The junctions between n-doped and p-doped regions are the key feature of transistors ([Mellisinos 1990]:14-20; [Simon 2013]:199-203). Electron diffusion at such junctions leads to depletion regions, stabilized by induced electric fields ([Mellisinos 1990]: Fig.1.9). Transistors are created by suitably shaped such junctions ([Simon 2013]:203-205).

**Electric voltage effects** In order to model the steady state situation when the transistor is in either a conducting state or not, depending on whether the applied voltage is above a threshold or not, electric field effects must be added. Thus the gate voltage $V(t)$ must be added to the model. This leads to an added potential energy term in the Hamiltonian of the electrons:

$$H_V(t) = \sum_i eV(\mathbf{r}_i, t) \tag{16}$$

where the higher level variable $V(t)$ determines the lower level variables $V(\mathbf{r}_i, t)$ in a downward way. The electrons reach an equilibrium where the electrical field created by the modified charge distribution cancels the electrical field due to the gate voltage. To calculate this equilibrium and so determine $H_V(t)$, a self-consistent calculation is needed, based on the charge density due to doping, the gate potential, and the thermal excitation.

This alters the band structure according to the bias voltage applied, and thereby either creates a conduction channel by changing the depletion zone, so current flows, or not, so no current flows ([Mellisinos 1990]:19), ([Simon 2013]:Fig.18.6).

> **In summary**: *This explicit example shows* ***(a)*** *how the basic Hamiltonian (11) representing the microdynamics* $\mathbf{m}$ *gets altered, in real world applications, to give the effective dynamics* $\mathbf{m'}$ *as in (4), affected by the higher level state;* ***(b)*** *the time dependent potential term (16) determines lower level dynamical outcomes to give an extension of (10) where now the time dependence of* $H_V(t)$ *is determined by the logic of algorithms;* ***(c)*** *this is a clear example*

*of downward causation enabled by the modular hierarchical structure of the computer [Tanenbaum 2006]. The word 'causation' is justified in both counterfactual and experimental terms: a different algorithm results in different electron flows.* ***(d)*** *It is a convincing case of strong emergence in Chalmer's sense [Chalmers 2000] because the physics summarised in (11) cannot in principle by itself lead to deduction of the kind of logical branching characterising the functioning of the emergent computational dynamics.*[11]

Algorithms produce real world outcomes by this chain of downward causation.

## 4.7 Multiple Realisability

A key point is that multiple realisability plays a fundamental role in strong emergence [Menzies 2003]. Any particular higher level state can be realised in a multiplicity of ways in terms of lower level states. In engineering or biological cases, a high level need determines the high level function and thus a high level structure that fulfills it. This higher structure is realised by suitable lower level structures, but there are billions of ways this can happen. It does not matter which of the equivalence class of lower level realisations is used to fulfill the higher level need, as long as it is indeed fulfilled. Consequently you cannot even express the dynamics driving what is happening in a sensible way at a lower level.

Consider for example the statements *The piston is moving because hot gas on one side is driving it* and *A mouse is growing because the cells that make up its body are dividing.* They cannot sensibly be described at any lower level not just because of the billions of lower level particles involved in each case, but because *there are so many billions of different ways this could happen at the lower level*, this cannot be expressed sensibly at the proton and electron level. The point is the huge different numbers of ways combinations of lower level entities can represent a single higher level variable. Any one of the entire equivalence class at the lower level will do. Thus it is not the individual variables at the lower level that are the key to what is going on: it is the equivalence class to which they belong. But that whole equivalence class can be describer by a single variable at the macro level, so that is the real effective variable in the dynamics that is going on. This is a kind of interlevel duality:

$$\{v_{\mathbf{L}} \in \mathbf{L}\} \Leftrightarrow \{v_{\mathbf{i}} : v_{\mathbf{i}} \in E_{\mathbf{L\text{-}1}}(v_{\mathbf{L\text{-}1}}) \in \mathbf{(L\text{-}1)}\} \tag{17}$$

where $E_{\mathbf{L\text{-}1}}(v_{\mathbf{L\text{-}1}})$ is the equivalence class of variables $v_{\mathbf{L\text{-}1}}$ at Level **L-1** corresponding to the one variable $v_{\mathbf{L}}$ at Level **L.** The effective law $\mathbf{EF_L}$ at Level **L** for the (possibly vectorial or matrix) variables $v_{\mathbf{L}}$ at that level is equivalent to a law for an entire equivalence class $E_{\mathbf{L\text{-}1}}(v_{\mathbf{L\text{-}1}})$ of variables at Level **L-1**. It does not translate into an Effective Law for natural variables $v_{\mathbf{L\text{-}1}}$ *per se* at Level **L-1**. The importance of multiple realisability is discussed in [Ellis 2019] and [Bishop and Ellis 2020]. The latter paper also discusses the underlying group theory relations, dealing in particular with the emergence of molecular structure.

**Essentially higher level variables and dynamics** *The higher level concepts are indispensible when multiple realisability occurs, firstly because they define the space of data* $d_{\boldsymbol{L}}$ *relevant at Level* ***L****, and secondly because of (17), variables in this space cannot be represented as natural kinds at the lower level. Effective Laws* $\boldsymbol{EF_L}$ *at level* ***L*** *can only be expressed at level* ***L-1*** *in terms of an entire equivalence class at that level. One can only define that equivalence class by using concepts defined at level* ***L****.*

[11] This is in parallel to the biological case, see Section 5.4 and equation (21) below.

To inject reality into this fact, remember that the equivalence class at the lower level is typically characterised by Avagadro's number.

Related to this, one should note that the dynamic or even static properties **P** of a macro system are the result of much faster dynamics at the microlevel. Electrons and molecules are always in motion, even ions in a crystal lattice are vibrating. Properties such as strength and ductility and brittleness are due to atoms moving past each other (as they are in metals), and this depends on dislocations, which are dynamic phenomena at the micro level, by which metallurgists mean microns rather than nanometres.

# 5 Emergence and Life

The argument for strong emergence is quite different in the cases of biological and physical systems. In this section I first comment on purpose and life (Section 5.1). Then I argue that the processes of Natural Selection are strongly emergent (Section 5.2). I use an argument based in the existence of time dependent constrains to show that molecular biology dynamics is strongly emergent (Section 5.3). In Section 5.4, I comment on whether there could be a dictionary from physics to biology. Putting this together, I argue that strong emergence certainly occurs in biology (Section 5.5).

## 5.1 Purpose and Life

Do biological organisms have purpose? It is strongly argued by Nobel Prize winning biologist Leland Hartwell and colleagues that this is indeed the case, in [Hartwell *et al* 1999]:

> *"Although living systems obey the laws of physics and chemistry, the notion of function or purpose differentiates biology from other natural sciences. Organisms exist to reproduce, whereas, outside religious belief, rocks and stars have no purpose. Selection for function has produced the living cell, with a unique set of properties that distinguish it from inanimate systems of interacting molecules. Cells exist far from thermal equilibrium by harvesting energy from their environment. They are composed of thousands of different types of molecule. They contain information for their survival and reproduction, in the form of their DNA. Their interactions with the environment depend in a byzantine fashion on this information, and the information and the machinery that interprets it are replicated by reproducing the cell."*

Even proteins have functions [Petsko and Ringe 2009]. One can claim that the interaction between structure and function whereby such function and purpose is realised is central to biology [Campbell and Reece 2008].

A central point as regards the nature of emergence is that words that correctly describe causation at higher levels simply do not apply at lower levels. It is not a point of confusion, it's a central aspect of emergence. As stated by [Anderson 1972], "*each level can require a whole new conceptual structure*". Just so; and that means new terminology and new variables $v_{\mathbf{L}}$ at higher levels **L**. To make this clear. I give a number of examples where this occurs in Section 5.2. The issue of whether there could be a dictionary deriving these variables from the underlying physics is considered in Section 5.4. In the case of biology, unless the higher levels include the concept of function, they will miss the essence of what is going on, as pointed out by [Hartwell *et al* 1999]. This leads to the key role of information flows [Davies 2019] and logic [Nurse 2008] [Ellis and Kopel 2019] in the processes of life. This crucially distinguishes biology from physics.

But how does this all emerge from physics? Through symmetry breaking, as in the physics case discussed above, together with time dependent constraints and the extraordinary dependence of molecular biology on the conformational shape of biomolecules [Ellis and Kopel 2019]. Wonderful examples are given in [Karplus 2014]. This all came into being via Darwinian processes of natural selection [Mayr 2001]: one of the best-established processes in biology, which is strongly emergent.

## 5.2 Why Natural Selection is Strongly Emergent

Developmental processes leading to the existence of proteins, and hence cells and bodies, take place by reading the genome via processes controlled by gene regulatory networks. This is emergence **E** of macro structure from microstructure. But it is based in the current existence of the genome and developmental systems. The genome has famously been shaped by processes of natural selection [Mayr 2001] taking place over evolutionary timescales. The key element is

> ***Evolutionary Emergence**: Genotypic mutations that provide some phenotypic advantages given some environmental conditions enable the species at stake to survive (relative to other species).*

When this occurs, the outcomes cannot even in principle be deduced from the underlying microphysics, because both of the random aspect of the processes of natural selection, and their top-down nature [Campbell 1974]. Apart from random drift, which is a significant effect in evolutionary history, life today is determined by adaptation to a variety of physical, ecological, and social contexts in the past. There was randomness in the process due to mutation, recombination, and drift at the genetic level, which was influenced by molecular randomness as well as by quantum events that cannot be predicted in principle (emission of cosmic rays that caused genetic mutations) so evolution from initial data as determined purely by the underlying physics cannot in principle determine unique outcomes (see the discussion in Section 6.1). Rather time-dependent downward influences took place [Campbell 1974] due to the time varying niches available for adaption, which determined what superior variation was selected. According to Ernst Mayr [Mayr 2001],

> *"Owing to the two step nature of natural selection, evolution is the result of both chance and necessity. There is indeed a great deal of randomness ('chance') in evolution, particularly in the production of genetic variation, but the second step of natural selection, whether selection or elimination, is an anti-chance process."* [12]

He then gives the eye as an example, which enhances relative survival probabilities because of the information processing and consequent contextually-dependent predictions of outcomes it allows. Neither step is predictable from physics, even in principle, because physics does not have the conceptual resources available to undertake the discussion. They are essentially biological in nature, and hence strongly emergent: the causal relations governing what happens (as described by Mayr) occur at emergent biological levels through interactions between biomolecules such as DNA, RNA, and proteins, determined by their specific molecular structures. Thus we have here **SB**(**NS**): the specific nature of the broken symmetries occurring in biomolecules today [Petsko and Ringe 2009], and so shaping current biological outcomes, has been determined by Darwinian processes of natural selection (**NS**) during our past evolutionary history leading to existence of these molecules.

[12] However note that the issue of randomness is complex and subtle; see for example [Wagner 2012].

To strengthen the point, I will now consider cases of strong emergence in biology due to such selection at the molecular level. Biological function at the molecular level is determined by the physical nature of biomolecules, with proteins being the workhorses of biology. Due to the nature of the underlying physical laws and the values of the constants occurring in the Laws of Physics (7), only certain proteins are possible in physics terms. Thus there is an abstract possibility space for the existence of proteins, as discussed by Andreas Wagner [Wagner 2014]. It is of enormous dimension, as he emphasizes. The proteins that actually exist on Earth, out of all the possible ones that might have existed but have not been selected for, have a very complex structure that is determined by the functions they perform [Petsko and Ringe 2009] (note the use of the word "function"). They were selected by Darwinian processes [Mayr 2001] so as to have a structure that will adequately perform a needed function.

An example is the Arctic Cod. Wagner discusses selection for proteins in Chapter 4 of his book ([Wagner 2014]:107). The Arctic cod lives within 6 degrees of the North Pole in sea water whose temperature regularly drops below 0 degrees Celsius. This ought to destroy the cells in the fish, as ice crystals should form within the cells and tear them apart. To deal with this problem, the arctic cod survives by producing antifreeze proteins that lower the temperature of its body fluids. What has been selected for is not just the genome that codes for the protein that will do the job, but the entire developmental system that produces the protein, involving gene regulatory cascades ([Wagner 2014]:Chapter 5). This is a case of downward causation from the environmental level (the cold nature of the sea water) to details of the DNA sequence that does the job [Ellis 2016]. Another example is the precisely patterned distribution of a particular family of proteins in squid eyes that leads to a graded refractive index in its lenses [Chang 2017]. A parabolic relationship between lens radius and refractive index allows spherical lenses to avoid spherical aberration [Sweeney *et al* 2007]. This physiological advantage results in evolutionary development of specific proteins that are then precisely positioned to attain this effect. Thus the need for clear vision in an underwater environment lead to selection of a genotype that produces this outcome via squid developmental systems.

The point of these examples is that causation is real at these biological levels. Tee underlying physics enables this all to happen, but does not determine the specific outcomes: it cannot do so as a matter of principle. The emergent relations are determined by biological needs, which is why it is essential to use language appropriate to the higher level in order to get a model that can provide a reductionist understanding of mechanisms.

Similar arguments apply at the cellular level, for example in determining the way neurons have specific voltage gated ion channels imbedded in axon membranes that underlie the Hodgkin-Huxley equations for action potential propagation [Hodgkin and Huxley 1952]. As discussed by [Scott 1999], the constants in those equations cannot in principle be determined by the underlying physics as they are determined by their physiological functioning. If they were determined in a bottom up way, those constants would be determined by the fundamental physics constants. This proves this is strong, not weak, emergence.

**The selection relation** When natural selection occurs, there is a phenotype level **LP** where phenotypic advantage leads to enhanced relative survival rates after genotype mutations occur. The genotype level **LG** upwardly produces outcomes at the phenotype level **LP** through developmental processes, in a contextually dependent way. Selective choice of advantageous phenotypes $v_{\mathbf{LP}}$ at the higher level **LP** (or at least ones that are not seriously disadvantageous) chains down to select preferred genotypes $v_{\mathbf{LG}}$ at the lower level **LG**. The effective theory $\mathbf{ET}_{\mathbf{LG}}$ at the genotypic level (reading of genes to produce proteins

via developmental systems) is altered because the set of variables $\{v_{\mathbf{LG}}\}$ at level **LG** has been changed by the selection process $\mathbf{S_{LG}}(v_{\mathbf{LP}})$ to give a new set of variables $\{v'_{\mathbf{LG}}\}$ at that level that will produce the preferred outcomes $\{v_{\mathbf{LP}}\}$ at the phenotype level:

$$\mathbf{S_{LG}}(v_{\mathbf{LP}}) : \{v_{\mathbf{LG}}\} \in \mathbf{LG} \to \{v'_{\mathbf{LG}}\} = \mathbf{S_{LG}}(v_{\mathbf{LP}})[v_{\mathbf{LG}}] \in \mathbf{LG} \tag{18}$$

This is thus a case of *Downward Causation by Adaptive Selection*, deleting lower level elements. This enables alteration of structures and functions at level **LG** so as to meet new challenges at level **LI**, in an interlevel closed causal loop. This is of course a highly simplified view of a complex process: in fact multiple phenotype levels are involved, and neutral evolution can occur. The fundamental point is unchanged.

## 5.3 Emergence of Properties P in Biology

Downwards effects in a biological system occur because of physiological processes, see *The Music of Life* by Denis Noble, and [Noble 2008], [Noble 2012]. These processes are mediated at the molecular level by developmental systems [Oyama et al 2001] operating through gene regulator networks [Wagner 2014] and cell signalling networks [Berridge 2014], guided by higher level physiological needs. They reach down to the underlying physical levels via time dependent constraints on the lower level dynamics [Ellis and Kopel 2019], which is why the lower level physics *per se* is not causally complete (cf. 4.4). The *set of interactions* between elements at that level is uniquely characterised by the laws of physics **L**, but their *outcomes* are determined by the biological context in which they operate, leading to effective laws **L'**. Equation (7) should, in parallel with (9), be modified to read

$$\mathbf{L'} : d \in \mathcal{D} \to \mathbf{L'}(C)[d] = o \in \mathcal{D},\ C = C(\mathbf{a}, \mathbf{p}, \mathbf{q}),\ \partial C(\mathbf{a}, \mathbf{p}, \mathbf{q})/\partial \mathbf{a} \neq 0 \tag{19}$$

where $C = C(\mathbf{a}, \mathbf{p}, \mathbf{q})$ are constraints on microlevel coordinates $\mathbf{p}$ and momenta $\mathbf{q}$ that are dependent on biological variable $\mathbf{a}$ representing conditions at a higher level. Now comes the essential point (cf.(10)): because $\mathbf{a}$ is a biological variable, it will be time dependent:

$$\mathbf{L'} : d \in \mathcal{D} \to \mathbf{L'}(C(\mathbf{a}(t), \mathbf{p}, \mathbf{q}))[d] = o \in \mathcal{D},\ \partial \mathbf{a}/\partial t \neq 0 \Rightarrow \partial C/\partial t \neq 0. \tag{20}$$

The lower level physics is not causally closed: outcomes depend on the time variation of the constraint C(t,**p**,**q**).[13] This accords with Juarrero's characterisation of constraints as causes [Juarrero 2002].

Examples are the voltage across a voltage gated ion channel [Ellis and Kopel 2019], and the presence or not of a ligand bound to a ligand gated ion channel [Berridge 2014], which both alter molecular conformations and so change constraints on ion flows. To model this in detail requires quantum chemistry methods well adapted to this biological context, such as those used in [Karplus 2014]. Putting this in a real biological context such as determination of heart rate , pacemaker activity of the heart is via cells in the sinoatrial node that create an action potential and so alter ion channel outcomes. This pacemaking circuit is an integrative characteristic of the system as a whole [Fink and Noble 2008] - that is, it is an essentially higher level variable - that acts down to this molecular level.

In this way downward causation takes place in biology and enables strong diachronic emergence of properties **P** in biology:

> **Strong Emergence of Properties (biology):** ***Emergence of Properties P in biology is strong emergence because time dependent constraints alter lower level dynamical outcomes in accord with higher biological needs. Equation (7) is replaced by (20).***

[13] A simple analogous model of how this effect works is a pendulum with time dependent length, see the Appendix of [Ellis and Kopel 2019].

## 5.4 A Dictionary from Physics to Biology?

Along with Ernst Mayr [Mayr 2001], it is my view that physics does not begin to have the resources to deal with this, *inter alia* because the concepts of being alive or dead cannot be captured in physics terms, where only interactions between particles are characterised, so 'survival'- central to evolutionary theory - is not a concept physics can deal with. Whether an animal is alive of dead is a biological issue, related to interlevel causal closure including biological variables [Ellis 2020]. Life has crucial characteristics of its own [Campbell and Reece 2008] that are orthogonal to concepts encompassed by physics.

Nevertheless I have been challenged by a referee as followings: might there be a dictionary relating physical variables to biological variables, thereby undermining my claim?

While there are indeed dictionaries relating aspects of physics to key aspects of life, for example relating the physics of thermodynamics to use of energy in biology, and relating physics scaling laws to biological scaling laws, such dictionaries fail to capture the aspect of purpose or function focused on by Hartwell et al [Hartwell *et al* 1999] and the use of information emphasized by [Nurse 2008] and [Davies 2019]. Regarding energy use, life is famously an open thermodynamic system [Peacock 1989] [Hoffmann 2012] where metabolism is taking place, but this feature *per se* does not characterise life, because even a candle flame is open in that sense, and it is certainly not alive. Indeed it is notoriously hard to state just what such conditions are. Various characterisations of life have been given, such as

- Life involves Homeostasis, Organization, Metabolism, Growth, Adaptation, Response to Stimuli, and Reproduction with variation
- "Life is a self-sustaining chemical system capable of Darwinian evolution."
- "Living things are self-organizing and autopoietic (self-producing)"
- Life is characterised by causal closure [Mossio 2013],
- Life is characterised by function and purpose [Hartwell *et al* 1999]
- Life is determined by information use [Nurse 2008] [Davies 2019] and associated information processing via logical operations [Hoffmann 2012] [Ellis and Kopel 2019]

and so on. None of these characterisations resemble any aspect of physics as commonly understood, either at the fundamental level of particles, forces and fields, or at emergent physics levels, except perhaps in regard to energy use and minimisation. I will pursue just two aspects here: life as characterised in energy terms, where one might claim there are such dictionaries, and life as characterised in information processing terms, where it is implausible there is such a dictionary.

**Life as characterised in energy terms?** Life certainly depends on energy, so some physicists have attempted to characterise life as being essentially described by energy or thermodynamic relations. I will comment on three different aspects here.

The most sustained effort is that by Karl Friston [Friston 2012], developed into a major theory ("The Free Energy Principle") *inter alia* encompassing the brain [Friston 2010]. However this is not derived in a bottom up way from the underlying physics: rather it is an effective higher level theory constructed in analogy with an important aspect of physics. It is about free energy in the information theory sense, not literally in the thermodynamic sense, thus it is in effect an analogue of how energy functions in the purely physical domain. It is unrelated to the emergence of biology from the underlying lower level phyics.

Second, one can suggest [Luu and Meißner 2019] that the "purpose" of some biological organism, seen from our limited point of view, is procreation and the continuation of its species, but at the same time is equivalent to the minimization of energy in some very complex phase space. Indeed so. This is the concept of a Fitness Landscape as proposed by Sewell Right [Wright 1932l], and developed by many others (see for example [McGhee 2006]). But this does nothing to further the reductionist project. It is an *Effective Theory* in the sense explained in [Castellani 2002] (quoted in §2.1), but has nothing to do with weak emergence for example via an Effective Field Theory relating microphysics to emergent phenomena by a power series expansions in terms of a physical parameter. A Fitness Landscape cannot be derived even in principle from the underlying physics by any deductive process that purely involves physics concepts, precisely because it depends on the concept 'fitness', and it does not map from the level of protons and electrons to the world of biological reality. It can however be developed as a productive effective theory if one introduces relevant contextually dependent higher level biological concepts.

Third, the way energy in the physical sense relates to biological activity is through the joint effects of the cardiovascular and digestive systems. These are macroscopic physiological systems, with extremely well established emergent laws of behaviour that underlie medical practice in these respective areas. They interact in an integrated with way with immensely complex metabolic networks at the cellular level, where logic of the form (21) determines what happens. An example is the Citric Acid Cycle (**Figure 1**). This is without doubt causally effective at the cellular level and is irreducible to any lower level because of the closure of the major loop in the interaction diagram (it does not follow from the interactions of protons and electrons *per se*). Notice also that this diagram establishes that a casual relation exists in Pearle's sense [Pearl 2009] [Pearl and Mackenzie 2018].

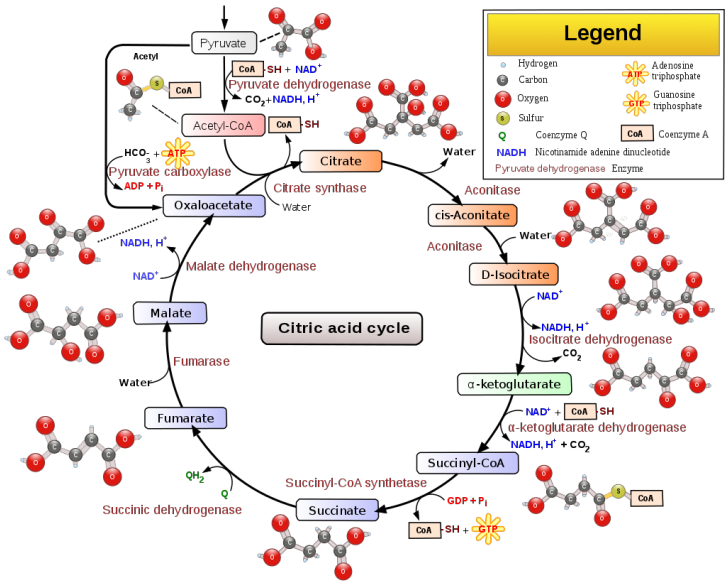


Figure 1: **Citric Acid Cycle** (Source: Wikipedia)

Metabolism at this level is driven by the dynamics at the emergent physiological system level [Noble 2012], which characterizes relations between macro variables such as cystolic blood pressure and heart rate [Fink and Noble 2008], and reaches down to control energy resources at the cellular and molecular levels [Noble 2012]. These are inherently higher

level variables, related to the detailed physiological structure and functioning of the heart. These effective higher level relations [Rhoades and Pflanzer 1989] are also experimentally verified to high precision. This is a far cry from a process of energy or entropy minimisation. It is a highly coordinated topologically closed set of far from equilibrium interactions. Indeed it is controlled by information and logic, which I now discuss.

**Life as characterised by information and logic** A contrasting view is that Information is key in biology [Nurse 2008] [Davies 2019], and the associated logic cannot emerge out of non-information because information conveys meaning only in a specific higher level context. At each level biology operates according to a branching logic of the following form ([Hoffmann 2012]:151-152) [Ellis and Kopel 2019]: a signal $X$ is evaluated by a Truth Function $T(X)$ in a context $\mathcal{C}$, as follows

$$\text{GIVEN context } \mathcal{C}, \text{ IF } T(X) \text{ THEN } O \text{ ELSE } O2. \tag{21}$$

where $O1$ and $O2$ are alternative dynamical paths. This is for example the way all cell signalling works [Berridge 2014] and so underlies metabolic networks and gene regulatory networks. This kind of logic is implemented by the conformational shapes of the relevant biomolecules and their interactions with each other, allowed by the underlying physics but not reducible to it because physical laws *per se* do not entail logical outcomes such as (21), and neither does specifying the detailed distribution of molecules (22). This dynamics only emerges in the context of interactions at the cell level [Hofmeyer 2017] [Hofmeyer 2018], allowing interactions of the form (21) to occur and be causally effective, reaching down to alter outcomes at Level **L1**. At the cellular level, the important networks are cell signalling networks [Berridge 2014], metabolic networks [Wagner 2014], and gene regulatory networks [Carroll 2008]. They each have their own logic, irreducible to lower levels precisely because they are networks characterised by their topology as well as interaction strengths and timescales. Because of multiple realisability, they can be realised in many billions of ways at the particle level.

Now one can perhaps claim (21) is somewhat like a phase transition (Section 4.3), associated with a change of symmetry. Indeed it is, but there are two key remarks here. First a phase transition is an example of downward causation associated with strong emergence, because a macro level change of conditions (temperature $T$ decreasing) causes the micro level change of symmetry ([Ellis 2016]:139) [Green and Batterman 2020]. Second, in the case of biology, (21) is associated with signalling [Berridge 2014], while in the case of phase changes in physics, that element is absent.

However one point remains: Quantum Computing contains logical branching such as is characterised by (21) when Qbits are manipulated. Why does this not disprove what I have just claimed, as Qbits are basic physical phenomena, in effect occurring in anytwo-state quantum system? Is that not an example of a potential dictionary from physics to biology? The response is that Qbits can only be read out to produce branching logic such as (21) in contexts that are extremely carefully engineered to make this outcome possible: in short, in the context of engineering systems. And such systems are strongly emergent [Ellis 2020]. The physics by itself cannot produce that kind of branching logic outside such a context. One can indeed produce dictionaries relating properties of such engineering systems to life and the brain, indeed that is what simulations are about. This ends up ultimately in the highly contested area of artificial intelligence. It is not a mapping between physics *per se*, as characterised by interactions between particles and fields, and biology; it is between engineering systems and biology, an entirely different proposition.

Thus it is my view that

> **Relating physics to life** ***There can be dictionaries relating physics concepts and interactions to some necessary conditions for life, but they cannot encompass sufficient conditions for life to exist. They are maps from some aspects of physics to some aspects of biology but they omit essential aspects of what it is to be alive, failing to capture the nature of biology as a whole, because they miss essential biological features. Equivalently, there are essential biological variables that cannot be captured by such a dictionary.***

From a physics viewpoint, this claim is justified by discussion in [Anderson and Stein 1987] regarding the contrast between equilibrium systems, dissipative systems, and life.

## 5.5 Strong Emergence in Biology

There are many reasons why biology is strongly emergent. I note first that biological emergence is based in supramolecular chemistry [Lehn 1993] [Lehn 1995] and the nature of biomolecules such as nucleic acids and proteins [Petsko and Ringe 2009]. Given this understanding, reasons biology is strongly emergent are as follows:

- Firstly, because organisms are made of biomolecules where the functional properties of those molecules **P** are strongly emergent due to their dynamic symmetry-breaking nature, see Section 5.3.
- Secondly, because of the functional emergence **P** of the properties of physiological process and organisms at any specific time out of the underlying biomolecules. These are the processes of physiology [Rhoades and Pflanzer 1989], which are dynamic processes based in essentially biological interactions and variables.
- Thirdly, because of the emergence **E** of each individual organism over time out of the component biomolecules through the processes of developmental biology [Carroll 2005] [Carroll 2008] controlled by irreducible hierarchical gene regulatory networks.
- Fourthly, because of the historical emergence **E** over evolutionary timescales of organisms, of the biomolecules out of which they are made [Wagner 2014], including the DNA that is our genetic material [Mayr 2001], and of the developmental processes whereby biomolecules give rise to organisms. This is an intricate intertwined process of evolution and development (hence *Evo-Devo* [Carroll 2008], [Oyama et al 2001]). Biology has an ineliminable historical element which results in its present day nature through environmental adaptation, a top-down process [Campbell 1974].

All these dimensions of biology are strongly emergent.

> **Conclusion**: ***Biology is not reducible to physics. It is an indubitable case of strong emergence of properties P because it involves function and purpose ([Hartwell et al 1999] associated with downward causation and information flows [Nurse 2008] [Davies 2019] implemented by time dependent constraints [Ellis and Kopel 2019], and because it involves strong emergence associated with molecular biology and with physiology. Its emergence E is also strong emergence because of the nature of the process of Natural Selection that led to life on Earth [Mayr 2001], which involves both randomness and downward causation [Campbell 1974].***

This emergence results in testable Effective Theories, such as the Hodgkin-Huxley equations for action potential propagation, Fitness Landscapes [Wright 1932l] [McGhee 2006], models of the heart [Fink and Noble 2008], and Darwinian evolution [Mayr 2001].

# 6 Conclusion

The discussion above has highlighted that while one can indeed obtain Effective Theories of almost any topic studied by physics — indeed that is what physics is about — one can almost always not make the link between the micro and emergent theories without explicitly introducing variables and concepts that were not implied by the underlying theory. A purely bottom-up attempt of derivation of the properties of emergent states such as semi-conductors cannot be done: these properties are strongly emergent. Downward causation [Ellis 2016] plays a key role in strong emergence in condensed matter physics and soft matter physics (Section 4), even if it is not explicitly identified as such.

Biology is strongly emergent for multiple reasons (Section 5.5). The Darwinian process of natural selection that leads to the existence of particular living forms, is strongly environmentally dependent, and hence unpredictable in a bottom up way. The processes of molecular biology are shaped in a downward way by signalling molecules that change protein conformation and hence alter constraints on ion and electron motion in a time dependent way, related to biological purpose and function.

I now comment on whether micro data dependence undermines strong emergence (Section 6.1), the nature of essentially higher level variables (Section 6.2), whether effective field theories disprove strong emergence (Section 6.3), and the equal validity of all emergent levels (Section 6.4). Section 6.5 summarise results obtained in this paper.

## 6.1 Does micro data dependence undermine strong emergence?

It has been strongly claimed to me that micro data dependence of all outcomes undermines my argument for strong emergence. The argument goes as follows. Suppose I am given the initial positions $\mathbf{r}_i$ and momenta $\mathbf{p}_i$ of all particles in the set $\mathcal{P}$

$$\mathcal{P} := (\text{protons, neutron, electrons}) \tag{22}$$

at a foundational level **L1**, which constitute an emergent structure **S** at a higher level **L2**. The details of **S** are determined by that microdata, even though its nature cannot be recognised or described at level **L1**. The forces between the particles at level **L1** completely determine the dynamics at that level. Hence the emergent outcomes at level **L2** are fully determined by the data at Level **L1**, so the emergence of dynamical properties and outcomes at level **L2** must be weak emergence and be predictable, at least in principle from the state (22) of level **L1**. This would apply equally to physical, engineering, and biological emergent systems. It is in effect a restatement of the argument from supervenience.

There are problems with the argument just stated as regards microphysics, macrophysics, biology, and causal effectiveness of higher levels.

**Microphysics** There is irreducible uncertainty of quantum outcomes when wave function collapse to an eigenstate takes place, with outcomes only predicted statistically via the Born rule [Ghirardi 2007]. This unpredictability has consequences that can get amplified to macrolevels, for example in terms of causing errors in computer memories due to cosmic rays. Cosmic rays also alter genes significantly enough to cause cancer and to

alter evolutionary history ([Ellis 2020] §7.2). Quantum uncertainty also affects specific structure formation outcomes in inflationary cosmology ([Ellis 2016]:p.398).

**Macrophysics** Consider a higher physical level **L3** in the context of a fluid where convection patterns take place. Because of the associated chaotic dynamics together with the impossibility of setting initial data to infinite precision [Ellis *et al* 2018], macroscopic outcomes are unpredictable in principle from micro data (22) [Del Santo and Gisin 2019]. [Anderson 2001] puts it this way:

> "*A fluid dynamicist when studying the chaotic outcome of convection in a Benard cell knows to a gnat's eyelash the equations of motion of this fluid but also knows, through the operation of those equations of motion, that the details of the outcome are fundamentally unpredictable, although he hopes to get to understand the gross behaviour. This aspect is an example of a very general reality: the existence of universal law does not, in general, produce deterministic, cause-and-effect behaviour*"

This fundamentally undermines the concept of a causally closed physical levels in the case of classical physics. This affects local fluid convection patterns as well as weather patterns, from where it chains down to affect macro biology, microbiology, and thus the underlying physics. Because of global warming, this is causing serious problems for farming.[14]

Furthermore chaotic motion is unpredictable because of gravitational effects. Michael Berry has pointed out that after about 120 collisions a billiard ball's future behavior would be effected by the gravitational pull of one electron at the edge of the universe.[15]

**Biology: Harnessing stochasticity** Consider the case of bio molecules in a cell where a molecular storm occurs [Hoffmann 2012]. In talking about the molecular machines that power cells, he states that every such molecular machine in our bodies is hit by a fast-moving water molecule about every $10^{-13}$ seconds. "*At the nanoscale, not only is the molecular storm an overwhelming force, but it is also completely random.*" To extract order from this chaos, "*one must make a machine that could 'harvest' favorable pushes from the random hits it receives.*" That is how biology works at this level. By selection for biological purposes from an ensemble of possibilities presented by this randomness, higher level needs may be selected for in a downward way, as stated by [Noble and Noble 2018]:

> "*Organisms can harness stochasticity through which they can generate many possible solutions to environmental challenges. They must then employ a comparator to find the solution that fits the challenge. What therefore is unpredictable in prospect can become comprehensible in retrospect. Harnessing stochastic and/or chaotic processes is essential to the ability of organisms to have agency and to make choices.*

This is the opposite of the Laplacian dream of ordered initial data at the micro level leading to outcomes uniquely predicted by that micro data. It is the detailed structure of molecular machines, together with the lock and key molecular recognition mechanism used in molecular signalling [Berridge 2014], that enables the logic of biological processes to emerge as effective theories governing dynamics at the molecular level and thereby reaching down to control the physical level **L1** via time dependent constraints [Ellis and Kopel 2019].

[14] See Climate change and farming: 'Unpredictability is here to stay'.
[15] I thank Martin Rees for that comment.

In this way downward causation by irreducible higher level variables takes place, and alters lower level dynamics so that the initial data does not determine outcomes, it is higher level needs that do so. But the issue then is, what is the nature of such variables? Do they indeed exist? I look at this in Section 6.2.

**Causal effectiveness of higher levels** Finally, there is what is in effect another version of this argument, this time in the backwards direction of time. The challenge is to explain how life as it exists nowadays arose out of the initial data in the early universe (to be simple, let's say data on the last scattering surface where matter and radiation decoupled from each other). It could not have arisen in a deterministic way from that initial data because of all the uncertainties and randomness just discussed. The obvious explanation is that it arose through the combination of evolutionary and developmental process ('Evo-Devo' [Carroll 2005] [Carroll 2008]) which are essentially biological processes which have real causal power at their emergent level, enabled but not determined by the underlying physics (see Section 5.2).

Martin Rees however says in relation to embryo development[16],

> "*It could be the case that if we 'run the film backwards' through successive hyper-surfaces back to the big bang, every event or state is emergent from its predecessors via a combination of micro events (like a breaking wave) even though we interpret it in terms of all the higher levels. ... you could trace back across 'hypersurfaces' to learn how viewed on an atomic scale, the embryo had evolved into the adult*"

As phrased, the underlying view is that both the evolutionary and developmental processes are epiphenomena floating on the surface of the real causal element, namely the atomic level physics operating on data (22). The underlying viewpoint is that all causation proceeds upwards in the hierarchy of complexity.

Of course you could trace the micro events in this way. But this would simply fail to tell you what was driving these developments: what was the key causal factor shaping these outcomes? You can do that micro level tracing whether the higher levels have zero causal powers, as is suggested by this phrasing, or they in fact have real causal powers because every level has causal powers ([Noble 2012] and Section 6.4). One either has to claim that genes and gene regulatory networks have real causal powers [Watson *et al* 2013] [Alberts 2007], as has apparently been demonstrated by many thousands of experiments, or that all these experimentally determined outcomes are just correlations not indicating causation because no causal powers reside at either the molecular or cellular level.

Tracing back what interactions took place at the atomic level as suggested by Rees has no ability to distinguish between these two fundamentally different possibilities. It simply records what happened but not why it happened. My position, following [Campbell 1974] and [Noble 2008] is that the higher levels have real causal powers based on essentially biological processes [Hartwell *et al* 1999] [Campbell and Reece 2008] [Nurse 2008] enabled by downward causation, for which their is ample experimental and counterfactual evidence (vary higher level variables and observe or argue for altered lower level outcomes).

## 6.2 Essentially higher level variables

Irreducible higher level variables cannot be obtained by coarse graining or in any other way upscaling lower level variables, that is, they are essentially higher level variables defined at a level $\mathbf{L} > \mathbf{L1}$. There are many examples

[16] Private communication

- **Structures** [Green and Batterman 2017] [Batterman 2018] give examples involving heterogeneous physical structures and biology. [Davidson *et al* 2009] give an example from developmental biology, the notochord:

  > "*The capacity of the notochord to resist bending as it extends the embryo comes from the structure of the whole notochord. Measurements at the level of the individual collagen fiber or fluid-filled cells that make up the structure would not reveal the mechanical properties of the whole notochord.*"

  This factor plays a key role in developmental dynamics, thereby affecting all levels down to **L1**. [Anderson and Stein 1987] cite rigidity as an example of true emergence: *"We emphasize that this rigidity is a true emergent property: none of the forces between actual particles are capable of action at a distance. It implies that the two ends cannot be decoupled completely without destroying the molecular order over a whole region between them."*

- **Feedback system goals** An important case is feedback control [Wiener 1948] (engineering), which is essentially the same as homeostasis [Guyton 1977] (biology) , implemented by a feedback control loop (**Figure 2**). This is one of the network motifs characterised by [Alon 2006]. Then the constraints $\mathcal{C}_{\mathbf{LI}}(t)$ in (19) depend on goals $G_{\mathbf{LI}}$ valid at level **L** but set at the Level of Influence **LI**. The goals $G_{\mathbf{LI}}$ are irreducible higher level variables because they obtain their causal power only by their undeniable effects via this closed loop.

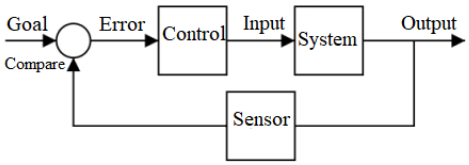


Figure 2: **Feedback control loop** *A goal determines outputs by being compared with the current system state, the difference being fed back to a controller that alters the state of the system (this is an essentially non-linear relation, which is why it cannot be characterised by any simple aggregation process).*

An engineering example is adjusting the (macro level) temperature setting on a thermostat, which determines the (micro level) motions of billions of molecules. The goals $G_{\mathbf{LI}}$ are irreducible macro level variables (their causal power vanishes if the feedback loop is broken). It is the dynamics at the feedback control loop level **LI** that determines what happens at that level, and hence at lower levels; this is the element of downward causation, where cause is understood as in [Pearl 2009] [Pearl and Mackenzie 2018]. This is a special case of a network.

- **Networks** Green and Batterman support the argument here by the general example of network dynamics. They say [Green and Batterman 2020]

  > "*The characteristics of functional equivalence classes are explained with reference to how network structures constrain dynamic outputs to enable generic types of functions such as sustained oscillations, noise filtering, robust perfect adaptation, signal amplification, etc.*"

These are functional principles that affect what networks actually exist. They lead to network motifs [Alon 2006] that represent higher level organising principles independent of the lower level dynamical details, that occur in many networks.

- **Dynamical system attractors** Consider a physical level **L3** whose dynamics can be characterised as a dynamical system [Arnold 1989]. It will have attractors at that level that are the causal element that order dynamic outcomes. They cannot in principle be characterised at Level **L1**, for they essentially involve the macro level interactions, but they reach down to shape dynamics at level **L1**, as for example in the case of fluid convection [Bishop 2008]. Paradoxically, they are not emergent rules, i.e. rules arising out of the lower level dynamics. Rather they are rules that characterise higher level dynamics of emergent entities quite independently of lower level dynamics [Green and Batterman 2020]. This is parallel to the claim by [Laughlin and Pines 2000] of existence of classical and quantum protectorates, governed by dynamical rules that characterise emergent systems as such.

- **Biological Needs** affect relative reproductive success, with selection reaching down to shape lower level outcomes [Campbell 1974] so as to meet higher level needs according to Eqn.(18). The selection criteria $c_{\mathbf{LI}}$ are higher level criteria related to biological purpose and function (Section 5.2). In effect, needs such as the ability to move and see act as attractors - higher level organising principles - for biological macro-evolution. This then reaches down to shape lower levels [Campbell 1974], with these needs leading to convergence to one or other of the restricted set of structures that can realise the needed function [McGhee 2006].

Underlying all these examples is the fact that Effective Theories $\mathbf{EF_L}$ at each level **L** are indeed effective. This is discussed further in §6.4.

## 6.3 Do Effective Field Theories disprove Strong emergence?

It has been suggested by [Hossenfelder 2019] and [Luu and Meißner 2019] that Effective Field Theories (EFTs) rule out strong emergence because they enable derivation of emergent properties **P** in a bottom up way,[17] in Hossenfelder's case even saying one should be able to in principle deduce election results in this way.

There is a wide variety of EFTs [Hartmann 2001], with some being closely related to the Renormalisation Group [Castellani 2002] [Burgess 2007]. The basic idea is to use a power series expansion of the Lagrangian in terms of $1/M$ where $M$ is the relevant mass scale of the microscopic theory, using symmetry principles to constrain the expansion, and using a power counting scheme. Using the EFT allows one to ignore the dynamics of fields relevant at high energies.

The philosophical literature on the topic includes [Bain 2013], [Butterfield 2014], and [Rivat and Gribaum 2020], who comment on the use of EFTs in condensed matter physics:

> "*Yet another example is the use of EFTs in condensed matter physics: even when the underlying theory is known, often the only tractable way to compute low-energy observables is to build an effective model as if the underlying theory were unknown*".

[17] Note that 'bottom-up' and 'top-down' are used exactly in the opposite way in discussions about EFTs and emergence. The top-down construction of an EFT means that we are constructing the EFT by deriving it from a more fundamental high energy theory. I will use the emergence appellation.

In other words it is derived as an EFT at the emergent level, without trying to derive it from the more fundamental level. Hence in this case it has nothing to say about emergence.

Most importantly, [Bain 2013] claims that if two successive theories related by an EFT are structurally different, for example if the Lagrangian has different symmetries, then the lower energy theory is essentially different from the underlying theory. He proposes

**(a) Ontological dependence**. *Physical systems described by an EFT are ontologically dependent on physical systems described by a high-energy theory.*

**(b) Failure of law-like deducibility**. *If we understand the laws of a theory encoded in a Lagrangian density to be its Euler-Lagrange equations of motion, then the phenomena described by an EFT are not deducible consequences of the laws of a high-energy theory.*

**(c) Ontological distinctness**. *The degrees of freedom of an EFT characterize physical systems that are ontologically distinct from physical systems characterized by the degrees of freedom of a high-energy theory.*

Thus Bain's paper supports both epistemological and ontological strong emergence. This viewpoint is controversial [Rivat and Gribaum 2020], but it suffices to prove that one cannot in an easy way state that EFTs disprove strong emergence. His view is supported by the fact that the broken symmetries result in the emergent state having an orthogonal Hilbert space to the Hilbert space of the underlying theory [Anderson 1984]. That is why they can be claimed to be ontologically different.

In any case [Adams *et al* 2006] show that some apparently perfectly sensible low-energy effective field theories governed by local, Lorentz-invariant Lagrangians, are secretly non-local, do not admit any Lorentz-invariant notion of causality, and are incompatible with a microscopic S-matrix satisfying the usual analyticity conditions. There are pitfalls in the way of claiming EFTs can characterise emergence of arbitrary low energy phenomena.

**Renormalisation group approaches** EFTs are often closely related to Renormalisation Group (RG) approaches. [Butterfield 2014] argues that the renormalisation group is compatible with Nagelian reduction, where as he explains, that idea has been extended to entail as regards the relation of the reducing theory T1 to the emergent theory T2 that

> (i) T2 may well have predicates, or other vocabulary, that do not occur in T1. So to secure its being a sub-theory, we need to augment T1 with sentences introducing such vocabulary, such that T2 is deducible from T1 as augmented.

Thus in my terms (Section 1.1), this is a case of weak derivation and hence a case of strong emergence. This is reinforced by the further extension [Butterfield 2014]

> (ii) What is deducible from T1 may not be exactly T2, but instead some part, or close analogue, of it. There need only be a strong analogy between T2 and what strictly follows from T1. Nagel called this approximative reduction.

This is simply not a genuine upward derivation of the emergent phenomena. It agrees with what [Leggett 1992] claims (see §3.2) and is compatible with strong emergence as defined by [Chalmers 2000]. Furthermore [Green and Batterman 2020] claim, on the basis of the universality classes that emerge, that a renormalisation group explanation extracts structural features that stabilize macroscopic phenomena irrespective of changes in or perturbations of microscopic details, hence supporting strong emergence. [Morrison 2012] strongly makes the same point:

"*What is truly significant about emergent phenomena is that we cannot appeal to microstructures in explaining or predicting these phenomena, even though they are constituted by them. Although this seems counter intuitive, RG methods reveal the nature of this ontological independence by demonstrating (1) how systems with completely different microstructures exhibit the same behavior, (2) how successive transformations give you a Hamiltonian for an ensemble that contains very different couplings from those that governed the initial ensemble, and (3) the importance of the physics behind the notion of an infinite spatial extension for establishing long-range order. ... RG reveals why the microdetails of symmetry breaking are neither ontologically nor epistemically necessary for emergence; the information is simply lost as the length scale changes.*"

This argument goes even further than the symmetry breaking arguments given above.

## 6.4 Equal Validity of Levels

Luu and Meissner in a paper on Effective Field Theories [Luu and Meißner 2019] state that in physics, one has an "*Equal validity of all levels*". This is indeed the case: this is the physics version of Denis Noble's "Principle of Biological Relativity" [Noble 2012], which states that in biology, no emergent level is privileged over any other.

This is expressed very nicely by Sylvan Schweber in [Schweber 1993], commenting on Phil Anderson's views:

> *"Anderson believes in emergent laws. He holds the view that each level has its own "fundamental" laws and its own ontology. Translated into the language of particle physicists, Anderson would say each level has its effective Lagrangian and its set of quasistable particles. In each level the effective Lagrangian - the "fundamental" description at that level - is the best we can do."*

This has to be the case because we don't know the underlying Theory of Everything (TOE) of physics, if there is one, and so don't - and can't - use it in real applications. So all the physics laws we use in applications are effective laws in this sense, applicable at the appropriate level (Section 4.5). It may be the Standard Model of Particle Physics, or Quantum Field Theory, or Quantum Theory, or nuclear physics, or atomic physics, and so on, depending on the problem at hand; each is a very well tested effective theory at the appropriate level. Similarly, there are very well tested effective theories at each level in biology: the molecular level, the cellular level, the physiological systems level for example.

The point then is this: each of them represents a causally valid theory, where causation can be defined as in [Pearl 2009] [Pearl and Mackenzie 2018], that holds at its level. None of them can be deemed to be more or less fundamental than any other, because *none of them is fundamental* (i.e. none is the hoped for TOE).

In terms of Effective Theories [Castellani 2002], this is the statement

> **Equal Causal Validity**: ***Each well established emergent level L in physics and biology represents an Effective Theory $ET_L$ as represented in (8), with variables $v_L$ appropriate to that level, so each level is equally valid with each other in a causal sense. None is the fundamental theory to which all else reduces.***

In particular, there is no known bottom-most physical level to which all of physics - or any other emergent level - can be reduced. No TOE can invalidate causation taking place at each emergent level. This equal causal validity occurs because higher levels are linked

to lower levels by a combination of upwards and downwards causation (sees Section 4 and [Noble 2012], [Ellis 2016]). I develop this all in more detail in a companion paper on interlevel causal closure [Ellis 2020]. The punchline is that whatever we may consider as the bottom-most physics level, it is not causally closed by itself. One must take into account higher levels in order for causal closure to occur.

## 6.5 Novel Results

New results in this paper include,

- The distinguishing in Section 2.3 of three different kinds of symmetry breaking mechanisms: **SSB**(**m**), Spontaneous Symmetry Breaking occurring at the micro level; **SSB**(**M**), Spontaneous Symmetry breaking occurring due to the emergence processes **E** creating the macro level from the micro level, and **SB**(**NS**), symmetry breaking due to Darwinian processes of natural selection.

On this basis, I provide

- The argument in Section 4.3 that emergence of properties based in **SSB**(**M**) is strong emergence. Such broken symmetries are the core of condensed matter physics [Anderson 1981], [Anderson 1989]. This implies solid state physics properties **P** are strongly emergent.

- The argument in Section 4.4 that lower level physics **m** is causally incomplete because it cannot by itself produce experimentally established outcomes of solid state physics. It only becomes causally complete if variables **a** based in higher level conditions are included to produce effective lower level dynamics **m'** that give the correct higher level outcomes, i.e. if downward causation $\mathbf{m} \rightarrow \mathbf{m'(m,a)}$ takes place. I develop the issue of causal closure in depth in a companion paper [Ellis 2020].

- The argument in Section 5.5 that strong emergence takes place in the case of living systems due to both **SSB**(**M**) and **SB**(**NS**).

- The argument in Section 6.4) that all levels are equally causally effective.

- Supporting arguments in Section 5.4 showing there is no dictionary relating basic physics variables to biology, and in Section 6.2 showing that essentially higher level variables do indeed occur.

- Section 6.1 shows that detailed initial data at the micro level does not uniquely determine outcomes at later times, allowing space for higher level variables to do so, and Section 6.3 shows that Effective Field Theories do not disprove strong emergence.

**Acknowledgments** I thank Barbara Drossel, Andrew Briggs, and Martin Rees for helpful comments. I thank Carole Bloch and Rob Adam for extremely helpful proposals regarding the scope of this paper. I thank two anonymous referees for comments that have greatly strengthened the paper.

# References


[Adams *et al* 2006] Adams, Allan, Nima Arkani-Hamed, Sergei Dubovsky, Alberto Nicolis, and Riccardo Rattazzi (2006).
"Causality, analyticity and an IR obstruction to UV completion." *Journal of High Energy Physics*10:014.

[Alberts 2007] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter (2007) *Molecular Biology of the Cell* (Garland Science)

[Alon 2006] Alon, Uri (2006).
*An introduction to systems biology: design principles of biological circuits* (CRC Press).

[Anderson 1972] Anderson, Philip W. (1972)
"More is different: Broken symmetry and the nature of the hierarchical structure of science". *Science* **177**: 393-396.

[Anderson 1981] Anderson, Philip W (1981).
"Some general thoughts about broken symmetry". In *Symmetries and Broken Symmetries in Condensed Matter Physics*, Ed N Boccaro (1981), pp:11-20. Reprinted in *A Career in Theoretical Physics* (World Scientific, 1994), pp.419-429.

[Anderson 1984] Anderson, Philip W (1984).
*Basic notions of condensed matter physics.* Reissued CRC Press, 2018.

[Anderson 1989] Anderson, Philip W (1989)
"Theoretical paradigms for the science of complexity" Nishina Memorial Lectures pp. 229-234. Reprinted in P W Anderson, *A Career in Theoretical Physics* (World Scientific), pp.584-593.

[Anderson 1994] Anderson, Philip W (1994)
*A Career in Theoretical Physics* (World Scientific).

[Anderson 2001] Anderson, Philip W (2001)
"Science: A 'dappled world' or a 'seamless web'?" *Stud. Hist. Phil. Mod. Phys.* **32**:487-494.

[Anderson and Stein 1987] Anderson, Philip W., and Daniel L. Stein (1987).
"Broken symmetry, emergent properties, dissipative structures, life." In *Self-organizing systems.* (Springer, Boston): 445-457 (Reprinted in [Anderson 1984]).

[Arnold 1989] V I Arnol'd (1989)
*Mathematical Methods of Classical Mechanics* (Springer)

[Bain 2013] Bain, Jonathan (2013).
"Emergence in effective field theories." *European journal for philosophy of science 3.3*: 257-273.

[Batterman 2018] Batterman, Robert W (2018)
"Autonomy of theories: An explanatory problem." *Noûs* **52**:858-873.

[Berridge 2014] M Berridge (2014)
*Cell Signalling Biology* Portland Press, doi:10.1042/csb0001001
http://www.cellsignallingbiology.co.uk/csb/

[Binder and Ellis 2016] Binder, Philippe M, and George F R Ellis (2016).
"Nature, computation and complexity." *Physica Scripta* **91**:064004.

[Binney et al 1992] Binney, James J, N Dowrick, A Fisher, and M Newman (1992).
*The theory of critical phenomena: an introduction to the renormalization group*.
(Oxford University Press).

[Bishop 2005] Bishop, Robert C. (2005)
"Patching physics and chemistry together." *Philosophy of Science* **72**:710-722.

[Bishop 2008] Bishop, Robert C (2008).
"Downward causation in fluid convection." *Synthese* **160**: 229-248.

[Bishop and Ellis 2020] Bishop, Robert C., and George FR Ellis.
"Contextual Emergence of Physical Properties." *Foundations of Physics*: 1-30.

[Blundell 2019] Blundell, Stephen J (2019).
"Phase Transitions, Broken Symmetry and the Renormalization Group."
In *The Routledge Handbook of Emergence*, Ed. S Gibbs, R Hendry, and T Lancaster (Routledge): pp.237-247.

[Bronowski 2011] Bronowski, Jacob (2011).
*The ascent of man*. (Random House).

[Burgess 2007] Burgess, Cliff P (2007).
"An introduction to effective field theory." *Annu. Rev. Nucl. Part. Sci.* **57**:329-362.

[Butterfield 2014] Butterfield, Jeremy (2014)
*"Reduction, emergence, and renormalization." The Journal of Philosophy 111.1*:5-49.

[Campbell 1974] Campbell, Donald T. (1974).
"Downward causation in hierarchically organised biological systems" in *Studies in the Philosophy of Biology: Reduction and Related Problems*, eds F. J. Ayala and T. Dobhzansky (Berkeley, CA: University of California Press), 179-186.

[Campbell and Reece 2008] Campbell, Neil A and Jane B Reece (2008)
*Biology* (San Francisco: Benjamin Cummings).

[Carroll 2005] Carroll, Sean B (2005).
*Endless forms most beautiful*. (WW Norton & Company).

[Carroll 2008] Carroll, Sean. B. (2008).
Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1), 25-36.

[Castellani 2002] Castellani, Elena (2002)
"Reductionism, emergence, and effective field theories." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **33**: 251-267.

[Chalmers 2000] Chalmers, David J. (2006)
"Strong and weak emergence", In P. Davies and P. Clayton, Eds, *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion* (Oxford University Press).

[Chang 2017] Chang, Sung (2017)
"How squid build their graded-index spherical lenses" *Physics Today* **70**:26-28

[Davidson *et al* 2009] Davidson, Lance, Michelangelo von Dassow, and Jian Zhou (2009). "Multi-scale mechanics from molecules to morphogenesis." *The international journal of biochemistry and cell biology* **41**:2147-2162.

[Davies 2019] Davies, Paul (2019).
*The Demon in the Machine*: *How Hidden Webs of Information are Solving the Mystery of Life.* (University of Chicago Press).

[Del Santo and Gisin 2019] Del Santo, Flavio, and Nicolas Gisin (2019)
"Physics without determinism: Alternative interpretations of classical physics."
*Physical Review* **A 100**:6 062107.

[Drossel 2020] Drossel, Barbara (2020)
"Strong emergence in condensed matter physics". https://arxiv.org/abs/1909.01134.

[Ellis 2016] Ellis, George (2016)
*How can physics underlie the mind? Top-down causation in the human context*
(Springer-Verlag, Heidelberg.)

[Ellis 2019] Ellis, George (2019)
"Why reductionism does not work" Essay for the Kurt Gödel prize and references.

[Ellis 2020] Ellis, George F R (2020).
"The Causal Closure of Physics in Real World Contexts". arXiv:2006.00972

[Ellis and Drossel 2019] Ellis, George, and Barbara Drossel (2019)
"How Downwards Causation Occurs in Digital Computers." *Foundations of Physics* **49**: 1253-1277. https://arxiv.org/pdf/1908.10186

[Ellis and Kopel 2019] Ellis, George and Jonathan Kopel (2019)
"The dynamical emergence of biology from physics" *Frontiers in physiology* **9**: 1966.

[Ellis *et al* 2018] Ellis, George FR, Krzysztof A. Meissner, and Hermann Nicolai (2018).
"The physics of infinity." *Nature Physics* **14**: 770-772.

[Fink and Noble 2008] Fink, Martin and Denis Noble (2008)
"Noble Model" Scholarpedia, 3(2):1803

[Franklin and Knox 2018] Franklin, Alexander, and Eleanor Knox (2018)
. "Emergence without limits: The case of phonons." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **64**:68-78.

[Friston 2010] Friston, Karl (2010).
"The free-energy principle: a unified brain theory?." *Nature reviews neuroscience* **11**:127-138.

[Friston 2012] Friston, Karl (2012).
"A free energy principle for biological systems." *Entropy* **14**:2100-2121.

[Ghirardi 2007] Ghirardi, G. (2007).
*Sneaking a Look at God's Cards: Unraveling the Mysteries of Quantum Mechanics*
(Princeton University Press).

[Gibb *et al* 2019] Gibb, S., Hendry, R. F., and Lancaster, T. (Eds.) (2019).
*The Routledge Handbook of Emergence*. (Routledge).

*The Matter of the Mind*, ed. M Schouton and H L de Jongh (Blackwell): 76.

[Gillett 2019] Gillett, Carl (2019).
"Emergence, Downward Causation and its Alternatives: Critically surveying a foundational issue." *The Routledge Handbook of Emergence*. (Routledge), 99-110.

[Goodstein 1985] Goodstein, David L. (1985).
*States of matter*. (Dover, New York )

[Green and Batterman 2017] Green, Sara, and Robert Batterman (2017).
*"Biology meets physics: Reductionism and multi-scale modeling of morphogenesis." Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **61**:20-34.

[Green and Batterman 2020] Green, Sara, and Robert Batterman (2020).
"Causal slack and top-down causation: Universality and functional equivalence in physics and biology". To appear in *Top-Down Causation and Emergence*, Ed Otávio Bueno (Springer Synthese Library Book Series).

[Grundmann 2010] Grundmann, Marius (2010)
*Physics of semiconductors. An Introduction including Nanophysics and Applications* (Berlin: Springer)

[Guay and Sartenaer 2018] Guay, Alexandre, and Olivier Sartenaer (2018)
"Emergent quasiparticles. Or how to get a rich physics from a sober metaphysics." In: Bueno, O., Fagan, M., Chen, R.-L., Ed, *Individuation, Process and Scientific Practices,* (Oxford University Press : New York) 2018, p.214-235.

[Guyton 1977] Guyton, Arthur C (1977)
*Basic Human Physiology: Normal Functions and Mechanisms of Disease* (W B Saunders, Philadelphia)

[Hartmann 2001] Hartmann, Stephan (2001)
"Effective field theories, reductionism and scientific explanation". *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **32**:267-304.

[Hartwell *et al* 1999] Hartwell, Leland H., John J. Hopfield, Stanislas Leibler, Andrew W. Murray (1999)
"From molecular to modular cell biology." *Nature* **402**: 6761: C47-C52.

[Hodgkin and Huxley 1952] Hodgkin Andrew L, and Andrew Huxley AF (1952)
"A quantitative description of membrane current and its application ...". *The Journal of Physiology* **117**:500-544.

[Hoffmann 2012] Hoffmann, P (2012) *Life's Ratchet: How Molecular Machines Extract Order from Chaos* (Basic Books)

[Hofmeyer 2017] Hofmeyr Jan-Hendrik (2017)
"Basic Biological Anticipation" In R. Poli (ed.), *Handbook of Anticipation* (Springer International Publishing AG)

[Hofmeyer 2018] Hofmeyr, Jan-Hendrik (2018)
"Causation, Constructors and Codes" *Biosystems* **164**:121-127.

[Hossenfelder 2019] Hossenfelder, Sabine (2019)
"The Case for Strong Emergence." In *What is Fundamental?*, Ed. A Aguirre, B Foster, and Z Merali (Springer, Cham): 85-94.

[Hunger 2006] Hunger, Johannes (2006).
"How Classical Models of Explanation Fail to Cope with Chemistry." *Philosophy Of Chemistry.* (Springer, Dordrecht), pp. 129-156.

[Juarrero 2002] Juarrero, Alicia (2002).
*Dynamics in action: Intentional behavior as a complex system.* (MIT Press).

[Karplus 2014] Karplus, Martin (2014).
"Development of multiscale models for complex chemical systems" *Angewandte Chemie International* **53**:9992- 10005.

[Kvorning 2018] Kvorning, Thomas Klein (2018)
*Topological Quantum Matter: A Field Theoretical Perspective* (Springer)

[Lancaster 2019] Lancaster, Tom (2019)
"The emergence of excitations in quantum fields" In Gibb et al (Eds.). *The Routledge Handbook of Emergence* (Routledge):275-286

[Laughlin 1999] Laughlin, Robert B (1999)
"Fractional Quantisation". *Reviews of Modern Physics* **71**: 863.

[Laughlin and Pines 2000] Laughlin, Robert and David Pines (2000)
"The Theory of Everything". *Proc Nat Acad Sci* **97**:28-31.

[Leggett 1992] Leggett, Anthony J (1992)
"On the nature of research in condensed-state physics". *Foundations of Physics* **22**:221-233.

[Lehn 1993] Lehn, Jean-Marie (1993).
"Supramolecular Chemistry". *Science.* **260**:1762-23.

[Lehn 1995] Lehn, Jean-Marie (1995).
*Supramolecular Chemistry* (Wiley-VCH).

[Luu and Meißner 2019] Luu, Thomas and Ulf-G. Meißner (2019)
"On the Topic of Emergence from an Effective Field Theory Perspective"
https://arxiv.org/abs/1910.13770

[Mayr 2001] Mayr, Ernst (2001).
*What evolution is.* (Basic books).

[McGhee 2006] McGhee, George R (2006)
*The geometry of evolution: adaptive landscapes and theoretical morphospaces.*
(Cambridge University Press).

[McLeish *et al* 2019] McLeish, Tom, Mark Pexton, and Tom Lancaster (2019)
"Emergence and topological order in classical and quantum systems"
*Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **66**:155-169.

[Mellisinos 1990] Mellisinos, Adrian C. (1990).
*Principles of Modern Technology* (Cambridge: Cambridge University Press)

[Menzies 2003] Menzies, Peter (2003)
"The causal efficacy of mental states." *Physicalism and Mental Causation* (Exeter: Imprint Academic): 195-224.

[Morrison 2012] Morrison, Margaret (2012).
"Emergent physics and micro-ontology." *Philosophy of Science* **79**:141-166.

[Mossio 2013] Mossio, Matteo (2013).
*Closure, causal.* In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, H. Yokota.(Eds) *Encyclopedia of Systems Biology* (Springer) pp.415-418.

[Noble 2008] Noble Dennis (2008)
*The music of life: biology beyond genes.* (Oxford: Oxford University Press).

[Noble 2012] Noble, Denis (2012)
"A theory of biological relativity: no privileged level of causation." *Interface focus* **2**: 55-64.

[Noble and Noble 2018] Noble, Raymond, and Denis Noblem(2018).
"Harnessing stochasticity: How do organisms make choices?." *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**:106309.

[Nurse 2008] Nurse, Paul (2008).
"Life, logic and information." *Nature* **454**:424-426.

[Oyama et al 2001] Oyama, Susan, Paul E Griffiths, and Russell D Gray (2001)
*Cycles of Contingency: Developmental Systems and Evolution* (Cambridge Mass: MIT Press)

[Peacock 1989] Peacocke, Arthur R (1989)
*An introduction to the physical chemistry of biological organization* (Oxford University Press, USA).

[Pearl 2009] Pearl, Judea (2009).
*Causality: Models, Reasoning, and Inference*. Cambridge university press, 2009.

[Pearl and Mackenzie 2018] Pearl, Judea, and Dana Mackenzie (2018).
*The book of why: the new science of cause and effect*. (Basic Books).

[Petsko and Ringe 2009] Petsko, Gregory A. and Dagmar Ringe (2009)
*Protein structure and function.* (Oxford University Press).

[Phillips 2012] Phillips, Philip (2012)
*Advanced Solid State Physics* (Cambridge University ress)

[Rhoades and Pflanzer 1989] R Rhoades and R Pflanzer (1989)
*Human physiology* (Fort Worth: Saunders College Publishing)

[Rivat and Gribaum 2020] Rivat, Sébastien, and Alexei Grinbaum (2020)
"Philosophical foundations of effective field theories." *The European Physical Journal* **A 56**:1-10.

[Schwabl (2007)] Schwabl, F. (2007).
*Quantum mechanics* (Springer Verlag Berlin Heidelberg)

[Schweber 1993] Schweber, Silvan S (1993)
"Physics, community, and the crisis in physical theory." *Physics Today* **46**: 34-34.

[Scott 1999] Scott, Alwyn (1999)
*Stairway to the mind: the controversial new science of consciousness*. (Springer Science and Business Media).

[Simon 2013] Simon, Steven H (2013)
*The Oxford solid state basics*. (OUP Oxford).

[Solyom (2009)] Solyom, J. (2009).
*Fundamentals of the physics of solids Volume II: Electronic Properties* (Springer Verlag Berlin Heidelberg)

[Sweeney *et al* 2007] Sweeney, Alison, D Marais, Y-En Ban, and S Johnsen (2007).
"Evolution of graded refractive index in squid lenses." *Journal of the Royal Society Interface* **4**: 685-698.

[Tanenbaum 2006] Tanenbaum, Andrew S (2006)
*Structured Computer Organisation* (Prentice Hall, Englewood Cliffs), 5th Edition.

[Wagner 2012] Wagner, Andreas (2012)
'The role of randomness in Darwinian evolution." *Philosophy of Science* **79**: 95-119.

[Wagner 2014] Wagner, Andreas (2014)
*Arrival of the fittest: solving evolution's greatest puzzle*. (Penguin).

[Watson *et al* 2013] J D Watson *et al* (2013)
*Molecular Biology of the Gene* (Pearson)

[Weinberg 2009] Weinberg, Steven (2009)
"Effective field theory, past and future." *Memorial Volume For Y. Nambu*. (World Scientific), 1-24.

[Wiener 1948] Wiener, Norbert (1948).
*Cybernetics or Control and Communication in the Animal and the Machine*. (MIT Press).

[Wilson 1982] Wilson, Kenneth G (1982)
"The renormalization group and critical phenomena"
https://www.nobelprize.org/uploads/2018/06/wilson-lecture-2.pdf

[Wright 1932l] Wright, Sewall (1932)
"The roles of mutation, inbreeding, crossbreeding, and selection in evolution" *Proceedings of the Sixth International Congress on Genetics*. **1** (8): 355-66.

[Zang *et al* 1989] Zhang, S. C., Th H. Hansson, and S. Kivelson (1989
"Effective-field-theory model for the fractional quantum Hall effect" *Physical Review Letters* **62**:82.

[Ziman 1979] Ziman, J M (1979)
*Principles of the Theory of Solids* (Cambridge University Press)

Version 2020/06/30

Foundations of Physics (2020) 50:1057–1097
https://doi.org/10.1007/s10701-020-00366-0

# The Causal Closure of Physics in Real World Contexts

**George F. R. Ellis**[1]

Received: 9 June 2020 / Accepted: 22 July 2020 / Published online: 18 August 2020


**Abstract**
The causal closure of physics is usually discussed in a context free way. Here I discuss it in the context of engineering systems and biology, where strong emergence takes place due to a combination of upwards emergence and downwards causation (Ellis, Emergence in Solid State Physics and Biology, 2020, arXiv:2004.13591). Firstly, I show that causal closure is strictly limited in terms of spatial interactions because these are cases that are of necessity strongly interacting with the environment. Effective Spatial Closure holds ceteris parabus, and can be violated by Black Swan Events. Secondly, I show that causal closure in the hierarchy of emergence is a strictly interlevel affair, and in the cases of engineering and biology encompasses all levels from the social level to the particle physics level. However Effective Causal Closure can usefully be defined for a restricted set of levels, and one can experimentally determine Effective Theories that hold at each level. This does not however imply those effective theories are causally complete by themselves. In particular, the particle physics level is not causally complete by itself in the contexts of solid state physics (because of interlevel wave–particle duality), digital computers (where algorithms determine outcomes), or biology (because of time dependent constraints). Furthermore Inextricably Intertwined Levels occur in all these contexts.

## 1 The Context

It is often supposed that causal closure occurs at the micro level in physical systems, and hence prevents the occurrence of strong emergence because the macrostate supervenes on the microstate [81, 82]. This is discussed in [27, 68, 89, 54]. In contrast, [20] shows by careful philosophical argument that

> One can have emergence with reduction, as well as without it; and emergence without supervenience, as well as with it.

✉ George F. R. Ellis
george.ellis@uct.ac.za

[1] Mathematics Department, University of Cape Town, Cape Town, South Africa

Here I want to examine the issue in a different way, by dealing in some detail with the hierarchical nature of emergence in real world contexts: the cases of engineering, based in the underlying solid state physics, and biology, based in the underlying molecular biology, in turn based in the underlying physics. The context is my paper [41] that establishes that strong emergence does indeed take place in both those cases, so that the argument against strong emergence has to be wrong in those contexts.

Reference [41] examines the issue of strong emergence of properties **P** of macrodynamics **M** out of the underlying microdynamics **m** in the context of condensed matter physics and biology. Following Anderson's lead [4, 5] that symmetry breaking is at the heart of emergence, its method was to identify five different kinds of symmetry breaking occurring in different contexts ([41], Sect. 2.3] and then to trace how broken symmetry states at the macro and micro levels interact with each other.

- The microscale dynamics **m**, based in the Laws of Physics $L$, obeys symmetries **S**: $\mathbf{S(m)=m}$. Spontaneous Symmetry Breaking **SSB**(**M**) leads to symmetry breaking of the macro scale dynamics **M** through the emergence process **E** whereby **M** emerges from the microscale dynamics **m**. This is weak emergence $\mathbf{E : m \rightarrow M : S(M) \neq M}$.
- This spontaneously broken macro state **M** reaches down to create quasiparticles such as phonons at the micro level, which play a key dynamical role at that level.[1] The base microdynamics **m** is altered to produce an effective microdynamics $\mathbf{m'}$ which breaks the symmetry of the underlying physical laws $L$. Thus $\mathbf{m \rightarrow m' : S(m') \neq m}$.
- To derive correctly the properties of macro dynamics **M** from the micro dynamical level, you must coarse grain the effective theory $\mathbf{m'}$ rather than **m**.
- Thus strong emergence takes place in this case: you cannot even in principle derive the macrodynamics **M** from the microdynamics **m** in a strictly bottom up way, because **m** satisfies the symmetry **S** and **M** does not.[2]

As a consequence[41]: Sect. **4.4**, in the case of solid state physics, the underlying microphysics **m** cannot be causally complete, because by itself it cannot lead to the emergence of known properties of solids such as electrical conductivity. The lower level physics only gives the correct outcome when modified by inclusion of terms **a(M)** arising from the higher level state **M** (so $\mathbf{S(a) \neq a}$). The same is true for living systems. That is, in both these cases, causal completeness is only attained by considering both the low level properties **m** and the higher level properties **M** (which lead to the alteration $\mathbf{m \rightarrow m'}$) together. The real causally closed system comprises both those levels.

The aim of this paper is to extend that result by investigating causal closure of physics in terms of determining dynamic properties[3] **P(d)** of entities in engineering

[1] See the discussion in Sect. 6.3.

[2] Unless the coarse-graining operation **C** breaks the symmetry **S**: then $\mathbf{CS \neq SC}$.

[3] I am using the classification in ([41], Sect. 2.2).

**Table 1** The emergent hierarchy of structure and causation for engineering (left) and life sciences (right) (developed from [39] and [43])

| | Engineering | Life Sciences |
|---|---|---|
| Level 9 (**L9**) | Environment | Environment |
| Level 8 (**L8**) | Sociology/Economics/ Politics | Sociology/ Economics/ Politics |
| Level 7 (**L7**) | Machines | Individuals |
| Level 6 (**L6**) | Components | Organs |
| Level 5 (**L5**) | Devices | Cells |
| Level 4 (**L4**) | Crystals | Biomolecules |
| Level 3 (**L3**) | Atomic Physics | Atomic Physics |
| Level 2 (**L2**) | Nuclear Physics | Nuclear Physics |
| Level 1 (**L1**) | Particle Physics | Particle Physics |

and biological contexts (I use the word closure rather than completeness for reasons that become apparent below as the theme develops). A separate very interesting project would extend this to considering the dynamic emergence of entities **E(d)** over time, see e.g. [24].

Section 2 sets the context for the discussion, which is the hierarchies of emergence in the cases of engineering and the life sciences respectively (Table 1). It introduce the idea of an *Effective Theory* $\mathbf{ET_L}$ at each level **L**, and discusses bottom up and top down causation in the hierarchy of emergence (Sect. 2.3).

Section 3 introduces the idea of a *Domain of Interest* **(DOI)**, and the concept of *Effective Spatial Closure* (Sect. 3.3). It is shown that in terms of spatial interactions, causal closure in engineering and biology only holds ceteris parabus. Yes of course philosophers know that this is the case; the point is that it has real consequences in real world contexts.

Section 4 defines *Levels of Interest* **(LOI)** and Sect. 4.4 introduces the need for *Restricted Domains of Interest*. Section 5 introduces *Interlevel Causal Closure* in the case of biology (Sect. 5.3), Sect. 6 extends this to the case of digital computers and physics.

Section 7 summarizes the main results of this paper, emphasizes that unavoidable unpredictability also undermines causal closure, and comments on ways people ignore the issues raised in this paper.

The novel concepts introduced are *Effective Theories* $\mathbf{ET_L}$ (Sect. 2.2), *Effective Spatial Closure* (Sect. 3.3), *Levels of Interest* (**LOI**) (Sect. 4), *Effective Causal Closure* (**ECC**) (Sect. 5.3), and *Inextricably Intertwined Levels* (**IIL**) (Sect. 5.4).

## 2 The Hierarchy of Emergence

The context of the discussion is the hierarchy of emergence. As stated by [4],

> At each level of complexity, entirely new properties appear ... At each stage entirely new laws, concepts, and generalizations are necessary.

In this section, I present the nature of the hierarchy (Sect. 2.1), and comment on Effective Theories and the Equal Validity of Levels (Sect. 2.2. The latter is enabled by a combination of upward and downward causation (Sect. 2.3), with the key feature of multiple realisability of higher levels in terms of lower level states (Sect. 2.4). The crucial relation between Effective Theories and Causal Closure is briefly commented on in Sect. 2.5. Going into more detail as regards the hierarchical structure, it is modular (Sect. 2.6), with the modules forming networks (Sect. 2.7).

### 2.1 The Nature of the Hierarchy

The emergent hierarchy is shown in Table 1 for the cases of engineering on the left and life sciences, in particular the case of humanity, on the right. The left hand side represent *The Sciences of the Artificial* as discussed by [126]. The right hand side represents the structures and processes of biology, as discussed by [23].

This hierarchy has important aspects.

- The bottom three levels **L1**–**L3** are the same on both sides. This is one of the great discoveries science has made: inanimate matter and living matter are made of the same stuff at the bottom. Electrons are at level **L3**, interacting with the nucleus.
- The atomic level **L3** is where new properties emerge out of the underlying physics, as characterised by the Periodic Table of the Elements, another great discovery.
- The components enabling complexity to arise occur at level **L4**. Both semiconductors and metals are crystals, and they are the key components of machines.[4] Solid State Physics covers levels **L3**–**L4**. Biomolecules such as DNA, RNA, and proteins are the foundations of biological emergence.
- Level **L5** is where the basic units of complexity arise, showing functional emergent properties, and being the basis for building complex entities. On the machines side, these are devices such as transistors, light emitting diodes, photodetectors, and lasers. On the life sciences side, they are cells: the basic building block of life, which come in many different types. This is the lowest level where the processes of life occur, entailing metabolism and information processing Hartwell et al. [61].
- Level **L6** is where on the machine side, devices are integrated into functional units, such as the Central Processing Unit in a Microprocessor, which is itself a device. On the life sciences side, organs comprising physiological systems occur.
- Level **L7** is where functional units occur that have an integrity of their own: they are effectively causally closed systems imbedded in a larger environment. On the engineering side, they are machines built to carry out some purpose, such as aircraft or digital computers or particle colliders. On the life sciences side, they are individuals with autonomy of action.

[4] Amorphous materials such as glasses may also occur, but they do not play a key role in the dynamic emergence of properties **P**(**d**) in machines.

- Level **L8** is the same on both sides. Both machines and individual human beings exist in the context of a society with social, economic, and political aspects, which sets the stage for their existence and functioning.
- Finally, level **L9** is again the same on both sides. It reflects the fact that each society exists in a natural environment with both ecological and geophysical aspects.

Note that this Table has chosen a particular set of levels to represent causation all the way from the Particle Physics Level **L1** to the Environmental Level **L9**. However most scientific studies will be interested in a much more restricted sets of levels: the *Levels of Interest* (**LOI**s) discussed in Sect. 4. Given such a choice, one will in general use a more fine-grained set of levels than represented in Table 1. Thus for example if the **LOI** is (**L4**–**L6**), one might divide that range into a finer set of sublevels.

### 2.2 Effective Theories and the Equal Validity of Levels

It is a common belief that the lower levels are more real than the higher levels, because bottom up causation from the lower to higher levels is the source of higher level properties. Arthur Eddington in *On the Nature of the Physical World* ([34], pp. 5–12) muses on the dual (solid macroscopic/atomic microscopic) nature of his writing desk, and concludes (p. 10) that because of the scientific world view,

> The external world of physics has thus become a world of shadows. In removing our illusions we have removed the substance, for indeed we have seen that substance is one of our great illusions.

However this view is subject to dispute. Richard Feynman in his book *The Character of Physical Law* (50], pp. 125–126) considers whether one level or another is more fundamental, and using a religious metaphor, argues that 'the fundamental laws are no nearer to God than emergent laws'.[5] Phil Anderson arguably had a similar view. Sylvan Schweber commented as follows [124]:

> Anderson believes in emergent laws. He holds the view that each level has its own "fundamental" laws and its own ontology. Translated into the language of particle physicists, Anderson would say each level has its effective Lagrangian and its set of quasistable particles. In each level the effective Lagrangian - the "fundamental" description at that level - is the best we can do.

Thus this does not recognize any level as more fundamental than any other.

Recently, Denis Noble has proposed a "Principle of Biological Relativity" [105]: all levels one deals with in studying emergence in biology are equally valid, there is no privileged level of causation.

[5] This passage is quoted in full in (39], pp. 454–455).

**Effective Theories** A good way to express this is that there is a valid *Effective Theory*[6] (ET) at each level. Elena Castellani gives this definition [25]:

> An effective theory (ET) is a theory which 'effectively' captures what is physically relevant in a given domain, where 'theory' is a set of fundamental equations (or simply some Lagrangian) for describing some entities, their behaviour and interactions... More precisely, an ET is an appropriate description of the important (relevant) physics in a given region of the parameter space of the physical world.

In parallel to the way the functioning of the Laws of Physics was sketched in [41], one can characterise an Effective Theory $\mathbf{ET_L}$ valid at some level $\mathbf{L}$ as follows;

> *An* **Effective Theory** $\mathbf{ET_L}$ *at a level* $\mathbf{L}$ *is a reliable relation between initial conditions described by effective variables* $v_\mathbf{L} \in \mathbf{L}$ *and outcomes* $o_\mathbf{L} \in \mathbf{L}$:
>
> $$\mathbf{ET_L} : v_\mathbf{L} \in \mathbf{L} \to \mathbf{ET_L}[v_\mathbf{L}] = o_\mathbf{L} \in \mathbf{L} \tag{1}$$
>
> *in a reliable way, whether* $\mathbf{ET_L}$ *is an exact or statistical law*.

It is important to note that an effective theory may have a randomisation element $\mathcal{R}$:

$$\mathcal{R} : v_\mathbf{L} \in \mathbf{L} \to \mathcal{R}(v_\mathbf{L}) = v'_\mathbf{L} \in \mathbf{L}, \tag{2}$$

where $\mathcal{R}$ might for example produce a Gaussian distribution.

**Equal Causal Validity** In terms of Effective Theories for emergent properties **P(d)**, Noble's principle [105] as extended in [41] can be restated:

> **Equal Causal Validity**: ***Each emergent level L in the hierarchy (characterised as in Table 1) represents an Effective Theory*** $\mathbf{ET_L}$, ***so each level is equally valid in a causal sense***.

This implies no level is a fundamental level with priority over the others, and particularly there is not a primary one at the bottom level. This is just as well, because there is no well-established bottom-most physical level to which physics can be reduced [102]. Every emergent level equally represents an effective theory.[7]

### 2.3 Upward and Downward Causation

Equality of validity of effective theories at every level is possible because causation is not just bottom-up. Rather higher level properties **P(d)** are linked to lower levels by a combination of upwards and downwards causation [39, 41, 105], which enables emergence of effective laws at each level.

**Upwards emergence** This has two different aspects ([41], §1.1).

[6] Not to be confused with an *Effective Field Theory* (EFT), see [19, 25, 60], which is a special case of an ET. Note that EFTs such as in [87] cannot deal with emergence in solid state physics, as they do not allow for symmetry breaking.

[7] While Luu and Meißner are critical of my claims on emergence [87], they agree on this point.

First there is the Emergence **E** of a macro system from its components. In terms of levels, this corresponds to creation of a higher level **LN** from a lower level **Ln**: that is, $\mathbf{Ln} \rightarrow \mathbf{LN}, \mathbf{N} > \mathbf{n}$. This may lead to topological non-trivial states emerging such as networks, or Quantum Entanglement may take place. The issue of phase transitions is important here. First order phase transitions occur when Spontaneous Symmetry Breaking **SSB** occurs leading to the emergent level **ET** having lower symmetries than the underlying **ET**. In terms of the associated micro dynamics **m** and macrodynamics **M**, if **S** is the symmetry set of **m**, then

$$\{\mathbf{E} : \mathbf{m} \rightarrow \mathbf{M},\ \mathbf{S(m)} = \mathbf{M}\} \ \Rightarrow\ \mathbf{S(M)} \neq \mathbf{M}. \tag{3}$$

Second there is emergence **P** of properties of the emergent level **LN** out of properties of the underlying constituent level **Ln** once **LN** has come into existence. This corresponds to emergence of a higher level $\mathbf{ET_L}$ out of a lower level one. Some form of coarse graining **C** of properties may suffice if the higher and lower levels have the same symmetries **S**, but not if their symmetries are different due to **SSB** (see [41]).

**Downward causation** A classification of different types of downward causation was given in [37, 39]. Here I will rather approach the issue from an $\mathbf{ET_L}$ viewpoint. There are essentially two kinds of downwards effects that can happen: downward alteration of lower level dynamics **L** via either constraints or effective potentials, and downward alteration of dynamics at level **L** by altering the set of lower level variables.

**Constraints and Effective Potentials** The way downward causation by constraints works is that the outcomes **P(d)** at Level **L** depend on constraints $\mathcal{C}_{\mathbf{LI}}$ at the level **L** arising from conditions at a Level of Influence **LI**. Thus when interlevel interactions are taken into account, relation (1) is modified (see [41, (30)]) to

$$\mathbf{ET_L}(\mathcal{C}_{\mathbf{LI}}) : v_{\mathbf{L}} \in \mathbf{L} \rightarrow \mathbf{ET_L}(\mathcal{C}_{\mathbf{LI}})[v_{\mathbf{L}}] = o_{\mathbf{L}} \in \mathbf{L}. \tag{4}$$

Essentially the same holds if the effect of the Level of Influence **LI** on the level **L** is expressed in terms of an effective potential $V(v_{\mathbf{LI}})$ at level **L** (see [41, (9)]). Thus constraints act as causes [76].

The constraints $\mathcal{C}_{\mathbf{LI}}$ may be time independent: $\partial \mathcal{C}_{\mathbf{LI}}/\partial t = 0$ in which case they are structural constraints; or they may be time dependent: $\mathcal{C}_{\mathbf{LI}} = \mathcal{C}_{\mathbf{LI}}(t), \partial \mathcal{C}_{\mathbf{LI}}/\partial t \neq 0$, in which case they are signalling or controlling constraints. An important case is feedback control (engineering), which is essentially the same as homeostasis (biology). Then the constraints $\mathcal{C}_{\mathbf{LI}}(t)$ depend on goals $G_{\mathbf{LI}}$ valid at level **L** but set at the Level of Influence **LI**. Similarly the potential $V(v_{\mathbf{LI}})$ may depend on time-dependent variables $v_{\mathbf{LI}}(t)$ at the Level of Influence **CL**. Then $\partial v_{\mathbf{LI}}(t)/\partial t \neq 0 \Rightarrow \partial V(v_{\mathbf{LI}})/\partial t \neq 0$. In both cases the level **L** is no longer causally complete on its own; at a minimum, only the combination {**L**,{**LI**} of levels can be causally complete.

**Altered variables** The causal effect due to the level of influence **LI** may rather be due to changes in the variables $v_{\mathbf{L}}$ at level **L** due to variables $v_{\mathbf{LI}}$ at the higher level **LI**:

$$\mathbf{ET_L}(v_{\mathbf{LI}}) : \{v_{\mathbf{L}}\} \in \mathbf{L} \rightarrow \mathbf{ET_L}(v_{\mathbf{LI}})[v_{\mathbf{L}}] = \{v'_{\mathbf{L}}\} \in \mathbf{L}, \tag{5}$$

where the new set $\{v'_{\mathbf{L}}\}$ of effective variables at level **L** may be smaller, larger, or altered.

They are *smaller* if they are changed by deleting lower level elements. This occurs when *Downward Causation by Adaptive Selection* takes place, altering or deleting selected lower level elements according to some selection criterion **c**. This enables alteration of structures and functions at level **L** so as to meet new challenges at level **LI**. This plays an important role in enabling by organisms to have agency and choice, enabled by stochasticity, as explained in [106]:

> Choice in the behavior of organisms involves novelty, which may be unpredictable. Yet in retrospect, we can usually provide a rationale for the choice. A deterministic view of life cannot explain this. The solution to this paradox is that organisms can harness stochasticity through which they can generate many possible solutions to environmental challenges. They must then employ a comparator to find the solution that fits the challenge. What therefore is unpredictable in prospect can become comprehensible in retrospect. Harnessing stochastic and/or chaotic processes is essential to the ability of organisms to have agency and to make choices.

They are *larger* if for example one has downward creation of quasiparticles such as phonons via interlevel wave–particle duality ([41] and Sect. 6.3), which underlies the properties of metals and semi-conductors. This is what Carl Gillett calls a "Foundational Determinative Relation" [55].

The are *altered* if the number is the same but the properties of an element changes. When they are bound in an emergent complex their own properties may change (for example, neutrons decay in 11 min when free but last for billions of years when bound in a nucleus), or their interactions with external entities may change (for example electrons bound in an atom interact with light quite differently than a free electron does).

Downward causation is related to Aristotle's *Formal Cause*, see [130], but I will not follow that strand here. To give these rather abstract statements flesh, see many examples given in [39, 104]. Downward causation in relation to the key physics–chemistry link is discussed in [86].

### 2.4 Multiple Realisability

Multiple realisability of higher level variables at lower levels plays a key role in downward causation [93]. Any particular higher level state can be realised in a multiplicity of ways in terms of lower level states. In an engineering or biological cases, a high level need determines the high level effective function that needs to be realised and thus the high level structure that fulfills it. This higher structure and function is then realised by suitable lower level structures and functions, but there are billions of ways this can happen

It does not matter which of the equivalence class of lower level realisations is used to fulfill the higher level need, as long as it is indeed fulfilled. Consequently you cannot even express the dynamics driving what is happening in a sensible way at a lower level.

The issue is not just the huge number lower level entities involved in realising a higher level systems, as characterised by Avagadro's Number It is the huge different numbers of ways combinations of lower level entities can represent a single higher level variable. Any one of the entire equivalence class at the lower level will do. Thus it is not the individual variables at the lower level that are the key to what is going on: it is the equivalence class to which they belong. But that whole equivalence class can be describer by a single variable at the macro level, so that is the real effective variable in the dynamics that is going on. This is a kind of interlevel duality:

$$\{v_{\mathbf{L}} \in \mathbf{L}\} \Leftrightarrow \{v_{\mathbf{i}} : v_{\mathbf{i}} \in E_{\mathbf{L-1}}(v_{\mathbf{L-1}}) \in (\mathbf{L-1})\}, \tag{6}$$

where $E_{\mathbf{L-1}}(v_{\mathbf{L-1}})$ is the equivalence class of variables $v_{\mathbf{L-1}}$ at Level $\mathbf{L-1}$ corresponding to the one variable $v_{\mathbf{L}}$ at Level $\mathbf{L}$. The effective law $\mathbf{ET}_{\mathbf{L}}$ at Level $\mathbf{L}$ for the (possibly vectorial or matrix) variables $v_{\mathbf{L}}$ at that level is equivalent to a law for an entire equivalence class $E_{\mathbf{L-1}}(v_{\mathbf{L-1}})$ of variables at Level **L-1**. It does not translate into an Effective Law for natural variables $v_{\mathbf{L-1}}$ per se at Level **L-1**.

### 2.5 Effective Theories and Causal Closure

It is important to note the following: one establishes the validity of an $\mathbf{ET}_{\mathbf{L}}$ for some chosen level $\mathbf{L}$ by doing experiments or making observations on phenomena occurring at that level. This involves the experimenter intervening at the level $\mathbf{L}$, hence it is an interlevel interaction. For example a particle physics experiment considers Effective Laws at level **L1** but involves scientists at level **L7** and organisations at level **L8** acting down to affect things at level **L1**. Consequently, one can make the following important observation:

> **Existence and functioning of Effective Theories $\mathbf{ET}_{\mathbf{L}}$ at level L does not necessarily imply causal closure of Level L**.

The issue is what determines constraints $\mathcal{C}$, potentials $V$, and effective variables $v'_{\mathbf{LI}}$ that may occur at that level. They may be influenced by other levels. That is what Sects. 5 and 6 are about. Determining an effective law at level $\mathbf{L}$ involves other levels then $\mathbf{L}$.

### 2.6 Modular Structure

Looking in more detail at the hierarchy, it is a hierarchy made of modules (this section) which form networks (next section). It is a modular hierarchy for very good reasons.

**Five principles of complex structure** ([16], §1.3) gives five principles of complex structure, developing from [126], starting from the idea

> **The Role of Decomposition**: The technique of mastering complexity has been known since ancient times: *divide et impera* (divide and rule).

The five principles, applicable to both engineering and biology, are stated by him to be,

(1) **Hierarchic Structure**: *Frequently, complexity takes the form of a hierarchy, whereby a complex system is composed of interrelated subsystems that have in turn their own subsystems, and so on, until some lowest level of elementary components is reached.*
(2) **Relative Primitives**: *The choice of what components in a system are primitive is relatively arbitrary and is largely up to the discretion of the observer of the system.*
(3) **Separation of Concerns**: *Intracomponent linkages are generally stronger than intercomponent linkages. This fact has the effect of separating the high-frequency dynamics of the components, involving the internal structure of the components, from the low frequency dynamics, involving interaction among components.*
(4) **Common Patterns** *Hierarchic systems are usually composed of only a few different kinds of subsystems in various combinations and arrangements.*
(5) **Stable Intermediate Forms** *A complex system that works is invariably found to have evolved from a simple system that worked. . . . A complex system designed from scratch never works and cannot be patched up to make it work. You have to start over, beginning with a working simple system.*

This underlies existence of levels such that each level is equally causally effective (Sect. 2.2). Booch says "Different objects collaborate with one another through patterns of interaction that we call mechanisms". These are what I am calling Effective Theories (ETs). The objects that collaborate are modules.

**Modules** Modularity is the property of a system that has been decomposed into a set of cohesive and loosely coupled modules ([16], p. 56). They can be represented by *Abstractions*, where "An abstraction denotes the essential characteristics of an object that distinguish it from all other kinds of objects and thus provide crisply defined conceptual boundaries, relative to the perspective of the viewer. An abstraction focuses on the outside view of an object and so serves to separate an object's essential behavior from its implementation" ([16], pp. 44–50). They involve *Encapsulation* ([16], pp. 50–53), that is, the internal details of the module's workings are hidden from the external world), and *Multiple realisability*: the required module functioning can be fulfilled in many ways by its internal structure and variables. Hierarchy is a ranking or ordering of abstractions ([16], p. 58).

### 2.7 Networks

A feature of particular interest is that emergent systems may give rise to Effective Theories that involve topological constraints. Indeed this happens quite often because emergent complexity in both engineering and biology often involves interaction networks, and a key feature of such networks is their topological connectivity, described by graph theory. Thus for example Arthur Peacocke points out that

> In electrical circuit theory there are certain topological constraints, the boundary conditions that one element imposes on another ([112], p. 74).

They obviously have strongly emergent properties: their functioning does not follow from any local characteristics of the elements that make up the circuit. The electric light won't work until you change its open circuit topology (isomorphic to an open interval) when the switch is off, to a closed topology (isomorphic to a circle) when the switch is on. This macro event than reaches done to alter the flow of billions of electrons at the micro level.

Networks can be physical networks, or interaction networks.

**Physical Networks** *Physical networks* are embodied in physical links between nodes, which constrain what interactions can take place by dictating what nodes can interact with what other nodes. Thus physical networks in fact create interaction networks by constraining interactions between links. This is the key structure–function relationship of engineering and biology. Examples in engineering are computer architecture [132], computer networks [83], and artificial neural networks [73]. The case of importance in biology is the nervous system [59] and neural networks [63].

**Interaction Networks** *Interaction networks* occur due to the presence of a variety of reagents that selectively interact with each other. This requires firstly a container that keeps the reagents within interaction distance of each other, rather than just diffusing away, and second the presence of an appropriate set of reagent that do indeed interact with each other. A key role is then played by selectively letting specific reagents enter or exit the container so as to control their interaction densities.

*On a large scale*, examples of importance in engineering are purification plants, chemical engineering reaction vessels, water treatment plants, sewage treatment plants. In biology, they arguably are the endocrine system, controlling signalling, and the digestive system, controlling metabolism at the systems level [59], and on a larger scale, ecological networks [75].

*On a small scale*, there are many interaction networks in cell biology [18]. These are crucially dependent on the existence of cells bounded by cell walls, that serve as the necessary reaction containers. They have ion channels imbedded in those walls that control movement of ions in and out of the cells, and molecular channels controlling movement of molecules in and out. They include

- Gene regulatory networks [24, 75, 139], also known as transcription networks [3],
- Metabolic networks [18, 75, 139],
- Cell signalling networks [11, 18],
- Protein interaction networks [18, 75],
- Signal transduction networks [18, 75].

These networks are the heart of cell biology [85] and underlie how information flows and logic underlie biological functioning as emphasized by [30, 45, 109].

**Networks and Hierarchy** Networks may have a hierarchical character in that subnetworks can often be identified within an overall network, and so define levels within the network [111, 121, 142]. This is an interesting topic I will

not develop further here except to remark that firstly, subnetworks include network motifs [3, 94], which are small subnetworks of particular functional significance. For example they include the *autoregulation motif*, which is nothing other than feedback control ([3], pp. 27–40) and the *feed-forward loop motif* ([3], pp. 41–73). They may contain higher-dimensional interactions characterised by *clique complexes* [118]. Networks may also contain hubs, central nodes of importance [75]. Their nature is highly dynamic [32].

**Causation** Because interaction networks are directed graphs (i.e. the edges between nodes have orientations), they represent causal effects, where causation is defined as [113, 114]. Reference [67] shows how such diagrams can be used to exemplify causal entailment in a diverse range of processes: enzyme action, construction of automata, and ribosomal polypeptide synthesis through the genetic code.

Their causal effects can be tested by *experiment*, where this is possible (vary conditions at one node and show that, ceteris parabus (i.e. conditions at other links to the node are unchanged) this results in a reliable change at another node. When this is not possible, one can use *counterfactual arguments*: demonstrating that as a result of the nature of the network links this should indeed be the outcome if one were to make such a change. This is the kind of argument I will use to claim that both upward and downward relations between levels are also causal (Table 3).

**Networks and strong emergence** Because of their systemic properties, biochemical networks display strong emergence [17].

## 3 Domains of Interest (DOI)

In examining the issue of causal closure of properties **P**, one must have the context clearly in mind. To do so, it is useful to define the *Domain of Interest* (**DOI**) of such study. This has three quite different aspects.

First, there will a specific *Topic of Interest* (**TOI**) one wishes to investigate. For example, it might be physics or engineering or chemistry or biology. In physics, one might have in mind atomic physics or condensed matter physics or plasma physics; in biology, molecular biology or physiology or neuroscience or population evolution. Or one might want to investigate relations between various of these topics.

In this paper, the interest is the nature of causal closure in the relation between physics, engineering, and biology.

Second, given a choice of topic of interest, the Domain of Interest **DOI** of a system of interest $S$ consists firstly of interaction limits for $S$ with its *surroundings*, and secondly of *time limits* on the duration when we are interested in the behaviour of $S$. Together these comprise spacetime limitations (Sect. 3.1), leading to *Effectively Isolated Systems* in the case of physics (Sect. 3.2) and *Effective Spatial Closure* in the case of biology and engineering (Sect. 3.3).

Thirdly, there will be a choice of *Levels of Interest* (**LOI**). The issue of **LOIs** is the focus of this paper, and is discussed in the following Sect. 4.

### 3.1 Spacetime Limitations

To be of physical interest, $S$ must be spatially limited. Although they are often talked about, systems of infinite extent do not occur in the real universe [46].[8]

Spacetime boundaries define the time and spatial domain we are interested in in relation to $S$. From a spacetime viewpoint, this is a world tube of finite radius $R$ that surrounds $S$, large enough to contain $S$ and all the elements strongly interacting with it, bounded by an initial time $t_i$ and final time $t_f$ defined in a suitable way. This governs the kinds of interactions it can have with other systems.

**Time limitations** We may be interested in short or long timescales characterised by the starting time $t_i$ and ending time $t_f$, depending on what we wish to study. We may be interested in,

- *Evolutionary processes* **E** whereby the family of systems of similar type to $S$ came into existence over long timescales via reproduction with variation followed by selection;
- *Developmental processes* **E** whereby a specific system $S$ came into existence through developmental or manufacturing processes, or perhaps by self assembly;
- *Functional processes* whereby the properties **P** of the system $S$ considered over short timescales emerge from the underlying physics. This is the focus of this paper.

Each involves very different choices of the timescale $\Delta t := t_f - t_i$ relevant to our study.

### 3.2 Effectively Isolated Systems

**Isolated systems** Causal closure of a system $S$ cannot happen if uncontrolled influences arrive from the surroundings:

$$\textbf{Sideways influences} : \{\textbf{Outside} \Rightarrow \textbf{S}\}. \tag{7}$$

As these influences vary with time, they will cause changes in the the state of the system with time that cannot be predicted from a knowledge of the properties of the system alone. The system is not causally closed.

Physics deals with this by introducing the idea of an *Isolated system*. This is usually expressed by giving limits on any incoming influences "at infinity", for example such conditions are imposed in studying electromagnetic and gravitational radiation.

However, as just stated, infinity is not a valid physics concept. One should instead refer to *Finite Infinity* $\mathcal{I}$ [36]: a world tube of finite radius $R_{\mathcal{I}} \gg R$ chosen so that

[8] Except perhaps for the Universe itself; but if this is indeed the case, it is of irrelevance to physics, because we can neither prove that this is the case or disprove it, and we cannot interact with or be affected by any regions outside our Particle Horizon [38].

incoming radiation and matter will not seriously interfere with $S$.[9] The dynamics of the system will then be autonomous except for small perturbations due to incoming matter and radiation crossing $\mathcal{I}$, which can be treated as small effects.

> **Effectively Isolated Systems (EIS)** ***What we can do is hope to find a world tube $\mathcal{I}$ of finite radius that serves as an effective infinity for the surface $S$. The dynamics of $S$ can be treated as an autonomous system, affected by small incoming perturbations over $\mathcal{I}$.***

However, there are two problems with this idea: one to do with physics, and one to do with engineering and biology.

**Causal Domains** The first is that famously, in general relativity, causal domains are determined by null cones rather than timelike tubes [62]. Why have I not defined the causal limits, which are basic in term of causal closure, in terms of null surfaces rather than a timelike world tube?

The answer is that on astronomical scales, effective causal limits are indeed given by timelike world tubes rather than null surfaces. On astronomical scales at recent times, the dynamic effects of radiation are very small compared with those of matter. We get a very usable $\mathcal{I}$ by choosing $R_{\mathcal{I}}$ to be about 1 Megaparsec in comoving coordinates [47]. Nothing outside there has had a significant effect on the history of our galaxy or the Solar System. Yes some radiation and matter is coming in, but it is negligible compared to the energies involved in daily life. The one form of radiation of significance for the world is light from the Sun, which comes from well within those limits: 1 Astronomical Unit might indeed suffice for local physics. The radiative energy coming from greater distances has negligible dynamical effects on Earth. A timelike world tube of radius 1 AU will do just fine in terms of considering causal closure of the Solar System.

**Isolated systems: Laboratories** However physics practice works in a different way: the key concept is an isolated system in a laboratory. It's a system that is in fact interacting strongly with the environment (Sect. 3.3), but that interaction is strictly controlled so that it is highly predictable. The system is shielded from influences outside the laboratory as far as possible. This then enables the results of experiments to also be highly predictable. And that is what enables the determination of the Effective Theories $\mathbf{ET_L}$ that hold at a Level **L**. Examples are the expensive isolation and cooling systems underlying the success of quantum optics experiments.

**Engineering and Biology as Open Systems** The real problem is different. It is that no biological system can be closed: they have by their nature to be open systems. And the same is true for engineering systems. In these cases, the 'isolated system' paradigm is simply wrong.

[9] We cannot shield from neutrinos and gravitational radiation no matter how we choose $\mathcal{I}$, but this does not matter as their effects are so weak.

### 3.3 Effective Spatial Closure

Life cannot exist as an isolated system. Biological systems are inherently open systems interacting with the environment [65, 112, pp. 10–11]) states,

> Biological organisations can only maintain themselves in existence if there is a flow of energy and this flow requires that the system not be in equilibrium and therefore spatially inhomogeneous.[10]

The effect of the outside world is not negligible. On the contrary, it is essential to biological functioning. It cannot be treated as a perturbation. The biosphere experiences incoming high grade radiation from the Sun and radiates outgoing low grade heat to the dark Sky, and this is its energy source enabling it to function thermodynamically [115]. Organisms need a flow of material in, and, because of the Second Law of Thermodynamics (essentially: as time progresses, matter and energy will be transformed from usable to unusable forms), need to dispose of waste matter and heat resulting from internal non-equilibrium metabolic processes. A living system **S** must take in materials and energy from the surrounding environment **E** and dispose of waste matter and energy to **E**:

$$\mathbf{E} \rightarrow \mathbf{S} \rightarrow \mathbf{E}. \tag{8}$$

In summary, living systems are essentially interacting systems. The same is true for engineering systems, because they do work of some kind.

**Reliable Interactions** They must therefore interact strongly with an environment that is stable enough that the interactions with the environment are reliable and reasonably constant so they do not disrupt the dynamics of the system over time.

In physics this is the concept of a Heath Bath or Thermal reservoir. You are in contact with an environment but don't need to take its dynamics into account because it is in a state usually assumed to be static, characterised only by a constant temperature $T$; and it is so large that the system $S$ has negligible influence on it state.[11]

**Life: Interaction limits**

How then does one limit those interactions to those that will enable life to sustain itself? Humberto Maturana and Francisco Varela essentially dealt with this by introducing the idea of *Autopoiesis* [91], which inter alia expresses the idea of causal closure in terms of system interaction with its environment.

**System Boundaries** The key point is how does one define the boundary of a system in this context. Instead of choosing a spacetime tube of some chosen radius $R$ as in the astronomy case, one chooses a *System Boundary* **B** that characterizes the system as being effectively autonomous. This underlies the meaning of Level **L8** in the hierarchy of emergence (Table 1, Sect. 2.1). A person has a skin that is her physical

[10] This is a kind of symmetry breaking different than the cases discussed in [41], and therefore may, by the same kind of arguments, be related to strong emergence in biology.

[11] Actually the assumption that $\partial T/\partial t = 0$ is problematic if one wants to explain the arrow of time [41]; but I'll leave that aside for the present.

boundary with the outside world, but still allows interaction with it. Energy and matter transfer takes place across the boundary. A machine similarly has a boundary that defines its limits, but will have some form of energy input enabling it to do work, and in many cases complex cooling devices to get waste heat out (paradoxically, they may consume large amounts of energy).

But this idea extends down to other levels, for example it holds also at the Device Level/Cellular Level **L5** and the Component Level/Organ Level **L6**. For example, a cell is the core of biology. It exists as an entity with its own integrity, characterised by the cell wall which allows controlled ingress and egress of materials and energy, yet interacts strongly with the environment.

How then does life handle functioning in this context? The environment must be sufficiently stable so as to allow effective predictability. This is the case when the system is not causally closed but has predictable interactions with the environment that makes its own functions predictable. It has an environment that can be treated as predictable up to perturbations. The environment may change with time, but if so, slowly enough to allow adaptation to the changing situation. This is often the case, and is what on the one hand allows living systems to flourish, and on the other allows biology to be a genuine scientific subject [23].

**The exceptions** But that is not always the case: take COVID-19 as an example. That started at the social level (**L9**) in one house, then spread worldwide via aircraft affecting life world wide and thereby affecting in a downward way all the biomolecules (**L4**) and electrons (**L1**) in the bodies of doctors and nurses and patients effected. But additionally it affected the engineering side by closing down thousands of flights across the world, thereby reaching down to affect all the billions of atoms (**L3**) and particles (**L1**) that make up those aircraft: a classic example of the inter-level causation I turn to next.

As far as predictability is concerned, the system $S$ has reliable interactions with the environment and so is predictable most of the time, except when we this is not the case. Predictability holds when all things are equal, but there is no guarantee this will be the situation. And that is the best we can do.[12]

> **Effective Spatial Closure (ESC)** ***Engineering and biological systems of necessity interact strongly with their environments, because that is necessary for their functioning. One can in practice usually set up a situation of*** **Effective Spatial Closure** ***where that interaction is by and large predictable so that the system will act in a predictable way. However there is no guarantee that this effective causal closure will always be as respected by the environment. Effective Spatial Closure is largely reliable, but holds ceteris parabus***.

Elaborating, Investopaedia states this concept as follows:

[12] And no this could not be predicted in principle in a strictly bottom up way if we knew the detailed positions and momenta of all the particles in our causal past to utmost precision, for multiple reasons [41, 42]. These are strongly emergent phenomena, unpredictable even in principle.

> Ceteris paribus *is a Latin phrase that generally means 'all other things being equal.' In economics, it acts as a shorthand indication of the effect one economic variable has on another, provided all other variables remain the same. ... Ceteris paribus assumptions help transform an otherwise deductive social science into a methodologically positive 'hard' science. It creates an imaginary system of rules and conditions from which economists can pursue a specific end. Put another way; it helps the economist circumvent human nature and the problems of limited knowledge'*.

In other words, Effective Spatial Closure works except when it doesn't. Unpredictability happens when *Black Swan Events* take place, possible when there is a fat-tailed rather than a Gaussian distribution [131], and with major significance at the macro level. And when it doesn't work, the effects chain all the way down from Level **L8** to the atomic level **L3** and particle level **L1** in Table 1, as in the case of the coronavirus pandemic.

## 4 Levels of Interest (LOI)

Setting this aside, the key conceptual issue I will deal with in this paper is the relation of the causal closure of emergent properties **P** to levels in the hierarchy of emergence.

Effective Theories and Existence of Levels is discussed in Sect. 4.1. Levels of Interest (**LOIs**) are defined in Sect. 4.2. Sensible choices for LOIs are discussed in Sect. 4.3. However there is a problem: Interactions span all levels: every LOI interacts with every other level by both upwards and downwards causation, so how can one get meaningful **LOIs**, or indeed a meaningful level? I elaborate on this problem in Sect. 4.4. A practical way out is by defining Effective Causal Closure, discussed later in Sect. 5.3. The choices one makes relate to whether one wants to answer *How* questions or *Why* questions (Sect. 4.5).

### 4.1 Effective Theories and Existence of Levels

**An Effective Theory $\mathbf{ET_L}$** (Sect. 2.2) is a set of variables and equations representing interactions and constraints at a particular level **L**, such that initial data implies a reliable outcome at that level. It is the possibility of existence of an Effective Theory $\mathbf{ET_L}$ (1) at each Level **L** (the dynamics at that level is determined at that level) that underlies the very concept of levels in the first place. As commented by Peacocke ([112], p. 10), following from [126]

> Natural hierarchies are more often than not 'nearly decomposable' - that is, the interactions among the sub-systems (the 'parts') are relatively weak compared withe the interactions among the subsystems, a property which simplifies there behaviour and description.

The fact that such levels exist is a consequence of the nature of the underlying physical laws and the values of the constants of nature [136]. It allows the existence of the

modular hierarchical structures that are the core foundation of complexity (Sects. 2.1, 2.6).

### 4.2 Definition of Levels of Interest LOIs

Reference [126] defines a hierarchical system as follows, quoted in ([112], p. 249):

> '***A hierarchical system***: A system of composed of inter-related subsystems, each of the latter being in turn, hierarchical in structure until we reach some lowest level of elementary subsystem (the choice of this lowest level he regarded as arbitrary).

This is what I have taken for granted above. While it has all the levels shown in Table 1 (Sect. 2.1), we usually do not want to consider them all at once. Reference [15] says it thus:

> Thus we take it as a given that when a portion of the universe is selected for study, be it a gas or a galaxy, we are allowed to blissfully ignore what is going on at scales that are much larger, or indeed much smaller, than the one we are considering.

That is what I am formalizing by defining the concept of Levels of Interest (**LOI**). They are defined as follows:

> **Levels of Interest** (**LOIs**) ***is a definition of the range of levels that will be covered by a theory*** .

A **LOI** is defined by its top level **TL** and its bottom level **BL**, thus

$$\mathbf{LOI}(\mathbf{TL} - \mathbf{BL}) := \{\mathbf{TL} \cup \mathbf{BL}\}, \tag{9}$$

where the Union sign "∪" means include all levels between **TL** and **BL**. Some studies are unilevel: **TL = BL**, and some are explicitly interlevel: **TL > BL**

- One can validly define such a LOI regardless of what levels you choose, because the levels are equally causally valid (Sect. 2.2).
- Your choice will depend on your Topic of Interest (**TOI**) (see Sect. 3). It is helpful if the choice of levels is made explicit, e.g. **LOI(3–5)** covers levels 3 to 5.
- Then one can for the purposes of studying the dynamics at those levels legitimately ignore higher and lower levels, in the sense to be explored below.
- This is possible because the levels effectively decouple in the sense of allowing valid Effective Theories $\mathbf{ET_L}$ at each level **L**.
- However this does not give you the right to deny the validity of levels that lie outside your Levels of Interest.
- There is no guarantee that causal completion will occur by including only those levels characterised by your choice of **LOI**.

That is the topic of Interlevel Causal Closure, which I discuss below (Sect. 5).

**Table 2** Levels of Interest **LOIs** for various academic disciplines

| | |
|---|---|
| Particle Physics [110] | **L1** |
| Nuclear Physics [64, §33–34] | **L1–L2** |
| Atomic Physics [64, §32] | **L2–L3** |
| Solid State Physics [127] | **L3–L4** |
| Computer Structure [132] | **L5–L7** |
| The Molecular Biology of the Gene [140] | **L3–L4** |
| The Molecular Biology of the Cell [2] | **L3–L5** |
| Neuroscience [80] | **L3–L6** |
| Physiology [120, 122] | **L4–L7** |
| Biology [23] | **L4–L8** |
| Major Transitions in Evolution [129] | **L4–L9** |
| Global Climate Change [72] | **L3–L9** |

### 4.3 Choice of LOIs

Are there limits on the **LOI**s one can choose? One can choose to investigate any desired LOIs, not investigating or take for granted the interactions that will inevitably occur from higher and lower levels. This is done to establish $\mathbf{ET_L}$s.

For example Noble [103] "Modeling the heart–from genes to cells to the whole organ" chooses to investigate the range of levels stated, namely Levels **L5** to **L8** in the Hierarchy of Emergence (Table 1), and not for example on the one hand the interactions between protons and electrons that make this possible, and on the other hand the mental and social influences that will inevitably be having an effect on how the heart is functioning. What he does do is investigate the interlevel relations within the levels **LOI(5–8)** he has chosen, because that is the domain of physiology.

Other examples are shown in Table 2.

Physics right down to Level **L1** always underlies what is happening, even if it lies outside the levels of interest to you. What happens in the real world right down to the physics levels **L1–L3** is always influenced by what happens at higher levels including **L9**, even if that is not what interests you in your particular studies.

What is a sensible DOI depends on conditions.

**Choice of Levels of Interest** *It all depends on what you want to understand. The suitable choice of **LOIs** will follow. This is the context within which "causal completeness of physics" must be evaluated.*

### 4.4 Interactions Span All Levels

**Strongly Interacting Levels** There is a basic problem with LOIs, however. That is the fact that, as mentioned above, every level interacts with every other level! The choices one makes relates inter alia as to whether one wants to answer *How* questions, *Why* questions, or both (Sect. 4.5). There is a practical way out that I discuss in Sect. 5.3 by introducing the idea of *Effective Causal Closure*.

### 4.5 The Whole Shebang

Table 1 omits two levels. At the bottom, it omits Level **L**: the level of a *Theory of Everything* (**TOE**). At the top, it omits Level **L10**: the level of Cosmology.

**At the bottom: The TOE and dynamic properties P(d)**

In order to examine the emergence of machines on the engineering side and organisms on the biology side, all one needs is Newton's Laws of motion, Maxwell's equations, Galilean gravity, and maybe the Schrödinger equation ([84], §4.1); in [41]. That is, Level **L3** is an adequate base. All engineering and biology emerges from this level, it is the lowest level engineers and biologists need to study. As explained in [41] and Sect. 2.3, conversely engineering and biology reach down to shape outcomes at Level **L3** via time dependent potentials or constraints.

However Levels **L3** and **L2** emerge from the Quantum Field Theory and the Standard Model of Particle Physics at Level **L1** [110]. Thus **L1** is a deeper foundation of emergence. However it is essentially decoupled from everyday life [4, 84]. Nevertheless outcomes at Level **L1** too must also be shaped in a downward way by engineering and biological variables, via outcomes at Level **L3**.

But this is not the bottom. Underlying **L1** is some theory of fundamental physics at Level **L0**, a "Theory of Everything" (**TOE**) [141]: maybe String Theory/M Theory, maybe not. There are variety of competing theories on offer [102]. This is the ultimate source of the emergence of dynamic properties **P**(**d**) in engineering and biology (using the notation of ([41], §2.2). It affects us every day.

These lie outside the **LOIs** of engineers and biologists. That is just as well, as we do not know what the answer is at Level **L0**, even though it underlies all physical emergence. As explained above, it suffices to deal with Effective Theories that hold at higher levels.

**At the top: Cosmology and Existence E** Consider Isaac Newton seeing an apple drop. This occurs for a variety of reasons: the Law of Gravity acting on the apple, the light rays that convey this image to his retina, the analysis of the image by his brain, and so on. But there are far deeper underlying issues. Why does the apple exist? Why does the Earth exist? Why does the Solar System and the Galaxy exist? Why does the Universe exist and have the nature it does? These are all the background reasons why the apple fell, and why Isaac Newton existed for that matter.

These questions of how everything came into being (**E**) is the domain of Cosmology [117] . It deals with issues such as, Where do elements such as Helium and Carbon come from? How did the Galaxy, the Sun, and the Earth arise? The Philosophy of Cosmology [38] considers issues such as Why is physics of such a nature as to allow life to exist? Reference [128]. Particularly: why are the constants of nature [136] of such a character as to allow the hierarchical structure in Table 1 to emerge? Thus cosmology affects us every day by underlying our existence.

**Everyday effects of cosmology P** There are more immediate issues as well, in the relation of cosmology to everyday life [36]: Why is the Sky dark at night, serving as a heat sink for the Earth's waste energy? This is crucial to the functioning of the biosphere ([115], pp. 411–417). Why is there an arrow of time? References [29, 44, 115]? This is crucial to all macro level physics, biology, and chemistry. Both are

due to the cosmological context. Thus cosmology affects the emergence of properties **P** in engineering and biology today.

The point then is that while it does indeed have a major causal effect on daily life [125], this is a rock solid relation that does not change with time. It is a fixed unchanging background that does not alter effective laws as time passes. It is thus not a case of ceteris parabus[13] (c.f. Sect. 3.3) and so can be taken for granted and not considered further when investigating causal closure in engineering, biology, and physics.

### 4.6 How Questions and Why Questions

An issue in choosing **LOIs** is if one is interested in *How* questions or *Why* questions.

**How questions** consider physical interactions on the one hand, and mechanisms on the other. Thus they will relate to levels **L1**–**L7** in Table 1, including **L7** because that relates to the integration of systems to produce the organism as a who [120, 122].

**Why questions** relate to motivation, meaning, and philosophical issues. Thus they will relate to levels **L7**–**L8** in Table 1, including **L7** because this is the level where as well as philosophy, psychology and motivation come in [33, 79] and **L8** where the causal power of social structures enters [35].

There is of course a trend for some strong reductionists to deny that the Why questions are valid or have any real meaning. From the viewpoint of this paper, that simply means that they themselves have a restricted set of **LOIs** that excludes those higher levels. Because of the equal validity of levels espoused in [41, 105], and in this paper, that restricted set of interests does not provide a justification for denying the validity of the levels with Effective Theories outside their particular set of interests.

## 5 Interlevel Causal Closure: Biology

The crucial concept in this paper is that of **Interlevel Causal Closure** of properties **P(d)**. I first consider Causation and Causal Closure (Sect. 5.1) and the nature of biology (Sect. 5.2). Then I consider Interlevel Causal Closure in the case of biology (Sect. 5.3), because this is where it is clearest and has been discussed most. I introduce here the key concept of ***Effective Causal Closure***. A stronger relation is the idea of ***Inextricably Intertwined Levels***(**IIL**s) which I discuss in Sect. 5.4.

[13] Unless we live in a false vacuum in which case local physics could suddenly change in a way that might wipe out all life [133].

**Table 3** The emergent hierarchy of structure and causation for engineering (left) and life sciences (right), indicating the upward and downward causation occurring

| | BU | Engineering | Life Sciences | TD |
|---|---|---|---|---|
| Level 9 (**L9**) | | Environment | Environment | ⇓ |
| Level 8 (**L8**) | ⇑ | Sociology/Economics/Politics | Sociology/Economics/Politics | ⇓ |
| Level 7 (**L7**) | ⇑ | Machines | Individuals | ⇓ |
| Level 6 (**L6**) | ⇑ | Components | Organs | ⇓ |
| Level 5 (**L5**) | ⇑ | Devices | Cells | ⇓ |
| Level 4 (**L4**) | ⇑ | Crystals | Biomolecules | ⇓ |
| Level 3 (**L3**) | ⇑ | Atomic Physics | Atomic Physics | ⇓ |
| Level 2 (**L2**) | ⇑ | Nuclear Physics | Nuclear Physics | ⇓ |
| Level 1 (**L1**) | ⇑ | Particle Physics | Particle Physics | |

### 5.1 Causation and Causal Closure

**Causation** In order to consider causal closure, one must first have a view on how one justifies claims of causation. This has been laid out in depth by Pearl [113, 114]. Causal inference is based in Causal Models (directed graphs) validated by experimental intervention, or when that is not possible, by Counterfactual Arguments.

**Causal Models** Here one consider causal models of the influences at work . In effect, the diagram of the hierarchy of Levels Table 1 in Sect. 2 *is* such a (very simplified) model, when one introduces the arrows of both upward emergence (left) and downward constraint or control (right).

To develop this approach more fully, one needs to expand Table 3 to a hierarchical diagram that represents the modular nature of the hierarchy (Sect. 2.6). This is a very worthwhile project, but I will not attempt it here. It is roughly indicated in ([112], pp. 8–11, 247–248), and examples are in ([18], pp. 6, 10–11, 22, 95, 110, 132, 160).

**Intervention** Here one actually intervenes at Level **LI** and reliably observes a resultant change at level **LF**. This has been done both for Effective Theories $\mathbf{ET_L}$ at each level **L**, and in many case for both upwards and downwards interlevel effects. One can do this also using digital computer models; for example [51, 105] have done this to show downwards causation occurring in computer models of heart function.

**Counterfactual views** [48] Here one considers what would happen if one intervened at Level **LI**, and plausibly argues that this will cause an actual difference at Level **LF**, when upward causation takes place: **LF > LI** (left column **BU**) or when downward causation takes place: **LI > LF** (right column **TD**). This has been used to establish that downward causal effects exist, e.g. [22, 43, 45].

**Causal Closure**Consider a multilevel system **S** (which could have only one level).

> **Causal Closure of the properties P of a system S(BL − TL) with a bottom level BL and top level TL** ***occurs when the set of Effective Laws*** $\mathbf{EF_L}$ ***governing outcomes at each level L, together with the upward and downward interactions between levels, are sufficient to determine the future state o of the system S at all levels {BL-TL} (the outcome) from an initial state*** **d** ***(the***

***data) given at a set of levels {LL-HL} with lowest level LL and highest level HL contained within or equal to {BL-TL}.***

Note that this includes the physicalist idea of causal closure where all follows from a single lowest physical level **LL**, for that is the case **LL = HL**, chosen as **L1** in Table 1.

Causal Closure requires Effective Predictability on the one hand (Sect. 3.3), and an Effectively Causally Closed set of levels on the other, which concept I now consider.

**The issue: Two opposing strands**

- There are no isolated sets of levels, as just discussed (Sect. 4.4). Causal Closure as just defined is an ideal that does not occur in practice unless one takes **LL = L1**, **HL = L9**: you are giving data for the whole thing.
- However there are in practice preferred restricted sets levels with a special integrity in terms of causal closure (see the comments just after Table 1).

How do we deal with this tension? The clearest domain in which to tackle this is biology (Sect. 5.3). The lessons from there carry over to engineering (Sect. 6.2), and physics (Sect. 6.4).

### 5.2 The Nature of Biology

Biological organisms have purpose, as stated by Nobel Prize winning biologist Hartwell et al. [61][14]:

> Although living systems obey the laws of physics and chemistry, the notion of function or purpose differentiates biology from other natural sciences. Organisms exist to reproduce, whereas, outside religious belief, rocks and stars have no purpose. Selection for function has produced the living cell, with a unique set of properties that distinguish it from inanimate systems of interacting molecules. Cells exist far from thermal equilibrium by harvesting energy from their environment. They are composed of thousands of different types of molecule. They contain information for their survival and reproduction, in the form of their DNA. Their interactions with the environment depend in a byzantine fashion on this information, and the information and the machinery that interprets it are replicated by reproducing the cell.

Consequently, as emphasized by Peacocke ([112], p. 13)

> *Many biological concepts and language are often* sui generis *and not reducible to physics and chemistry, certainly not in the form to which they apply to simpler and restricted atomic and molecular systems.*

[14] And see also [97].

In the case of biology, unless the concepts considered include purpose and function, it will miss the essence of what is going on, as pointed out by [61]. You also need to introduce the concepts "alive" and "dead", which do not occur at any lower level than the cellular level in biology, and do not occur at any physics level. Without this concept you cannot for example discuss the theory of natural selection [92].

**Upward and downward causation** As just stated, all biological entities have purpose or function, and that controls in a top-down way what happens at lower levels [105] reaching down to the underlying physical levels [45]. The physics does not control the higher levels, rather—without any violation of the laws of physics—it does what the biology asks it do. This functioning occurs via a combinations of upwards and downwards causation [104, 105], for example gene regulation taking place on the basis of the state of the heart [51] or the brain [78]. This dynamic reaches down to the molecular level and then the underlying electron level.

The enabling factors are black boxing [7] to get higher level logic out of lower level logic, together with time dependent constraints at the lower level that are regulated by higher level biological variables [45].

Together they underlie the emergent effective laws $\mathbf{ET_L}$ at each level **L** in biology.

**Preferred levels** The cellular level **L5** is a key level in biology: cells have an organisation and integrity of their own, and are living integral entities that are the basic units of life. They interact with other cells at the same level, and react to their environment in appropriate ways. In multicellular organisms they depend on higher levels for nutrition, materials, waste disposal, and signals as to what to do. But they can be treated as modules (Sect. 2.6) with an integrity of their own that responds to inputs and produces outputs. The associated set of biological levels, taking the underlying physics for granted, is the set of levels **L4–L5**.

Similarly the level of individual organisms **L7** again represents a level of emergent integrity. Individuals are entities that can be treated as autonomous entities that respond to environmental cues (from the levels above) and other individuals (at the same level). The associated set of biological levels, taking the underlying physics for granted, is the set of levels **L4–L7**.

So the issue is, how does this kind of autonomy emerge at these particular levels, given that all levels are interacting?

### 5.3 Effective Causal Closure in Biology

What characterizes these special sets of levels? The key point as to what occurs is organisational closure in biological organisms [100, 101]:

> The central aim of this paper consists in arguing that biological organisms realize a specific kind of causal regime that we call 'organisational closure'; i.e., a distinct level of causation, operating in addition to physical laws, generated by the action of material structures acting as constraints. We argue that organisational closure constitutes a fundamental property of biological systems since even its minimal instances are likely to possess at least some of the typical features of biological organisation as exhibited by more complex organisms.

This is a distinct causal regime, as explained in [98]:

> In biological systems, closure refers to a holistic feature such that their constitutive processes, operations and transformations (1) depend on each other for their production and maintenance and (2) collectively contribute to determine the conditions at which the whole organization can exist. According to several theoretical biologists, the concept of closure captures one of the central features of biological organization since it constitutes, as well as evolution by natural selection, an emergent and distinctively biological causal regime.

This is developed further in [95], identifying biological organisation as closure of constraints

> We propose a conceptual and formal characterisation of biological organisation as a closure of constraints. We first establish a distinction between two causal regimes at work in biological systems: processes, which refer to the whole set of changes occurring in non-equilibrium open thermodynamic conditions; and constraints, those entities which, while acting upon the processes, exhibit some form of conservation (symmetry) at the relevant time scales. We then argue that, in biological systems, constraints realise closure, i.e. mutual dependence such that they both depend on and contribute to maintaining each other.

Thus biological organisation is an interlevel affair, involving downward causation as well as upwards emergence, thus enabling teleology [12, 99]. From the viewpoint of this paper, these authors are identifying specific sets of levels where effective interlevel causal closure occurs: the topmost level links to the bottom-most level to close the dynamic loop that leads to biological emergence.

I will quote three more papers that have essentially the same view. Reference [66] emphasizes this property in the case of the cell:

> [The] property of self-fabrication is the most basic expression of biological anticipation and of life itself. Self-fabricating systems must be closed to efficient causation... I identify the classes of efficient biochemical causes in the cell and show how they are organized in a hierarchical cycle, the hallmark of a system closed to efficient causation. Broadly speaking, the three classes of efficient causes are the enzyme catalysts of covalent metabolic chemistry, the intracellular milieu that drives the supramolecular processes of chaperone-assisted folding and self-assembly of polypeptides and nucleic acids into functional catalysts and transporters, and the membrane transporters that maintain the intracellular milieu, in particular its electrolyte composition.

You need all these components and levels for the thing to work. Reference [49] emphasizes that multi-level homeostasis is part of the mix:

> Two broad features are jointly necessary for autonomous agency: organisational closure and the embodiment of an objective-function providing a ‘goal’: so far only organisms demonstrate both. Organisational closure has been studied (mostly in abstract), especially as cell autopoiesis and the cybernetic prin-

> ciples of autonomy, but the role of an internalised 'goal' and how it is instantiated by cell signalling and the functioning of nervous systems has received less attention. Here I add some biological 'flesh' to the cybernetic theory and trace the evolutionary development of step-changes in autonomy: (1) homeostasis of organisationally closed systems; (2) perception-action systems; (3) action selection systems; (4) cognitive systems; (5) memory supporting a self-model able to anticipate and evaluate actions and consequences. Each stage is characterised by the number of nested goal-directed control-loops embodied by the organism, summarised as will-nestedness.

Finally [107] argue for circular causality:

> We argue that (1) emergent phenomena are real and important; (2) for many of these, causality in their development and maintenance is necessarily circular; (3) the circularity occurs between levels of organization; (4) although the forms of causation can be different at different levels, there is no privileged level of causation a priori: the forms and roles of causation are open to experimental investigation; (5) the upward and downward forms of causation do not occur in sequence, they occur in parallel (i.e. simultaneously); (6) there is therefore no privileged direction of emergence - the upper levels constrain the events at the lower levels just as much as the lower levels are necessary for those upper-level constraints to exist. Modern biology has confirmed [...] that organisms harness stochasticity at low levels to generate their functionality. This example shows in fine detail why higher-level causality can, in many cases, be seen to be more important than lower-level processes.

This is closely related to the idea of *autonomous systems*, characterized by their organizational and operational closure [138], where [135]

> ***Organizational closure*** *refers to the self-referential (circular and recursive) network of relations that defines the system as a unity, and* ***operational closure*** *to the re-entrant and recurrent dynamics of such a system*

which is just the idea above. This discussion is related to the idea of Autopoiesis—a system capable of reproducing and maintaining itself—mentioned above, and to the idea of Autocatalytic sets [69, 70]. However I will not develop those links here. Rather my purpose is to claim that exactly the same applies in engineering systems in general, and even in physics in some cases. That is what I develop below.

Given that it is understood I am considering causal closure in terms of Levels and **LOIs**, I can summarise as follows[15]:

> **Effective Causal Closure (ECC) in Biology**: ***We have Effective Causal Closure of properties P(d) in a biological context when the considered set of levels {BL-TL} and data {HL-LL} is large enough to allow causal closure leading to autonomous biological functioning. It is "Effective"***

[15] This is related to the idea of inter level *causal entanglement*[137].

***because (i) we know other levels do indeed have an influence, but can regard those influences as inputs to an autonomous system that do not destroy its autonomy, and (ii) it is a ceteris parabus relation, as discussed in*** §3.3. ***It can be destroyed by unpredictable Black Swan events*** [131] ***that lie outside the normal operating environment.***

Thus this characterizes the set of levels needed for an entity (a cell or an organism) to function successfully. There is then no preferred level enabling the system to function: they all equally enable this to happen [107]. Reference [57] argue that in such cases, while the Effective Theories $\mathbf{ET_L}$ are contained in the range {**BL–TL**}, one needs different models at each level:

> No single mathematical model can account for behaviors at all spatial and temporal scales, and the modeler must therefore combine different mathematical models relying on different boundary conditions.

These are the different $\mathbf{ET_L}$s for each level. But note that one then needs data $d_{\mathbf{L}}$ for each level **L** too. Thus the set of levels where data is given has to be the same as the set of levels where **ECC** occurs. Thus ***in the definition of Effective Causal Closure just given, one should set*** {**BL–TL**} = {**LL–HL**}. Reference [57] give the examples of epithelial sheets and mechanical modeling of gastrulation.

To be clear: one is free to work with an Effective Theory $\mathbf{ET_L}$, with appropriate data for that level, at any level chosen **L** in the range {**BL–TL**}; but one only gets Effective Causal Closure by including that full set of levels and data.

**When does this occur in biology?** There are two cases where **ECC** occurs in biology.

- **Cells** The cellular level is the lowest level showing all the attributes of life. It is a case of **ECC** involving Levels **L4–L5**.
- **Individuals** The organism level is the major coherent emergent level in life, assuming it is a multicellular organism such as a human being. This is a case of **ECC** involving levels **L4–L7**.

However there is an interesting different view: that human beings are essentially social beings, so that in fact it is a mistake to view them as being capable of living on their own, as is implied by that categorisation, Thus Berger and Luckmann [10] wrote about the *Social Construction of Reality*: our worldview—an inescapable part of our nature shaping our actions—is crucially shaped by the society in which we live. Merlin Donald's book *A Mind so Rare* [33] essentially agrees, as does Andy Clark's book *Supersizing the Mind* [26] and [108]. In short, top-down effects from society so crucially shape our being that they are not just perturbations of independent existence: they are essential, and that characterisation is wrong. The correct **ECC** statement is

- **Social human beings** Human beings are essentially social, and are in fact a case of **ECC** involving levels **L4–L8**.

**Ignoring the lower levels**A further key comment regards the other end of the scale: why is it legitimate to ignore levels **L1–L3** here? The answer is the existence of *quantum and classical protectorates* that are governed by emergent rules and are insensitive to microscopics [84]. This is another way of affirming the causal efficacy of the Effective Theories $\mathbf{ET_L}$ at each emergent level **L**. However the Effectively Causally Closed levels will reach down to determine what happens at those levels via time dependent constraints (4) [45].

> **In summary:** ***Interlevel Effective Causal Closure as identified here is a key feature of biological functioning, emphasized in*** [8] ***and*** [96]. ***As well as being key in terms of emergence P(d) of properties, it is also key in terms of evolutionary and developmental processes E(d), see*** [24] ***(where it is not identified as such, but is there) and*** [123], ***where the relation is made explicit.***

**Setting the data** One final issue remains: we don't in practice set data at all the levels **LL–HL** required in a particular context.

> **Setting data**: ***In practice one sets data at the highest level HL=TL that is relevant to a particular problem, and lets downward causation cascade data down to choose any set of data in the required equivalence class at each lower level down to the lowest level LL=BL.***

There is no way we could in fact set the data at the lower levels. And there is no need for us to do so. This comment applies both in theory and in practice. That is, it is a statement both about epistemology (what we can know) and ontology (what we can do).

### 5.4 Inextricably Intertwined Levels: Biology

**A higher level may be essential to a lower level** An important possibility is that the properties **P** of two levels {**BL,TL**} may have an essential relationship with each other: each level cannot function without the other, as in some cases of symbiosis.

> **Inextricably Intertwined Levels** ***Two levels BL, TL are inextricably intertwined levels (IIL) if the effective dynamics*** $\mathbf{ET_{BL}}$, $\mathbf{ET_{TL}}$ ***at each of the two levels cannot occur without involving the other.***

Consider an individual human being. It is no surprise that the level **L7** of the individual cannot exist without the level **L5** of cells, for human beings are made out of cells and depend on them for their existence and physiological functioning. But the fact is that the converse is also true: the cells cannot exist and function without the existence of the body that they comprise. The reason is that cells have specialised for specific functions, and cannot survive on their own. They are supplied with oxygen-laden blood by the lungs, heart, and indeed the entire circulatory system, without which they die in a matter of minutes (as happens if a heart attack occurs). Thus levels **L5** and **L6** are inextricably intertwined. But organs are part of the individual and won't function without systemic integration at that level. Hence levels **{L5–L7}** are in fact inextricably intertwined.

**Table 4** Computer Implementation Hierarchy (schematic)

| Level | Entity | Nature |
|---|---|---|
| Level **L9** | Global Society | Global Social and Economic Context |
| Level **L8** | Country | Social and Economic Context |
| Level **L7** | Internet | Maximal Network |
| Level **L6** | Network | Linked computers, printers, file servers |
| Level **L5** | Computer | Integrated Circuits, I/O devices, Memory devices |
| Level **L4** | Integrated circuits | ALOE, CPU, Memory, linked by bus |
| Level **L3** | Gates | Boolean logic: AND, OR, NOT |
| Level **L2** | Transistors | Binary ON/OFF function |
| Level **L1** | Crystalline structure | Symmetry, Band Structure |
| Level **L0** | Electrons, Ions, Carriers | Structure, Current Flow |

This is the physical context within which upward emergence and downward causation takes place in the case of digital computers. For a full discussion, see ([39], Chap. 2)

Now an interesting issue arises: **ECC** occurs for levels **L5–L7**. Should I have included **L4** in the inextricably intertwined levels? Certainly Level **L4** is required in order that cells exist at level **L5**, but is the other way round true also? I believe one can claim it is, because the gene regulatory networks that control production of proteins at Level **L4** are at Level **L5**, and they would not exist if it were not for their functioning. Thus the real inextricably intertwined set of levels is {**L4–L7**}: the same as the **ECC** set of levels. There is however this difference: the **ECC** relation is ceteris parabus, as explained above. The **IIL** relation is not, it is essential, whatever happens at other levels, these levels are crucially dependent on each other.

## 6 Interlevel Causal Closure: Digital Computers, Physics

The discussion in the last two sections makes clear a set of principles that apply equally to engineering, and that is what I will show in this section.

To make the discussion concrete, I will consider the case of digital computers. But it will apply equally to other branches of engineering: automobiles, aircraft, chemical plant, water supply systems, sewerage systems, and so on.

I consider the nature of digital computers (Sect. 6.1), where Interlevel Effective Causal Closure again occurs (Sect. 6.2). Inextricably Intertwined Levels again occur in computers (Sect. 6.3), and in physics and chemistry (Sect. 6.4).

### 6.1 Digital Computers

The relevant hierarchy [43, 132][16] is shown in Table 4.

[16] The labeling of levels is a bit different than in Tables 1 and 3 because the focus here is specifically on computers.

**Upward and downward causation** The dynamics of a computer is driven by the algorithms encoded in the programs loaded, together with the data used by those programs. These control the flow of electrons through gates at the transistor level via a combinations of upward and downward causation ([39], Chap. 2, [43]). This enables the emergent effective laws $\mathbf{ET_L}$ at each level **L**. Different algorithms result in different flows of electrons, as can be demonstrated by running different computer programs which produce different patterns of electron flows through transistors at Level **L2**, and cause major effects at social levels **L8** and **L9** [88].

**The bigger picture** The fact that the higher levels {**L8,L9**} reach down to affect what happens at the lower levels {**L0–L6**} is stated in [43] as follows:

> ***Causal closure in the case of computers***: *In the real world, it is only the combination of physics with its logical, social, psychological, and engineering contexts (which includes the values guiding policy) that can be causally complete, because it is this whole that determines what computer programs will be written and what data utilised, hence what electron flows will take place in integrated circuits, as per the discussion in this paper*.

This is in parallel to the interlevel causal closure that takes place in biology, as discussed in Sect. 5.3. Reference [43] gives a specific example:

> As a specific example: the amount of money that can be dispersed to you from an ATM will be limited by an agreement you have reached with your bank. The program used to control the ATM will take into account the existence of such limits, and the specific amount you are able to take out in a given time period will be limited by a logical AND operation linking this agreed amount to the amount of money in your account. Thus these abstract variables will control electron flows in both the bank computers and the ATM dispenser mechanism. Every relevant abstract variable has physical counterparts; in other words, it?s realized by some physical properties on some relevant physical substrate.

But crucially there is much more than this: there is the whole issue of the purposes computers are used for in society, from controlling manufacturing to enabling the internet, cell phones, and social media, and they way that this whole enterprise is shaped by the values of those that control the system. The book *Coders* [134] considers "the morality and politics of code, including its implications for civic life and the economy. Programmers shape our everyday behavior: When they make something easy to do, we do more of it. When they make it hard or impossible, we do less of it." All this is expressed in the flows of electrons through gates at the digital levels **L2–L3**.

### 6.2 Effective Causal Closure in Computers

Nevertheless, just as in the case of biology, Effective Causal Closure can occur when interlevel causation results in a high degree of autonomy of operation. Analogously to the case of biology, one can state

**Effective Causal Closure (ECC) in Computers**: ***Effective Causal Closure of properties P in a digital computer occurs when the set of levels considered are large enough to allow causal closure leading to autonomous functioning. It is "Effective" because (i) we know other levels do indeed have an influence, but believe we can regard those influences as inputs to an autonomous system that do not destroy its autonomy, and (ii) it is a*** **ceteris parabus** *relation, as discussed in* §3.3. ***It can be destroyed by Black Swan events that lie outside the normal operating environment.***

The two emergent levels with their own causal integrity emerging through **ECC** are the integrated circuit level (**L4**), the equivalent of the cell in biology, with ECC given by Levels **L2–L4**; and the computer level (**L5**), the equivalent of the individual in biology, with ECC given by levels **L2–L5**. But just as in the case of biology one can make a case that one should really include the societal level, the same applies here too. The quotes above suggest that ECC for computers really only occurs for Levels **L2–L9**, including the highest level because of the effect of the World Wide Web.

**Engineering and Applied Physics** The same kind of considerations apply to all branches of engineering, and equally to all branches of applied physics. The applications (socially determined at Level **L8**) determine what physical effects occur (Levels **L1–L3**).

### 6.3 Inextricably Intertwined Levels: Computers

**Inextricably Intertwined Levels (IILs)** Do these occur in this case too, as they did in biology? Here there is a major difference: the transistors do not depend on the computer for their continued existence, whereas cells depend on the organism for their existence. While in biology **IILs** link the individual as a whole to the molecular level, here they also occur, but only at the levels **L0–L1** in Table 4.

The reason for that relation is that downward emergence of key properties at the electron level **L0** takes place, due to properties of the crystal level **L1**, as explained in detail in [41]. This is called a "Foundational Determinative Relation" (FDR) by Carl Gillett, see [55]. In more detail, quasiparticles such as phonons exist due to the broken symmetries of the emergent lattice structure. They come into being as effective particles at the lower level **L0** because they are dynamically equivalent to collective oscillations of a level **L1** structure (the crystal lattice) ([127], pp. 82–83).

This is an essentially quantum theory phenomenon. One can think of it as an interlevel **wave(macro)–particle(micro) duality.** Reference [52] say it this way:

> Phonons [are] quasi-particles that have some claim to be emergent, not least because the way in which they relate to the underlying crystal is almost precisely analogous to the way in which quantum particles relate to the underlying quantum field theory.

Stephen Blundell states the key point thus ([15], p. 244):

> So now we come to the key question: Are these emergent particles real? From the perspective of quantum field theory, the answer is a resounding yes. Each of these particles emerges from a wave-like description in a manner that is entirely analogous to that of photons. These emergent particles behave like particles: you can scatter other particles off them. Electrons will scatter off phonons, an interaction that is involved in superconductivity. Neutrons can be used to study the dispersion relation of both phonons and magnons using inelastic scattering techniques. Yes, they are the result of a collective excitation of an underlying substrate. But so are 'ordinary' electrons and photons, which are excitations of quantum field modes.

As a consequence, the levels {**L0, L1**} are inextricably intertwined. Another way of stating this is the way that Bloch's Theorem [14] shows how the crystal structure causes the electron wave functions to have a basis consisting of Bloch eigenstates with the same periodicity as the crystal.

This is a proof that the solid state physics occurring in digital computers is a case where causal closure is impossible at the micro level **L0** alone. It also shows that in general the set of **IILs** is not the same as the **ECCs**.

### 6.4 Inextricably Intertwined Levels: Physics and Chemistry

While these considerations apply to digital computers, of course they also apply in particular to the solid state physics itself that underlies their operation, due to the nature of crystals. For the reasons just discussed

> **In Solid State Physics**, ***as a consequence of interlevel wave-particle duality, levels {L1,L4} in Table 1 are inextricably intertwined levels.***

However the **IIL** phenomenon is not confined to this case. The laser is another example ([119], §2)

> The laser involves a kind of 'circular' causality which occurs in the continuous interplay between macrolevel resonances in the cavity guiding, and being reinforced by, self-organization of the molecular behaviour.

A quite different example is Resonance Energy Transfer (RET) in the transport of electronic energy from one atom or molecule to another [74]. As described in that paper,

> The individual electrons do not migrate between molecules during the transfer process, since the molecular orbitals (the wavefunctions) do not overlap, but instead move between individual electronic states within the molecules. ... energy transfer, through dipole coupling between molecules, mostly depends on two important quantities: spectral overlap and intermolecular distance.

This lead to a $r^{-6}$ distance-dependence for the resonance energy transfer rate in the short-distance regime. The behaviour results from interwining between the electron level **L1** and the molecular level **L4**. Because of these interactions, Second-order perturbation theory is the minimal needed to describe RET. While this paper refers

to 'molecules', RET occurs in atoms, chromophores, particles and carbon nanotubes. Its applications include nanosensors and photodynamic therapy.

This raises the issue that whenever molecular physics [21] is concerned, there is inextricable intertwining between the molecular structure at Level **L4**, bound by electrons, and the motions of the electrons at Level **L1**, controlled by that structure. This is manifest in binding energy ([21], pp. 17–19, [90], p. 569) and the existence of covalent bonds between atoms ([21], pp. 17–19).

> **Molecular physics and quantum chemistry** ***also exhibit IILs between levels L1 and L4, which is why chemistry is a classic example of downward causation*** [86].

The group theory underlying this intertwining is discussed in [13].

## 7 Conclusion

It is believed by many that because the bottom-most physics level is causally complete, and only upward causation takes place, higher levels are purely derivative: they have no real causal validity. In this paper and its companion [41], I argue against that position. It is invalid because it treats physics in a way that ignores context, whereas physics outcomes always depend on context. In fact downward effects imply the opposite: in real world contexts, the bottom-most physics level is not by itself causally complete.

In this section, I look at the contextual nature of causal closure of physics (Sect. 7.1), the way that unpredictability undermines causal closure of physics per se (Sect. 7.2), and comment on ways people ignore the issues discussed in this paper (Sect. 7.3).

### 7.1 The Contextual Nature of Causal Closure of Physics

We can consider physics per se, or in relation to the natural world, or in relation to biology, or in relation to engineering. Within physics, the issue of causal closure depends on what aspects we are considering: Particle Physics, Nuclear Physics, Condensed Matter Physics, Cosmology for example. Firstly, we have no reliable tested **TOE** at the very bottom level **L0**. We don't try to reduce to that most fundamental physical level (Sect. 4.4). Rather we reduce to a level that is convenient. That that can work is due to the existence of Quantum Protectorates, as explained in [84].

But then it is common to assume that Level **L1** (particle physics) is causally complete. Is that indeed so? I have argued that this is not the case in the contexts of physics and biology (Sect. 5.3); physics and engineering, as exemplified by digital computers (Sect. 6.2); solid state physics (Sect. 6.3); and physics and chemistry (Sect. 6.4). In each case the real causal closure that takes place is an interlevel affair, as emphasized in particular in the case of biology by many perceptive writers (Sect. 5.3). Effective Causal Closure in real world contexts spans many levels, in the case of biology reaching down from the level of the organism to the underlying

physics via time dependent constraints. This implies how it works in terms of physics in relation to society. The causal effects of the coronavirus pandemic at the social level reaches down to cause major changes at the physical levels **L1–L3** through a complex interaction between social behaviours, virology, and microbiology that for example has temporarily destroyed international air travel and so the trajectories of the billions of particles that make up aircraft. Causal closure only occurs when we take all these factors and levels into account.

Considering only disembodied physical laws seriously misleads about the nature of causation and causal closure in real world contexts. In summary,

> **Causal closure of physics** ***In the real world context of engineering and biology, physics at the lowest level considered, whatever that is, is not by itself causally complete. Interlevel causal closure involving engineering or biological variables, in those respective cases, is required in order to have an effectively causally closed system.***

This is formalized by the concepts of Effective Causal Closure (Sects. 5.3, 6.2), and Inextricably Intertwined Levels (Sects. 5.4, 6.3). Within solid state physics itself, the lower levels **L1–L3** are not causally closed by themselves because of *interlevel particle–wave duality* between the particle level **L1** where electrons and phonons live, and the crystal level **L4** where lattice vibrations take place (Sect. 6.3). Properties of Level **L4** decouple from the lower physics levels. As stated by [84],

> The crystalline state is the simplest known example of a quantum protectorate, a stable state of matter whose generic low-energy properties are determined by a higher organizing principle and nothing else.

### 7.2 Unpredictability Undermines Causal Closure of Physics Per Se

Ignoring interlevel issues, if by causal closure one means that data specified as precisely as possible leads to unique outcomes, then unavoidable unpredictability undermines the possibility of physics per se (whether quantum or classical) being causally closed.

**Quantum Physics** Quantum effects doubly cause uncertainty in outcomes.

Firstly, the Heisenberg Uncertainty Principle states that the standard deviations of position $\sigma_x$ and momentum $\sigma_p$ obeys

$$\sigma_x \sigma_p \geq \hbar/2 \tag{10}$$

so one cannot even in principle apply Laplace's dream of setting initial data precisely at level **L1**. Consequently, outcomes are also uncertain.

Secondly, collapse of the wave function introduces an irreducible uncertainty in classical outcomes at this level when interactions take place [53]. This can reach up to macro levels through various amplifiers such as photon multipliers and CCDs. In the engineering case it causes predictability issues at macro scales in terms of digital computer reliability because cosmic rays cause errors in computer memories [56, 143], and the emission of a cosmic ray by an excited atom is a quantum event that is

unpredictable even in principle. As regards biology, cosmic rays have had a significant effect on evolutionary history by causing genetic mutations [116].

**The Classical Case** Uncertainty of outcomes occurs in this case too, because one can't set initial data to infinite precision [31]. This is an outcome of the fact that infinity never occurs in physical reality [46]. Thus physics is not causally closed in the classical case at higher levels because of chaotic dynamics (the butterfly effect), together with the impossibility of specifying initial data to infinite accuracy. This occurs for instance at the level at which fluid motion is determined.

Reference [6] puts it this way:

> A fluid dynamicist when studying the chaotic outcome of convection in a Benard cell knows to a gnat's eyelash the equations of motion of his fluid but also knows, through the operation of those equations of motion, that the details of the outcome are fundamentally unpredictable, although he hopes to get to understand the gross behaviour. This aspect is an example of a very general reality: the existence of universal law does not, in general, produce deterministic, cause-and-effect behaviour.

This fundamentally undermines the concept of a causally closed physical levels in the case of classical physics. As was already known to Poincare, this occurs even in the 3-body gravitational case.

**Microbiology** In the case of microbiology, interactions take place in the context of what Hoffmann [65] has called "The Molecular Storm". Molecular machines use ratchet-like mechanisms to harness energy from that storm, and organisms use it to provide an ensemble of options from which they can choose preferred lower level states and so attain biological objectives [106] (see the quote in Sect. 2.3). One has the opposite of the calm relation between initial data and outcomes supposed by Laplace.

### 7.3 How to Ignore the Issue

Here are some ways that the nature of causal closure as discussed in this paper is avoided.

**Partial reduction** This is very common.

Francis Crick in *The astonishing hypothesis* [28] states,

> You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules.

In other words, he is reducing **L7** to {**L5–L4**}. Now my physics colleagues who believe that all that matters is the particle interactions at level **L1** will just laugh and say, cells at Level **L4** and molecules at **L5** are nothing but particles interacting with each other. Thus Crick believes in the reality and effectiveness of causality at Levels **L4** and **L5** that for example [71, 58], who believe that all causality resides at Level **L1**, clearly must deny.

So why did Crick emphasize causality at those levels? The answer of course is that those were the levels at which he worked—and experienced the effectiveness of causality in terms of the interactions between entities (molecules, neurons) at those levels.

From a strictly reductionist viewpoint, this is an illegitimate move. It is however fine if you accept Noble's Principle of Biological Relativity [105], as extended in [41] and this paper: then causality is real at the levels {**L4, L5**} he studies. But that removes Crick's justification for denying the reality of causation at Level **L7**.

**Ignoring context** Crucial contextual effects are simply ignored by some writers. For example [58, 71]. The view is "You are nothing but a bag of particles, it's just a matter of particles interacting via a known set of forces". Context has nothing to do with it. The physicists holding this view all come from the particle physics/cosmology side, where this is to some extent true. Physicists from the solid state physics side (the largest section of the physics community) do not hold this view, see e.g. [5, 127]. This leads to the large divide in the physics community between these two groups, as discussed by Schweber [124].

**Denying top-down causation** There is a frequent denial of the possibility of top-down causation, even though it occurs in physics and cosmology. In the latter case it occurs in the context of primordial nucleosynthesis in the early universe ([117], §4.3) and structure formation in the later universe ([117], §5), which are both dependent on the cosmological context at level **L10**. That is the reason that primordial element abundances on the one hand and matter power spectra and Cosmic Background Radiation angular power spectra on the other can be used to place strong limits on the background model parameters [1], as discussed in ([39], pp. 275–277). If there were not such downward causation, this would not be possible.

Top down causation clearly occurs in biology [22, 51, 104]. A recent example is [119]:

> Top-down approaches focus on system-wide states as causal actors in models and on the computational (or optimality) principles governing global system dynamics.

That paper gives fascinating examples of downward causation from morphogenesis and regenerative medicine.

Downward causation is obvious in subjects other than particle physics and cosmology, and in particular in the functioning of the brain [39, 40, 77, 79]. For example, our minds are shaped by society, and shape society, including material outcomes in terms of architecture, engineering, art, and so on. Society does not exist without mind, and mind does not exist without society [9, 10, 33]. That is the nature of inter-level causal closure.

## References

1. Aghanim, N., et al.: (The Planck Team) "Planck 2018 results. VI. Cosmological parameters" (2018). arXiv:1807.06209

2. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P.: Molecular Biology of the Cell. Garland Science, New York (2007)
3. Alon, U.: An Introduction to Systems Biology: Design Principles of Biological Circuits. CRC Press, London (2019)
4. Anderson, P.W.: More is different: broken symmetry and the nature of the hierarchical structure of science. Science **177**, 393–396 (1972)
5. Anderson, P.W.: A Career in Theoretical Physics. World Scientific, Singapore (1994)
6. Anderson, P.W.: Science: a 'dappled world' or a 'seamless web'? Stud. Hist. Philos. Mod. Phys. **32**, 487–494 (2001)
7. Ashby, W.R.: Design for a Brain: The Origin of Adaptive Behaviour. Springer, Dordrecht (2013)
8. Bechtel, W.: Biological mechanisms: organized to maintain autonomy. In: Systems Biology: Philosophical Foundations, pp. 269–302. Elsevier, Amsterdam (2007)
9. Berger, P.L.: Invitation to Sociology: A Humanistic Perspective. Anchor Books, New York (1963)
10. Berger, P.L., Luckmann, T.: The Social Construction of Reality: A Treatise in the Sociology of Knowledge. Penguin, New York (1991)
11. Berridge, M.: Cell Signalling Biology. Portland Press (2014). https://doi.org/10.1042/csb0001001
12. Bich, L., Mossio, M., Soto, A.M.: Glycemia regulation: from feedback loops to organizational closure. Front. Physiol. **11**, 69 (2020)
13. Bishop, R.C., Ellis, G.F.R.: Contextual emergence of physical properties. Found. Phys. **50**, 1–30 (2020)
14. Bloch, F.: Über die Quantenmechanik der Elektronen in Kristallgittern. Z. Phys. **52**, 555–600 (1929)
15. Blundell, S.J.: Phase transitions, broken symmetry and the renormalization group. In: Gibb, S., et al. (eds) The Routledge Handbook of Emergence, pp. 237–247. Routledge, Oxford (2019)
16. Booch, G.: Object Oriented Analysis and Design with Application, 2nd edn. Addison Wesley, Boston (2006)
17. Boogerd, F.C., Bruggeman, F.J., Richardson, R.C., Stephan, A., Westerhoff, H.V.: Emergence and its place in nature: a case study of biochemical networks. Synthese **145**, 131–164 (2005)
18. Buchanan, M., Caldarelli, G., De Los Rios, P., Rao, F., Vendruscolo, M. (eds.): Networks in Cell Biology. Cambridge University Press, Cambridge (2010)
19. Burgess, C.P.: An introduction to effective field theory. Annu. Rev. Nucl. Part. Sci. **57**, 329–362 (2007)
20. Butterfield, J.: Emergence, reduction and supervenience: a varied landscape. Found. Phys. **41**, 920–959 (2011)
21. Buyana, T.: Molecular Physics. World Scientific Publishing Co., Singapore (1997)
22. Campbell, D.T.: Downward causation in hierarchically organised biological systems. In: Ayala, F.J., Dobhzansky, T. (eds.) Studies in the Philosophy of Biology: Reduction and Related Problems, pp. 179–186. University of California Press, Berkeley (1974)
23. Campbell, N.A., Reece, J.B.: Biology. Benjamin Cummings, San Francisco (2005)
24. Carroll Sean, B.: Endless Forms Most Beautiful: The New Science of Evo Devo. WW Norton and Company, New York (2005)
25. Castellani, E.: Reductionism, emergence, and effective field theories. Stud. Hist. Philos. Sci. B **33**, 251–267 (2002)
26. Clark, A.: Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press, New York (2008)
27. Clayton, P.C., Davies, P.C.W. (eds): The Emergentist Hypothesis from Science to Religion, pp. 79–110. Oxford University Press, Oxford (2006)
28. Crick, F.: The Astonishing Hypothesis: The Scientific Search for the Soul. Touchstone, New York (1994)
29. Davies, P.: The arrow of time. Astron. Geophys. **46**, 1.26–1.29 (2004)
30. Davies, P.: The Demon in the Machine: How Hidden Webs of Information are Solving the Mystery of Life. University of Chicago Press, Chicago (2019)
31. Del Santo, F., Gisin, N.: Physics without determinism: alternative interpretations of classical physics. Phys. Rev. A **100**(6), 062107 (2019)
32. Deritei, D., Aird, W.C., Ercsey-Ravasz, M., Regan, E.R.: Principles of dynamical modularity in biological regulatory networks. Sci. Rep. **6**, 21957 (2016)
33. Donald, M.: A Mind So Rare: The Evolution of Human Consciousness. WW Norton and Company, New York (2001)

34. Eddington, A.S.: The Nature of the Physical World. Cambridge University Press, Cambridge (1929)
35. Elder-Vass, D.: The Causal Power of Social Structures. Cambridge University Press, Cambridge (2010)
36. Ellis, G.F.R.: Cosmology and local physics. N. Astron. Rev. **46**, 645–657 (2002). arXiv:gr-qc/01022017
37. Ellis, G.F.R.: Top-down causation and emergence: some comments on mechanisms. Interface Focus **2**, 126–140 (2012)
38. Ellis, G.F.R.: On the philosophy of cosmology. Stud. Hist. Philos. Sci. B **46**, 5–23 (2014)
39. Ellis, G.: How Can Physics Underlie the Mind? Top-Down Causation in the Human Context. Springer, Heidelberg (2016)
40. Ellis, G.: Top-down effects in the brain. Phys. Life Rev. **31**, 1–30 (2018)
41. Ellis, G.F.R.: Emergence in Solid State Physics and Biology (2020). arXiv:2004.13591
42. Ellis, G.F.R.: Does Physics Deny Free Will? Aeon Essay, Week 8th June 2020 (2020)
43. Ellis, G., Drossel, B.: How downwards causation occurs in digital computers. Found. Phys. **49**, 1253–1277 (2019). https://arxiv.org/pdf/1908.10186
44. Ellis, G., Drossel, B.: Emergence of time. Found. Phys. **50**, 161–190 (2020)
45. Ellis, G., Kopel, J.: The dynamical emergence of biology from physics. Front. Physiol. **9**, 1966 (2019)
46. Ellis, G.F.R., Meissner, K.A., Nicolai, H.: The physics of infinity. Nat. Phys. **14**, 770–772 (2018)
47. Ellis, G.F.R., Stoeger, W.R.: The evolution of our local cosmic domain: effective causal limits. Mon. Not. R. Astron. Soc. **398**, 1527–1536 (2009)
48. Epstude, K., Roese, N.J.: The functional theory of counterfactual thinking. Personal. Soc. Psychol. Rev. **12**, 168–192 (2008)
49. Farnsworth, K.D.: How organisms gained causal independence and how it might be quantified. Biology **7**, 38 (2018)
50. Richard, F.: The Character of Physical Law. MIT Press, Cambridge (2017)
51. Fink, M., Noble, D.: Noble model. Scholarpedia **3**(2), 1803 (2008)
52. Franklin, A., Knox, E.: Emergence without limits: the case of phonons. Stud. Hist. Philos. Mod. Phys. **64**, 68e78 (2018)
53. Ghirardi, G.: Sneaking a Look at God's Cards: Unraveling the Mysteries of Quantum Mechanics. Princeton University Press, Princeton (2007)
54. Gibb, S., Hendry, R.F., Lancaster, T. (eds.): The Routledge Handbook of Emergence. Routledge, Abingdon (2019)
55. Gillett, C.: Emergence, downward causation and its alternatives: critically surveying a foundational issue. In: The Routledge Handbook of Emergence, pp. 99–110. Routledge, Abingdon (2019)
56. Gorman, T.J., et al.: Field testing for cosmic ray soft errors in semiconductor memories. IBM J. Res. Dev. **40**, 41–50 (1996)
57. Green, S., Batterman, R.: Biology meets physics: reductionism and multi-scale modeling of morphogenesis. Stud. Hist. Philos. Sci. C **61**, 20–34 (2017)
58. Greene, B.: Until the End of Time: Mind, Matter, and Our Search for Meaning in an Evolving Universe. Knopf, New York (2020)
59. Guyton, A.C.: Basic Human Physiology: Normal Functions and Mechanisms of Disease. W B Saunders, Philadelphia (1977)
60. Hartmann, S.: Effective field theories, reductionism and scientific explanation. Stud. Hist. Philos. Sci. B **32**, 267–304 (2001)
61. Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W.: From molecular to modular cell biology. Nature **402**(6761), C47–C52 (1999)
62. Hawking, S.W., Ellis, G.F.R.: The Large Scale Structure of Spacetime. Cambridge University Press, Cambridge (1973)
63. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, Upper Saddle River (1994)
64. Hewitt, P.G.: Conceptual Physics. Pearson, London (2002)
65. Hoffmann, P.: Life's Ratchet: How Molecular Machines Extract Order from Chaos. Basic Books, New York (2012)
66. Hofmeyr, J.-H.: Basic biological anticipation. In: Poli, R. (ed.) Handbook of Anticipation. Springer, Cham (2017)
67. Hofmeyr, J.-H.: Causation, constructors and codes. Biosystems **164**, 121–127 (2018)

68. Hohwy, J., Kallestrup, J. (eds.): Being Reduced: New Essays on Reduction, Explanation, and Causation. Oxford University Press, Oxford (2008)
69. Hordijk, W.: Autocatalytic sets: from the origin of life to the economy. BioScience **63**, 877–881 (2013)
70. Hordijk, W., Steel, M.: Autocatalytic networks at the basis of life? Origin and organization. Life **8**, 62 (2018)
71. Hossenfelder, S.: The case for strong emergence. In: Aguirre, A., Foster, B., Merali, Z. (eds.) What is Fundamental?, pp. 85–94. Springer, Cham (2019)
72. Houghton, J.: Global Warming: The Complete Briefing. Cambridge University Press, Cambridge (2009)
73. Jain, A.K., Mao, J., Moidin Mohiuddin, K.: Artificial neural networks: a tutorial. Computer **29**, 31–44 (1996)
74. Jones, G., Bradshaw, D.: Resonance energy transfer: from fundamental theory to recent applications. Front. Phys. (2019). https://doi.org/10.3389/fphy.2019.00100
75. Junker Björn, H., Schreiber, F.: Analysis of Biological Networks. Wiley, Hoboken (2011)
76. Juarrero, A.: Dynamics in Action: Intentional Behavior as a Complex System. MIT Press, Cambridge (2002)
77. Kandel, E.R.: A new intellectual framework for psychiatry. Am. J. Psychiatry **155**, 457–469 (1998)
78. Kandel, E.R.: The molecular biology of memory storage: a dialogue between genes and synapses. Science **294**, 1030–1038 (2001)
79. Kandel, E.R.: The Age of Insight: The Quest to Understand the Unconscious in Art, Mind, and Brain, from Vienna 1900 to the Present. Penguin Random House, New York (2012)
80. Kandel, E., Schwartz, J.H., Jessell, T.M., Siegelbaum, S.A., Hudspeth, A.J.: Principles of Neural Science. McGraw Hill Professional, New York (2013)
81. Kim, J.: Mind in a Physical World. MIT Press, Boston (1998)
82. Kim, J.: Supervenient properties and micro-based properties: a reply to Noordhof. Proc. Aristot. Soc. **99**, 115–118 (1999)
83. Kurose James, F., Ross, K.W.: Computer Networking: A Top-Down Approach Featuring the Internet. Addison Wesley, Boston (2005)
84. Laughlin, R., Pines, D.: The theory of everything. Proc. Natl Acad. Sci. USA **97**, 28–31 (2000)
85. Loewenstein Werner, R.: The Touchstone of Life: Molecular Information, Cell Communication and the Foundations of Life. Oxford University Press, Oxford (1999)
86. Luisi, P.L.: Emergence in chemistry: chemistry as the embodiment of emergence. Found. Chem. **4**, 183–200 (2002)
87. Luu, T., Meißner, U.-G.: On the Topic of Emergence from an Effective Field Theory Perspective (2019). arXiv:1910.13770
88. MacCormick, J.: Nine Algorithms that Changed the Future. Princeton University Press, Princeton (2011)
89. Macdonald, G., Macdonald, C. (eds.): Emergence in Mind. Oxford University Press, Oxford (2010)
90. McQuarrie, D.A.: Quantum Chemistry. University Science Books, Sausalito (2008)
91. Maturana, H., Francisco, V.: Autopoiesis and Cognition: The Realization of the Living. Springer, Dordrecht (1980)
92. Mayr, E.: What Evolution is. Basic Books, New York (2001)
93. Menzies, P.: The causal efficacy of mental states. In: Physicalism and Mental Causation, pp. 195–224. Exeter: Imprint Academic, Charlottesville (2003)
94. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Science **298**, 824–827 (2002)
95. Montévil, M., Mossio, M.: Biological organisation as closure of constraints. J. Theor. Biol. **372**, 179–191 (2015)
96. Moreno, A., Mosseo, M.: Biological Autonomy: A Philosophical and Theoretical Enquiry. Springer, Dordrecht (2015)
97. Moss, L., Nicholson, D.J.: On nature and normativity: normativity, teleology, and mechanism in biological explanation. Stud. Hist. Philos. Biol. Biomed. Sci. **43**, 88–91 (2012)
98. Mossio, M.: Closure, causal. In: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., Yokota, H. (eds.) Encyclopedia of Systems Biology, pp. 415–418. Springer, New York (2013)
99. Mossio, M., Bich, L.: What makes biological organisation teleological? Synthese **194**, 1089–1114 (2017)

100. Mossio, M., Moreno, A.: Organisational closure in biological organisms. Hist. Philos. Life Sci. **32**, 269–288 (2010)
101. Mossio, M., Saborido, C., Moreno, A.: An organizational account of biological functions. Br. J. Philos. Sci. **60**, 813–841 (2009)
102. Murugan, J., Weltman, A., Ellis, G.F.R. (eds.): Foundations of Space and Time: Reflections on Quantum Gravity. Cambridge University Press, Cambridge (2012)
103. Noble, D.: Modeling the heart—from genes to cells to the whole organ. Science **295**, 1678–1682 (2002)
104. Noble, D.: The Music of Life: Biology Beyond Genes. Oxford University Press, Oxford (2008)
105. Noble, D.: A theory of biological relativity: no privileged level of causation. Interface Focus **2**, 55–64 (2012)
106. Noble, R., Noble, D.: Harnessing stochasticity: how do organisms make choices? Chaos Interdiscip. J. Nonlinear Sci. **28**, 106309 (2018)
107. Noble, R., Noble, D.: A-mergence of biological systems. In: Gibb, S., et al. (eds.) Handbook of Emergence, pp. 387–399. Routledge, London (2019)
108. Noble, R., Tasaki, K., Noble, P., Noble, D.: Biological Relativity requires circular causality but not symmetry of causation: so, where, what and when are the boundaries? Front. Physiol. **10**, 827 (2019)
109. Nurse, P.: Life, logic and information. Nature **454**, 424–426 (2008)
110. Oerter, R.: The Theory of Almost Everything: The Standard Model, the Unsung Triumph of Modern Physics. Penguin, New York (2005)
111. Papin, J.A., Reed, J.L., Palsson, B.O.: Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. Trends Biochem. Sci. **29**, 641–647 (2004)
112. Peacocke, A.: An Introduction to the Physical Chemistry of Biological Organization. Oxford University Press, Oxford (1990)
113. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, Cambridge (2009)
114. Pearl, J., Mackenzie, D.: The Book of Why: The New Science of Cause and Effect. Basic Books, New York (2018)
115. Penrose, R.: The Emperor's New Mind. Oxford University Press, New York (1990)
116. Percival, I.: Schrödinger's quantum cat. Nature **351**, 357 (1991)
117. Peter, P., Uzan, J.-P.: Primordial Cosmology. Oxford University Press, Oxford (2013)
118. Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P.J., Vaccarino, F.: Homological scaffolds of brain functional networks. J. R. Soc. Interface **11**, 20140873 (2014)
119. Pezzulo, G., Levin, M.: Top-down models in biology: explanation and control of complex living systems above the molecular level. J. R. Soc. Interface **13**, 20160555 (2016)
120. Randall, D., Burggren, W., French, K.: Eckert Animal Physiology: Mechanisms and Adaptations. W H Freeman, New York (2002)
121. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.-L.: Hierarchical organization of modularity in metabolic networks. Science **297**, 1551–1555 (2002)
122. Rhoades, R., Pflanzer, R.: Human Physiology. Saunders College Publishing, Fort Worth (1989)
123. Ruiz-Mirazo, K., Moreno, A.: Autonomy in evolution: from minimal to complex life. Synthese **185**, 21–52 (2012)
124. Schweber, S.S.: Physics, community, and the crisis in physical theory. Phys. Today **46**, 34–34 (1993)
125. Sciama, D.W.: The Unity of the Universe. Courier Corporation, New York (2012)
126. Simon, H.A.: The Sciences of the Artificial. MIT Press, Cambridge (2019)
127. Simon, S.H.: The Oxford Solid State Basics. OUP, Oxford (2013)
128. Sloan, D., Batista, R., Hicks, M., Davies, R. (eds.): Fine Tuning in the Physical Universe. Cambridge University Press, Cambridge (2020)
129. Smith, J.M., Szathmary, E.: The Major Transitions in Evolution. Oxford University Press, Oxford (1997)
130. Tabaczek, M.: The metaphysics of downward causation: rediscovering the formal cause. Zygon **48**, 380–404 (2013)
131. Taleb, N.N.: The Black Swan: The Impact of the Highly Improbable. Penguin, London (2010)
132. Tanenbaum, A.S.: Structured Computer Organisation, 5th edn. Prentice Hall, Englewood Cliffs (2006)
133. Tegmark, M., Bostrom, N.: Is a doomsday catastrophe likely? Nature **438**, 754 (2005)

134. Thompson, C.: Coders: The Making of a New Tribe and the Remaking of the World. Penguin, New York (2019)
135. Thompson, E.: Mind in Life: Biology, Phenomenology, and the Sciences of Mind. Harvard University Press, Cambridge (2007)
136. Uzan, J.-P.: The fundamental constants and their variation: observational and theoretical status. Rev. Mod. Phys. **75**, 403 (2003)
137. Vecchi, D., Miquel, P.-A., Hernández, I.: From biological determination to entangled causation. Acta biotheor. **67**, 19–46 (2019)
138. Villalobos, M., Dewhurst, J.: Enactive autonomy in computational systems. Synthese **195**, 1891–1908 (2018)
139. Wagner, A.: Arrival of the Fittest: Solving Evolution's Greatest Puzzle. Penguin, New York (2014)
140. Watson, J.D., et al.: Molecular Biology of the Gene. Pearson, Upper Saddle River (2013)
141. Weinberg, S.: Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature. Penguin Random House, New York (1994)
142. Wuchty, S., Ravasz, E., Barabási, A.-L.: The architecture of biological networks. In: Complex Systems Science in Biomedicine, pp. 165–181. Springer, Boston (2006)
143. Ziegler, J.F., Lanford, W.A.: Effect of cosmic rays on computer memories. Science **206**, 776–788 (1979)

# Physics, Determinism, and the Brain

George F R Ellis

September 8, 2020

**Abstract**

*This paper responds to claims that causal closure of the underlying microphysics determines brain outcomes as a matter of principle, even if we cannot hope to ever carry out the needed calculations in practice. Following two papers of mine where I claim firstly that downward causation enables genuine causal powers to occur at higher emergent levels in biology (and hence in the brain) [Ellis 2020a], and that secondly causal closure is in reality an interlevel affair involving even social levels [Ellis 2020b], Carlo Rovelli has engaged with me in a dialogue where he forcefully restates the reductionist position that microphysics alone determines all, specifically the functioning of the brain. Here I respond to that claim in depth, claiming that if one firstly takes into account the difference between synchronic and diachronic emergence, and secondly takes seriously the well established nature of biology in general and neuroscience in particular, my position is indeed correct.*

## Contents

# 1 Physics, The Brain, and Predictability

After I published articles in *Foundation of Physics* on emergence in solid state physics and biology [Ellis 2020a] and on causal closure in biology and engineering contexts [Ellis 2020b], I have had an interesting exchange with Carlo Rovelli questioning my arguments. This paper responds as regards brain function. In Section 1.1 I will give an edited version of that interchange. In section 1.2 I outline my response. The key point is that while his argument does indeed apply in the case of synchronic emergence - emergence at one point in time - it does not apply in the case of diachronic emergence, that is, emergence as it unfolds over time. The paper is summarised in Section 1.3.

## 1.1 A dialogue

Here is a summary of that dialogue.

**CR** - You list a large number of situations in which we understand phenomena, cause and effects, and we make sense of reality, using high level concepts. I think this is great.

- You point out that nobody is able to account for these phenomena in terms of microphysics. I think this is true.

- You emphasize the fact that we *need* these high level notions to understand the world. I think this is important and is an observation which is underappreciated by many. I agree.

- You point out that to some extent similar points have been made by a number of biologists, solid state physicists, etcetera. I definitely think this is true.

⇒ Then, you jump to a conclusion about microphysics, which does not follow from any of the above. And I think that the large majority of physicists also think that it does not follow. Given a phenomenon (condensed matter, biology, anything else), happening in a specific situation in a specific context, it is the belief of the near majority of physicists today that the initial microphysical data (which we may not know), determines uniquely the probability distribution of the later outcomes.

How does this square with all your examples, and why there is no contradiction between all your examples and the points you make, and this? The reason is that in all your examples the high-level cause is also a microphysics configuration. For instance, a crystal configuration determines the motion of the electrons, smoking causes cancer, my getting excited changes the motion of the electrons in my body. All this is true of course. But a crystal configuration, smoking, my getting excited are also microphysical states of affairs. Once we fold them in the microphysics, there is no reason whatsoever for the microphysics not to be sufficient to determine what happens. There isn't something "in addition" to the microphysics. It is only that the microphysics is far more complicated than what we can manage directly in our calculations of simple minded understanding.

Where is our disagreement? It is not big. But I think it is crucial, and I think it can be summarized in the case of the effect of Jupiter and the drop of water on the beach of Marseille affected by its gravitational pull. Here we can take the microphysics to be the Newtonian gravitation of all the small grains of matter in the Solar system, governed by classical mechanics with Newtonian interactions and short scale pressure. This is a good level of approximation. Can we compute the motion of the drop of water using the Newton equations and all the forces? Of course we cannot. There are far too many grains of matter in the Solar system. But we can go to a higher level description, where we ignore a huge amount of details and we represent everything in terms of a few planets in their orbits. "Planet" is not a notion in the microphysics. It is a simple calculation to account for the tidal forces due to the Moon and the Sun and to estimate the small correction due to Jupiter. And we find that the drop of water last Tuesday came a bit higher than expected because of Jupiter.

Here is the point: there is nothing I see, in all the examples you mention that distinguishes them from this version of"top-down-causation": The high-level effect of Jupiter affects the motion of a drop of water. I do not see any difference between this and the crystalline structure affecting motion of electrons in condensed matter, or a biological molecule reacting to an evolutionary pressure, or my excitement affecting the motion of an electron.

**GE** You are then agreeing that high level causation is real?

**CR** "Causal" mean all sort of different things for all sort of different people. I am not being pedantic, I think it is one of the cruxes of the matter. Instinctively, I am with Russell that notices that there are no "causes" in physics: there are just regularities expressed by laws. But of course I am aware (with Cartwright ) that we do use causes heavily and effectively . I think that "causes" make sense with respect to an agent that can act, and that the agency of the agent is ultimately rooted in entropy growth. If "causal" concepts are understood in this high level and sort of non-fundamental manner, then suddenly your entire project makes sense to me I do not think that anybody would deny that smoking causes cancer. Therefore, yes, I agree that high level causation is real. I think that "smoking" denotes a large

ensemble of microphysical configurations, all of which (actually, in this context: many of which), evolving according to microphysical laws without any reference to higher order notions, evolve into later microphysical configurations that belong to the ensemble that we call "a person with cancer".

**GE** "The reason is that in all your examples the high-level cause is also a microphysics configuration" Yes indeed This sounds very close to Denis Noble's Principle of Biological relativity, extended to include the physics level, which is what I propose in my two papers [Ellis 2020a] [Ellis 2020b]. I certainly state that the micro levels are effective, as you do - there is no disagreement there. Maybe then we are not so far apart. Question: would you agree that you can invert that statement to get: "in all these examples the microphysics configuration is also a high-level cause"?

**CR** Here we are back to the ambiguity of the notion of "cause". I do not know what you mean. The microphysical configurations belonging to "smoking" do evolve into those belonging to "cancer", while the microphysical configurations "almost everything the same but not smoking", do not. If this is what we mean by "cause", I agree. If there is some other meaning of "cause", I sincerely do not understand it. I think that this is what we more more less mean usually, (what else means "smoking causes cancer" if not that by not not smoking we can get less cancer?), then I am happy to use "cause" here.

**GE** This is the key phrase: "Once we fold them in the microphysics". This sounds very much like what I am saying. The higher level situation is setting the context for the microphysics to act, and hence shaping the specific outcomes that occur through the microphysics If that is what you mean, then we agree! Otherwise what does that phrase mean?

**CR** No, here we disagree, I think. Because there isn't the microphysics, and then something else that sets the context. The macro-physics is just a way of talking about microphysical states. "Smoking" is not something that that is added to the microphysics: it is that one possible configuration of the microphysics. More precisely: an ensemble of many possible configurations of the microphysics. This is the main point: you cannot have the same microphysics with different "contexts". Different "contexts" always require different microphysics. It is the key question: the question I asked you when we started talking about that at a conference somewhere: suppose there are two Chess games on two different planets: everything looks the same but the rules of Chess are a bit different on the two planets (in one you can castle after having moved the king, in the other you cannot). Here a very high order difference (different rules of Chess) make the evolution different in the two planets. BUT (here is the point): could the rules in the two planets be exactly the same if the microphysics was the same? NO, of course, because different rules of the games means different memories stored in the players brains, hence different synapses, hence different physics. To get a difference, you need different microphysical configurations. You cannot have a high level source of difference without having a microphysics level difference that achieves the same result.

**GE** To get full clarity, please clarify two questions:

1. Consider one person and the issue of smoking; or (for your neuroscientist friends) one person's brain states when they play chess. Are you envisaging the total cause of their future behaviour being the microphysics configuration of their own body/brain,

or of some larger ensemble of particles? If so, what set? What set does "causal closure"", in your terms, refer to?

**CR** A larger ensemble of course, because what happens to a person depends on plenty of exterior influences.

**GE** The second question.

2. Do you consider "folding them in the microphysics" (first reply above) as an ongoing process that is taking place all the time, or not?

**CR** "Folding then in the microphysics" means recognizing what they are, once translated into microphysics. They are ensembles of microphysical states. If you do this at the start of a time interval and if you include enough degrees of freedom to account for anything that matters, then the microscopic evolution does the job, whether or not we can compute. If this was NOT the case, we would have found cases whether these evolution laws fail. When people say that a theory is "causally closed" they mean that the initial conditions determine the following (of course: in principle. In practice we cannot do the calculations, nor know all relevant initial conditions.)

**GE** To be clear, is it a process that takes place once at the start of some interaction (the brain starts considering a chess problem) and then the physics by itself determines all outcomes; or is it an ongoing process that is taking place all the time ?

**CR** It is not a process. I can in principle describe the set of events without any reference to high level concepts, or I can describe the set of events using high level concepts. Both work. The difference is that one may be unmanageable and the other may be manageable. We have these different levels of description everywhere in life: I can think of my trip from Verona to Marseille accounting for the instantaneous changes of velocities, or I can represent it just as average constant velocity with some snack-pauses. The first is unmanageable, the second uses high level concepts, is more useful. But it has less information than the first, not more.

**GE** Finally, of course I believe the outcomes lies in the space of possibilities allowed by the microphysics. It could not be otherwise, and the marvel is that that set of laws is able to produce such complexity. There is no way I am underestimating what the physics allows. In the case of biology, I see it as happening because the physics allows the existence of the Platonic space of all possible proteins discussed by Andreas Wagner in his marvelous book *Arrival of the Fittest* [Wagner 2014].

**CR** Yes, this is what I meant. I am glad we agree here. This space of possibility is immense, and too hard to explore theoretically even if we know the microscopic laws and the fact these laws are not violated and they determine (probabilistic) evolution unequivocally. Still, we carve it out for our understanding by recognising high level patterns and using them. But then why do you need the microphysics to be affected by something outside it? You do not need it. And we have zero evidence for anything like this. The autonomy of the higher level logic that you keep citing is no argument against the autonomy of the microphysics.

But you seem to mean more. Do you mean that they select which dynamical histories are realised and which not? That is: which initial conditions are allowed and which not? Or they select which quantum outcomes become actual, over and above the quantum probability amplitudes? Or what else?

It seems to me you confuse the richness of the tools of a physicist, and the complexity of reality, with a statement about lack of causal closure of the microphysics, that does not follow.

Finally, a last response:

**CR** Today the burden of the proof is not on this side. Is on the opposite side. Because:

**(i)** There is no single phenomenon in the world where microphysics has been proven wrong (in its domain of validity of velocity, energy, size...).

**(ii)** By induction and Occam razor, is a good assumption that in its domain validity it holds.

**(iii)** There are phenomena too complex to calculate explicitly with microphysics. These provide no evidence against (ii), they only testify to our limited tools.

There is much more, but I will not repeat it all. Rather I summarise points of agreement, of misunderstanding, and of disagreement, and then take up that challenge.

**Points of agreement**

- All is based in immutable lower level physical laws, unaffected by context
- These laws allow immensely complex outcomes when applied to very complex microstates, such as those that underlie a human brain
- Testable higher level laws correctly express dynamics at higher levels
- They make processes at that level transparent in a way that is completely hidden when one traces that same dynamics at the lower levels

**Point of misunderstanding**

- **Downward causation**

  [**CR**] From your examples it does not follow that our current elementary theories, such GR and the SM, have free parameters that are controlled by something else that we understand. That would be a wild speculation unsupported by anything.

  [**GE**] The theories themselves of course do not have such parameters, and I have never claimed that they do. But Lagrangians used in specific contexts do.

  Those theories *per se* have generic Lagrangians that apply to anything at all, and so say nothing detailed about anything specific. They do not by themselves determine outcomes in biology or engineering. A particular context determines the details of the terms in the Lagrangian, and that happens in a contextually dependent way: after all, in a particular case it represents a specific context. Once the Lagrangian has been determined at time $t_1$ then the next emergent step is indeed fully determined purely at the micro physical level, as Carlo claims. But macro conditions can then change parameters in the Lagrangian. That is where the downward causal effects come in. For specific examples, the case of transistors in digital computers is discussed in [Ellis and Drossel 2019], and voltage gated ion channels in the brain in [Ellis and Kopel 2019].

  The key issue is whether downward causation is real, having real causal powers. I argue that it is; and that this kind of causation does not required any compromising of the underlying physics. It works by changing constraints [Juarrero 2002].

**Points of disagreement**

**CR** **(i)** In principle, we could entirely deduce what happens even ignoring the high-level concepts

**(ii)** In practice, we use high-level concepts

**(iii)** In addition, we get a better grasp, a better sense of "understanding", a better control, in terms of higher level notions.

Hence: high-level notions are far more relevant for comprehension of the world. I think that (ii) and (iii) are very important.

(**iv**) But if you add the negation of (i), you get everybody disagreeing (me included) and the message about (ii) and (iii) does not go through.

**GE** **1** Causation in biology is an interlevel concept [Mossio 2013]. Physics underlies this but does not by itself give causal closure. What Carlo calls "causal closure of physics" is in fact the statement that at its own level, it is a well-posed theory: a completely different affair.

**2** Carlo's statement can be defended in the synchronic case (at a time), but is not always true for individual brains in the diachronic case (unfolding over time).

**3** Statistics don't cut it. Individual events occur. We have to explain why specific individual brain events occur, for example leading to the specific words in Carlo's emails. Specific events in individual lives occur and need to be accounted for.

**4** Microphysics enables this but does not determine the outcomes. The basic physics interactions of course enable all this to happen: they allow incredible complexity to emerge. Higher level organising principles such as Darwin's theory of evolution then come into play. That then changes the macro level context and hence the micro level context. This downward process [Campbell 1974] relies on concepts such as 'living' that simply cannot be represented at the microlevel, but determine outcomes.

**5** The reductionist physics view is based in a linear view of causation. Central to the way biology works are the closely related ideas of *self cause* and *circular causation.*

## 1.2 A response

In the rest of this paper, I give a full response to Carlo's arguments in the case of the brain, based in the nature of causation in biology. It rests on three things. First, taking seriously the nature of biology in general [Campbell and Reece 2008] and neuroscience [Kandel 2012] [Kandel *et al* 2013] in particular, demanding that whatever overall theory we propose must respect that nature. Second, requiring that individual events and outcomes are what need to be accounted for, not just statistics. Third, noting the key difference between synchronic and diachronic emergence, which we did not make in our email interchange. The answer is very different in these two cases.

**Synchronic and diachronic emergence** Carlo's argument - the microstate uniquely determines macro level outcomes - is correct when we consider synchronic emergence. That is what a lot of neuroscience is about. It is not valid however when one considers diachronic emergence. The issue here is one of timescales.

*Synchronic emergence* is when the timescale $\delta t := t_b - t_a$ of the considered microdynamic outcomes is very short relative to the timescale $\delta T$ of change of structures at the micro scale: $\delta T \gg \delta t$. It is the issue of emergent dynamics when parameters are constant

and constraints unchanging. In the case of the brain this would for example be the flow of electrons in axons leading to mental outcomes at that time, with this micro structure taken as unchanging. Electrons and ions flow in a given set of neural connections.

*Diachronic emergence* is when the timescale of micro dynamic outcomes considered $\delta t$ is of the same order or larger than the timescale $\delta T$ of change of structure at the micro scale: $\delta T \leq \delta t$, so microdynamics contexts alters significantly during this time. It is the case when parameters or constraints change because of interactions that are taking place. In the case of the brain this would for example be when something new is learned so that strengths of neural connections are altered.

**Consider first a single brain** Dynamic outcomes at the molecular scale are due to the specific structures at the cellular scale, neural connectivity for example, and the way that they in turn constrain electron and ion activity. Three points arise.

- First, ***the brain is an open system***. It is not possible for the initial physical state to determine later states because of the flood of data incoming all the time. The last round of microlevel data does not determine the initial data that applies at the next round of synchronic emergence. The brain has evolved a set of mechanisms that enable it to cope with the stream of new data flowing in all the time by perceiving its meaning, predicting futures, and planning how to respond. This is what determines outcomes rather than evolution from the last round of initial data.

- Second, ***the brain is a plastic brain*** that changes over time as it learns. Neural connections are altered as learning takes place in response to the incoming stream of data. This change in constraints alters future patterns of electron and ion flows. This learning involves higher level variables and understandings such as "A global Coronavirus pandemic is taking place", that cannot be characterised at lower levels and cannot be predicted from the initial brain microdata.

- Third, ***there is a great deal of stochasticity at the molecular level*** that breaks the ideal of Laplacian determinism at that level. Molecular machines have been evolved that take advantage of that stochasticity to extract order from chaos. From a higher level perspective, this stochasticity enables organisms to select lower level outcomes that are advantageous to higher level needs. From a systems perspective, this enables higher level organising principles such as existence of dynamical system basins of attraction to determine outcomes.

This argument applies to all biology, as all biological systems are by their nature open systems [Peacocke 1989]. The initial physics data for any organism by itself cannot in principle determine specific later outcomes because of this openness.

The fundamental physical laws are not altered or overwritten when this happens; rather the context in which they operate - for example opening or closing of ion channels in axons - determine what the specific outcomes of generic physical laws will be as alter configuration. From a physics viewpoint this is represented by time dependent constraint terms or potentials in the underlying Hamiltonian [Ellis and Kopel 2019].

**The whole universe gambit** The ultimate physicalist response is "Yes the brain may be an open system but the whole universe is not; and the brain is just part of the universe, which is causally complete. Hence brain dynamics is controlled by the microphysics alone when one takes this into account, because it determines all the incoming information to the brain". However this argument fails for the following reasons:

- Firstly there is ***irreducible quantum uncertainty*** in outcomes, which implies the lower physics levels are in fact not causally complete. This can get amplified to macroscales by mechanisms that change mental outcomes, such as altered gene expression due to damage by high energy photons.

- Secondly, this downward process - inflow of outside information to individual brains - does not uniquely determine brain how microstructures change through memory processes because of ***multiple realisability***. But such uniqueness is required to sustain a claim that the causal closedness of microphysics determines specific brain outcomes over time.

- Thirdly, ***chaotic dynamics*** associated with strange attractors occurs, which means the emergent dynamics of weather patterns is not in fact predictable even in principle over sufficiently long timescales. This affects decisions such as whether to take an umbrella when going to the shops or not.

- Fourthly, ***microbiome dynamics*** in the external world affects brain outcomes in unpredictable ways, for example when a global pandemic occurs

- Fifthly, this all takes place in a social context where ***social interactions*** take place between many brains, each of which is itself an open system. Irreducible uncertainty influences such contexts due to the real butterfly effect (weather) and the impossibility, due to the molecular storm, of predicting specific microbiome mutations that occur (e.g. COVID-19), leading to social policy decisions, that are high level variables influencing macro level brain states. The outcomes then influence details of synaptic connections and hence shape future electron and ion flows.

This downward causation from the social/psychological level to action potential spike chains and synapse activation is essential to the specific outcomes that occur at the physical level of electron and ion flows in individual brains. Causal closure only follows when we include those higher level variables in the dynamics.

## 1.3 The argument that follows

Section 2 sets the scene by discussing the foundations for what follows, in particular the fact that life is an ongoing adaptive process. In the following sections I discuss the key issues that support my view.

Firstly, an individual brain is an open system, and has been adapted to handle the problems this represents in successfully navigating the world (Section 3). This rather than the initial brain micro data determines outcomes.

Secondly, the brain learns: it is plastic at both macro and micro levels, which continually changes the context within which the lower level physics operates (Section 4).

Third, the kind of Laplacian view of determinism underlying Carlo's position is broken at the molecular level because of the huge degree of stochasticity that happens at that level (Section 5). Biological processes - such as Darwinian evolution, action choices, and the brain pursuing a line of logical argumentation - are what in fact determine outcomes, taking advantage of that stochasticity. Biological causation occurs selects preferred outcomes from the molecular storm, and the brain selects from action options.

In section 6 I counter the whole universe gambit by claiming that this will not work because of quantum wave function collapse, macro level chaotic dynamics, multiple realisability of macro brain states, and unpredictable microbiome interactions that affect brain dynamics both directly and via their social outcomes.

Section 7 consider how higher level organising principles - the effective laws that operate at higher levels - are in fact what shapes outcomes. This is what enables causal closure - an interlevel affair - in practice. I also comment on the issue of freewill (Section 7.2).

# 2 Foundations

As stated above, the premise of this paper is that when relating physics to life, one should take seriously the nature of biology as well as that of physics. I assume the standard underlying microphysics for everyday life, based in the Lagrangian for electrons, protons, and nuclei, see [Laughlin and Pines 2000] and [Bishop 2005]. This section sets the foundation for what follows by discussing the nature of biology and of causation.

Section 2.1 discusses the basic nature of biology. Section 2.2 outlines the biological hierarchy of structure and function. Section 2.3 discusses the nature of Effective Theories at each emergent level **L**. Section 2.4 discusses the equal validity of each level in causal terms. Section 2.5 discusses the various types of downward causation, and Aristotle's four types of causes as well as Tinbergen's 'Why' questions. Section 2.6 discusses the important issue of multiple realisability of higher level structure and function at lower levels. Finally Section 2.7 discusses the key role of Higher Level Organising Principles.

## 2.1 The basic nature of biology

All life [Campbell and Reece 2008] is based in the interplay between structure (that is, physiology [Hall 2016] [Rhoades and Pflanzer 1989]) and function. For good functional, developmental, and evolutionary reasons, it is composed (**Table 1**:§2.2) of ***Adaptive Modular Hierarchical Structures*** [Simon 2019] [Booch 2006] based in the underlying physics. It comes into being via the interaction between evolutionary and developmental (Evo-Devo) processes [Carroll 2005] [Carroll 2008], and has three key aspects. [1]

**1. Teleonomy: function/purpose** Life has a teleonomic nature, where Jacques Monod defines teleonomy as the characteristic of being "endowed with a purpose or project" ([Monod 1971]:9) He points out the extreme efficiency of the teleonomic apparatus in accomplishing the preservation and reproduction of the structure. As summarised by Nobel Prizewinner Leland Hartwell and colleagues [Hartwell *et al* 1999],

> "*Although living systems obey the laws of physics and chemistry, the notion of function or purpose differentiates biology from other natural sciences. Organisms exist to reproduce, whereas, outside religious belief, rocks and stars have no purpose. Selection for function has produced the living cell, with a unique set of properties that distinguish it from inanimate systems of interacting molecules. Cells exist far from thermal equilibrium by harvesting energy from their environment. They are composed of thousands of different types of molecule. They contain information for their survival and reproduction, in the form of their DNA*".

Function and purpose emerge at the cell level. Francois Jacob says [Jacob 1974][2]

> *"At each level of organisation novelties appear in both properties and logic. To reproduce is not within the power of any single molecule by itself. This faculty*

[1] An excellent introduction to the relevant mechanisms is given in [Noble 2016].
[2] Quoted in [Peacocke 1989]:275.

*appears only within the power of the simplest integron*[3] *deserving to be called a living organism, that is, the cell. But thereafter the rules of the game change. At the higher level integron, the cell population, natural selection imposes new constraints and offers new possibilities. In this way, and without ceasing to obey the principles that govern inanimate systems, living systems become subject to phenomena that have no meaning at the lower level. Biology can neither be reduced to physics, nor do without it.*"

**2. Life is a process** Being alive is not a physical thing made of any specific elements. It is a *process* that occurs at macro levels, in an interconnected way. In the case of human beings it involves all levels[4] from **L4** (the cellular level) to Level **L6** (individual human beings), allowing causal closure [Mossio 2013] [Mossio and Moreno 2010] and hence self-causation [Juarrero 2002] [Murphy and Brown 2007].

> **Life is an ongoing adaptive process *involving metabolism, homeostasis, defence, and learning in the short term, reproduction, growth, and development in the medium term, and evolution in the long term. It uses energy, disposes of waste heat and products, and uses contextual information to attain its purposes.***

The claim I make is that this process of living has causal power, making things happen in an ongoing way. High level processes take place via an interlevel dialogue between levels [Noble 2008], higher levels continually altering the context of the underlying physical levels in order to carry out these functions [Ellis and Kopel 2019]. Yes of course the resulting physical processes can be traced out at the physics level. But my claim will be that biological imperatives [Campbell and Reece 2008] enabled by physiological systems [Rhoades and Pflanzer 1989] [Hall 2016] shape what happens. Evolutionary processes [Mayr 2001] [Carroll 2008] have enabled this synergy to occur [Noble 2016].

**3. Basic biological needs and functions** In the case of animal life,[5] the basic biological functions are,

**B1:** Metabolism (acquiring energy and matter, getting rid of waste),
**B2:** Homeostasis and defence,
**B3:** Reproduction and subsequent development,
**B4:** Mobility and the ability to act,
**B5:** Information acquisition and processing.

They serve as attractors when variation takes places ([Ginsburg and Jablonka 2019]:245). They are the higher level organising principles that evolution discovers and then embodies in hierarchically structured physiological systems, where the macro functions are supported at the micro level by metabolic networks, gene regulatory networks, and cell signalling networks, selected from an abstract space of possibilities and realised through specific proteins [Wagner 2014]. Information is central to what happens [Nurse 2008] [Davies 2019].

These principles cannot be described or identified at the underlying microphysical levels not just because the relevant variables are not available at that level, but because their multiple realisability at lower levels means they do not correspond to specific patterns of interactions at the ion and electron level. They correspond to a whole equivalence class of such patterns of interactions (Section 2.6).

[3] An 'Integron' is each of the units in a hierarchy of discontinuous units formed by integration of sub-units of the level below [Jacob 1974]:302.

[4] See **Table 1**, Section 2.2.

[5] Other forms of life share **B1-3.**

**4. Interaction networks** These processes are realised by means of immensely complex *interaction networks* at the molecular level [Buchanan *et al* 2010] [Junker & Schreiber 2011]:

**N1**: Metabolic Networks ([Wagner 2014] §3) [Noble 2016]

**N2:** Gene Regulatory Networks ([Wagner 2014] §5)

**N3:** Signalling Networks [Junker & Schreiber 2011] [Buchanan *et al* 2010]

**N4:** Protein Interaction Networks [Junker & Schreiber 2011]

based in very complex molecular interactions [Berridge 2014] and with higher level design principles shaping their structure [Alon 2006], and at the cellular level,

**N5:** Neural Networks [Kandel *et al* 2013] [Churchland and Sejnowski 2016]

These networks compute in the sense of ([Churchland and Sejnowski 2016]:69-74)

**5. Branching causal logic** In order to meet these needs, the dynamics followed at each level of biological hierarchies is based on contextually informed dynamical branching $L$ that support the functions $\alpha$ of a trait $T$ in a specific environmental context $E$ [Ellis and Kopel 2019]. Thus biological dynamics can be functionally-directed rather than driven by inevitability or chance:

$$\texttt{Biological dynamics tends to further the function } \alpha \texttt{ of a trait } T \texttt{ through contextually informed branching dynamics } L \quad (1)$$

where the dynamics $L$ in its simplest form is branching logic of the form [Hoffmann 2012]

$$\texttt{L: given context } C, \quad \texttt{IF } T(\mathbf{X}) \texttt{ THEN } F1(\mathbf{Y}), \quad \texttt{ELSE } F2(\mathbf{Z}) \quad (2)$$

(a default unstated "ELSE" is always to leave the status quo). Here $\mathbf{X}$ is a contextual variable which can have many dimensions, $\mathbf{Y}$ and $\mathbf{Z}$ are variables that may be the same variables as $\mathbf{X}$ or not. $T(\mathbf{X})$ is the truth value of arbitrary evaluative statements depending on $\mathbf{X}$. It can be any combination of Boolean logical operations (NOT, AND, OR, NOR, etc.) and mathematical operations, while $F1(\mathbf{Y})$ and $F2(\mathbf{Z})$ are outcomes tending to further the function $\alpha$. Thus they might be the homeostatic response "If blood sugar levels are too high, release insulin", or the conscious dynamic "If the weather forecast says it will rain, take an umbrella". At the molecular level, these operations are based in the lock and key molecular recognition mechanism ([Noble 2016]:71), [Berridge 2014]. This mechanism is how information [Nurse 2008] [Davies 2019] gets to shape physical outcomes.

**6. Brain Function** The human brain supports all these activities by a series of higher level processes and functions. These are [Purves *et al* 2008] [Gray and Bjorklund 2018]

**BR1:** Sensation, perception, classification

**BR2:** Prediction, planning, making decisions, and action

**BR3:** Experimenting, learning, and remembering

**BR4:** Experiencing and responding to emotions

**BR5:** Interacting socially, communicating by symbols and language

**BR6:** Metacognition, analysis, and reflection, 'off-line' exploration of possibilities.

It does so via its complex adaptive modular hierarchical structure [Kandel *et al* 2013] [Scott 2002]. Brains compute [Marr 2010] [Churchland and Sejnowski 2016], but they are not digital computers [Piccinini and Shagrir 2014].

## 2.2 The hierarchy

The framework for the following is the hierarchy of structure and function for the biological sciences shown in (**Table 1**), based in the underlying physics.

| | Biology Levels | Processes |
|---|---|---|
| Level 8 (**L8**) | Environment | Ecological, environmental processes |
| Level 7 (**L7**) | Society | Social processes |
| Level 6 (**L6**) | Individuals | Psychological processes, actions |
| Level 5 (**L5**) | Physiological systems | Homeostasis, emergent functions |
| Level 4 (**L4**) | Cells | Basic processes of life |
| Level 3 (**L3**) | Biomolecules | Gene regulation, metabolism |
| Level 2 (**L2**) | Atom, ion, electron Physics | Atomic, ionic, electron interactions |
| Level 1 (**L1**) | Particle and Nuclear Physics | Quark, lepton interactions |

**Table 1**: *The hierarchy of structure for biology (left) and corresponding processes (right).* **L2** *is the relevant physics level of emergence,* **L4** *the fundamental biological level, made possible by* **L3** *(in particular proteins, RNA, DNA), in turn made possible by* **L2** and so **L1**.

The first level where the processes of life occur is **L4**, the level of cells. At level **L6** one finds the integrated processes of an individual organism. At level **L7** one finds sociology, economics, politics, and legal systems.

## 2.3 Effective theories

I am assuming that each of these levels exists as a matter of fact - they exist ontologically. The key issue is, if we propose a specific level $\mathbf{L}$ exists ontologically, there should be a valid Effective Theory $\mathbf{ET_L}$ applicable at that level which characterizes that level. 'Valid' means it either makes testable predictions that have been confirmed, or at least characterizes the variables that would enter such a relation.[6] Here following [Ellis 2020a] and [Ellis 2020b], one can characterise an Effective Theory $\mathbf{ET_L}(a_\mathbf{L})$ valid at some level $\mathbf{L}$ as follows:

> *An* **Effective Theory** $\mathbf{ET_L}(a_\mathbf{L})$ *at an emergent level* $\boldsymbol{L}$ *is a reliable relation between initial conditions described by effective variables* $v_{\boldsymbol{L}} \in \boldsymbol{L}$ *and outcomes* $o_{\boldsymbol{L}} \in \boldsymbol{L}$*:*
>
> $$\mathbf{ET_L}(a_\mathbf{L}) : v_\mathbf{L} \in \mathbf{L} \to \mathbf{ET_L}(a_\mathbf{L})[v_\mathbf{L}] = o_\mathbf{L} \in \mathbf{L} \tag{3}$$
>
> *where* $a_{\boldsymbol{L}}$ *are parameters of the relation, and* $\boldsymbol{ET_L}(a_{\boldsymbol{L}})$ *may be an exact or statistical law. The parameters* $a_{\boldsymbol{L}}$ *may be vectorial or tensorial*

Thus I will define a meaningful level to exist if there is such a relation. Determining that relation is in effect epistemology, but what it indicates is the underlying ontology.

The effective theory $\mathbf{ET_L}(a_\mathbf{L})$ is **well posed** if for specific choices of the parameters $a_\mathbf{L}$ it provides a unique mapping (3) from $v_\mathbf{L}$ to $o_\mathbf{L}$. This is the concept one should use instead or referring to the theory as being causally complete. That is a misnomer because firstly, the idea of causality does not apply to the physics laws *per se* (although effective theories do), and secondly because causal completion - the set of conditions that actually determine what outcomes will occur in real-world contexts - is always an interlevel affair, no single level $\mathbf{L}$ by itself is causally complete (Section 6.6). Effective Theories represent verifiable patterns of causation at the relevant level, not causal closure [Ellis 2020b].

[6] The cautionary note reflects the difficulty in establishing reliable relations at levels **L6**-**L8**. The theories may have to be described in terms of propensities rather than mathematical laws. They are nevertheless well established fields of study, for example [Gray and Bjorklund 2018] at Level **L6**, [Berger 1963] at Level **L7**, and [Houghton 2009] at Level **L8**.

**Effective theory examples** It is useful to give some examples of effective theories at different levels. It is my contention, in agreement with [Noble 2012] [Noble 2016], that real causal processes are going on at each of these levels, even though this is enabled by underlying levels, including the physics ones. The relevant effective theories are more than just useful descriptions of high level processes. In all but the last two cases this is demonstrated by the fact that evolution has selected genomes that result in them happening. Their causal effectiveness is a driver of evolutionary selection.

1. **Gene regulation** The kind of gene regulatory processes discovered by Jacob and Monod [Jacob and Monod 1961] [Monod 1971] represent real causal processes at the cellular level (they require the relevant molecular processes, but can only take place in the context of a functioning cell [Hofmeyer 2018]). Their importance is that they underlie the Evo-Devo processes discussed in [Carroll 2005] [Carroll 2008].

2. **Action potential propagation** Brain processes are supported at the micro level by propagation of action potential spikes according to the Hodgkin-Huxley Equations [Hodgkin and Huxley 1952]. This is an emergent phenomenon that cannot be deduced from the underlying physics *per se* because they involve constants that are not fundamental physical constants. [Woodward 2018] defends the view that the explanation the equations provide are causal in the interventionist sense.

3. **The brain** The way the brain works overall [Kandel 2012] [Gray and Bjorklund 2018] is based in the underlying neuroscience [Kandel *et al* 2013]. It has been arrived at by an evolutionary process based in the advantages its specific functioning provides. Two key issues are the ability to function under uncertainty [Clark 2013] [Clark 2016] [Hohwy 2013] [Hohwy 2016] and the existence of a symbolic ability [Deacon 1997] that allows language, culture, and technology to arise [Ginsburg and Jablonka 2019].

4. **Natural Selection** Natural selection [Mayr 2001] is a meta-principle: it is a process of downward causation [Campbell 1974] that allows the others listed above to come into being. Because the biological needs listed above are attractor states in the adaptive landscape [McGhee 2011], evolutionary convergence takes place [McGhee 2006]: that is, there are multiple ways they can be met. Any physiological implementation in the equivalence class that satisfies the need will do. Thus this is an example of multiple realisability (Section 2.6), which characterizes topdown causation [Ellis 2016].

5. **Smoking, lung cancer, and death** The relation between smoking and lung cancer is an established causal link, as discussed in depth in [Pearl and Mackenzie 2018]. It can certainly be redescribed at the physics level, but the key concepts in the correlation - smoking, cancer - cannot. Therefore, starting off with an initial state described at the microphysics level, one cannot even in principle determine the probabilities of cancer occurring on the basis of those variables alone, let alone when death will occur as a result of the cancer, because death also cannot be described at that level.

   Once cancer occurs (at the genetic/cellular levels **L3/L4**) leading to death (at the whole organism level **L6**) this will alter physical outcomes at the ion/electron level **L2** because the process of life (see above) has ceased. This is a real causal chain, not just a handy redescription of micro physics: smoking causes cancer and then death as a matter of fact. The physics allows this of course, but the actual physical trajectories and outcomes follows from the essential higher level dynamics of the cessation of being alive.

## 2.4 Equal Validity of Levels

There is a valid Effective Theory $\mathbf{ET_L}$ at each level **L**, each of them represents a causally valid theory holding at its level, none more fundamental than the others. This is expressed nicely in [Schweber 1993], commenting on Phil Anderson's views:

> *"Anderson believes in emergent laws. He holds the view that each level has its own "fundamental" laws and its own ontology. Translated into the language of particle physicists, Anderson would say each level has its effective Lagrangian and its set of quasistable particles. In each level the effective Lagrangian - the "fundamental" description at that level - is the best we can do."*

None of them can be deemed to be more fundamental than any other, *inter alia* because none of them is *the* fundamental level, i.e. none is the hoped for Theory of Everything (TOE). This has to be the case because we don't know the underlying TOE, if there is one, and so don't - and can't - use it in real applications. So all the physics laws we use in applications are effective theories in the sense of [Castellani 2002], applicable at the appropriate level. Similarly, there are very well tested effective theories at levels **L3-L5** in biology: the molecular level, the cellular level, the physiological systems level for example. Whenever there are well established laws at the higher levels (for example the laws of perception at Level **L6**) the same applies to them too.

More fundamentally, this equal causal validity occurs because higher levels are linked to lower levels by a combination of upwards and downwards causation [Noble 2012] [Noble 2016] so no level by itself is causally complete. They interact with each other with each level playing a role in causal completeness. Hence ([Noble 2016]:160),

> ***The Principle of Biological Relativity**: There is no privileged level of causation in biology: living organisms are multi-level open stochastic systems in which the behaviour at any level depends on higher and lower levels and cannot be fully understood in isolation*

This is because of circular causality which for example necessarily involves downward causation from the whole cell to influence the behaviour of its molecules just as much as upward causation form the molecular level to the cellular level [Noble 2016]:163-164). This applies to all levels in **Table 1**, i.e. it includes the underlying physics levels as well [Ellis and Kopel 2019] [Ellis 2020b], as has to be the case for physical consistency.

In the case of the brain, after having set out in depth the hierarchical structure of the brain ([Churchland and Sejnowski 2016]:11,27-48), Churchland and Sejnowski state ([Churchland and Sejnowski 2016]:415)

> "*An explanation of higher level phenomena in terms of lower level phenomena is usually referred to as a* reduction, *though not in the perjorative sense that implies the higher levels are unreal, explanatorily dismissable, or somehow empirically dubious",*

which agrees with the view put here. Brain computational processes have real causal power [Marr 2010] [Scott 2002] [Churchland and Sejnowski 2016].

## 2.5 Types of causation

Causation can be characterised either in an interventionist or a counterfactual sense, either indicating when causation takes place [Pearl 2009] [Pearl and Mackenzie 2018]. The first key claim I make is that as well as upward causation, downward causation takes place

[Noble 2012] [Ellis 2016]. The second one is that as well as efficient causation, Aristotle's other forms of causation play a key role in real world outcomes.

**Downward causation** Physicists take for granted upward causation, leading to emergence through aggregation effects such as coarse graining. However one can claim there is also downward causation that occurs via various mechanisms [Noble 2008] [Ellis 2012] [Ellis 2016], allowing strong emergence [Chalmers 2000] to occur. Carlo agrees downward causation takes place, but believes it can be rewritten purely in terms of low level physics, and hence does not represent strong emergence.

Downwards effects in a biological system occur because of physiological processes [Noble 2008], [Noble 2012]. These processes [Hall 2016] are mediated at the molecular level by developmental systems [Oyama et al 2001] operating through metabolic and gene regulator networks [Wagner 2014] and cell signalling networks [Berridge 2014], guided by higher level physiological needs. They reach down to the underlying physical level **L2** via time dependent constraints [Ellis and Kopel 2019]. The *set of interactions* between elements at that level is uniquely characterised by the laws of physics $L$, but their *specific outcomes* are determined by the biological context in which they operate.

An example is determination of heart rate. Pacemaker activity of the heart is via cells in the sinoatrial node that create an action potential and so alter ion channel outcomes. This pacemaking circuit is an integrative characteristic of the system as a whole [Fink and Noble 2008] - that is, it is an essentially higher level variable - that acts down to the molecular level [Noble 2012] [Noble 2016]. In the synchronic case - nothing changes at either macro or micro levels - it is correct that one can predict the lower level and hence the higher level dynamics purely from the lower level initial state. However if the higher level state changes - an athlete starts running - the higher level state changes, and this alters lower level conditions. Nothing about the initial molecular level state of the heart or the underlying physics state could predict this happening. Neither could initial knowledge of both the athletes heart and brain micro states determine this outcome, because it depended on an external event - the firing of the starting gun, another macro level event which the athlete's initial states cannot determine.

Considering the individual athlete, causation at the macro level is real: the firing of the starting gun led to her leaving the starting post. Downward causation that alters motion of ATP molecules in her muscles via metabolic networks is real: that is a well established physiological process [Rhoades and Pflanzer 1989]. The result is altered electron flows in the muscles, in a way consistent with the laws of physics but unpredictable from her initial microphysical state. Regression to include the brain state of the person firing the gun will not save the situation, as one then has to include all the influences on his brain state [Noble *et al* 2019] as well as all the stochastic elements in his brain (Section 5.3).

A similar example of a rhythmic pattern determined by a network as a whole is the stomatogastric ganglion of the spiny lobster ([Churchland and Sejnowski 2016]:4-5):

> "*The network in question contains about 28 neurons and serves to drive the muscles controlling the teeth of the gastric mill so that food can be ground up for digestion. The output of the network is rhythmic, and hence the muscular action and the grinders movements are correspondingly rhythmic. The basic electrophysiological and anatomical features of the neurons have been catalogued, so that the microlevel vitae for each cell in the network is impressively detailed. What is not understood is how the cells in the network interact to constitute a circuit that produces the rhythmic pattern. No one cell is a repos-*

*itory for the cells rhythmic output; no one cell is itself the repository for the properties displayed by the network as a whole. Where then does the rhythmicity come from? Very roughly speaking, from the patterns of interactions among cells* and *the intrinsic properties of the component cells.*

The network produces rhythmic patterns in the cells, which produce rhythmic activity in the constitutive electrons and ions. This is a classic example of higher level order controlling both macro and micro level outcomes.

**Types of downward causation** The basic type of downward causation are as follows (developed from [Ellis 2012] [Noble 2012] [Noble 2016] [Ellis 2016]):

**TD1A Boundary conditions** are constraints on particles in a system arising from the environment[7] as in the case of a cylinder determining pressure and temperature of the enclosed gas, or the shape of tongue and lips determining air vibrations and so spoken words. **Structural Constraints** are fairly rigid structures that determine possible micro states of particles that make up the structure, as in the case of a cylinder constraining the motion of a piston, or a skeleton that supports a body.

**TD1B Channeling and Containing constraints** are key forms of contextual causation shaping microbiological and neural outcomes. **Channeling constraints** determine where reactants or electrical currents can flow, as in blood capillaries in a body, wires in a computer, or neural axons and dendrites in a brain. **Containing constraints** confine reactants to a limited region, so preventing them from diffusing away and providing the context for reaction networks to function. A key case is a cell wall.

**TD2A Gating and signalling constraints** Gating constraints control ingress and egress to a container, as in the case of voltage gated ion channels in axons, or ligand gated ion channels in synapses. They function via conformational changes controlled by voltage differential in the former case, and molecular recognition of ligands in the latter case, thus underlying cell signalling processes [Berridge 2014].

**TD2B Feedback control to attain goals** is a cybernetic process where the difference between a goal and the actual state of a system generates an error signal that is fed back to a controller and causes corrective action, as in thermostats and engine governors [Wiener 1948]. In biology this is *homeostasis*, a crucial feature of physiology at all levels [Hall 2016]. Because of this closed causal loop, goals determine outcomes. Changing the goals changes both macro and micro outcomes, as in altering the setting on a thermostat. In biology, multilevel homeostatic systems are continually responding to internal changes and external perturbations [Billman 2020].

**TD3A Creation of New Elements** takes place in two ways. **Creation of new lower level elements** occurs in physics when crystal level conditions create quasiparticles such as phonons that play a key role in dynamics at the electron level [Ellis 2020a]. This is what [Gillett 2019] calls a *Downward Constitutive relation.* It occurs in biology when genes are read to create proteins, a contextual process [Gilbert and Epel 2009] controlled by gene regulatory networks according to higher level needs [Noble 2016]. **Creation of new higher level elements** restructures lower level relations and so alters lower level dynamics. In engineering this takes

[7] Carlo's example of Jupiter causing tides on Earth fits here: Jupiter is part of the Earth's environment, causing a detectable gravitational field at Marseilles.

place by manufacturing processes such as making a transistor. In biology this occurs when cell division takes place at the cellular level, and when an organism gives birth to progeny at the organism level. The context of lower level dynamics changes completely in both cases. In the latter case, as Darwin already recognised, sexual selection takes place and determines outcomes, involving very complex social and psychological interactions that alter outcomes at the genetic and physical levels.

**TD3B Deleting or Altering Lower Level elements** is the complementary process that is crucial in biology. In developmental biology, apoptosis (programmed cell death) plays a key role for example in digit formation (separating fingers and thumbs), while in neural development, synaptic connections are pruned as development takes place [Wolpert *et al* 2002]. Cell are specialised to perform specific functions as growth takes place, altering their nature and behaviour. A fundamental biological process is **Adaptive selection due to selection criteria** which alters either the set of lower level elements by deletion as in Darwinian selection [Campbell 1974] and the functioning of the immune system, or selecting optimal configurations, as in neural network plasticity involved in learning.

The higher level types of downward causation: **TD4** (Adaptive selection of goals) and **TD5** (Adaptive selection of selection criteria) build on these ones [Ellis 2012] [Ellis 2016].

The key issue is whether any of these types of downward causation are really causally effective, or just redescriptions in convenient form of microphysical causation.

**Aristotle's kinds of causation** There is an important further point as regards causation. As Aristotle pointed out [Bodnar 2018], there are four kinds of causation that occur in the real world. This is discussed by ([Juarrero 2002]:2,125-128,143) ([Noble 2016]:176-179) and ([Scott 2002]:298-300) They are

- **Material Cause**: the physical stuff that is needed for an outcome; the stuff out of which it is made, e.g., the bronze of a statue. In biology this is the physical stuff, the chemical elements as characterised by the periodic table, that make biology possible.
- **Formal Cause**: which makes anything what it is and no other; the material cause necessary for some outcome must be given the appropriate form through the way in which the material is arranged e.g., the shape of a statue. In biology, this is the structure at each level that underlies function at that level: physiological systems [Hall 2016] and the underlying biomolecules such as proteins [Petsko and Ringe 2009].
- **Efficient Cause**: The primary source of the change or rest, the force that brings an action into being; nowadays in the Newtonian case taken to be the effect of forces on inert matter, in the quantum chemistry case, Schrödinger's equation.
- **Final Cause**: the goal or purpose towards which something aims: "that for the sake of which a thing is done".

Physics only considers efficient causes [Juarrero 2002]. Biology however needs material, formal, and efficient causes. [Hofmeyer 2018] gives a careful analysis of how the relation between them can be represented and how they are are realised in biology, giving as an example an enzyme that catalyses a reaction. He explains that a set of rules, a convention or code, forms an interface between formal and efficient cause.

All four kinds of causation are needed to determine specific outcomes in social contexts, which is the context within which brains function. Without taking them all into account, one cannot even account for existence of a teapot [Ellis 2005].

A network of causation is always in action when any specific outcomes occurs. When we refer to 'The Cause', we are taking all the others for granted - the existence of the Universe, of laws of physics of a specific nature, and of the Earth for example.

**Tinbergen's 'Why' questions** In biology, an alternative view on causation is provided by Tinbergen's four 'Why' questions. [Bateson and Laland 2013] summarise thus:

> *"Tinbergen pointed out that four fundamentally different types of problem are raised in biology, which he listed as 'survival value', 'ontogeny', 'evolution', and 'causation'. These problems can be expressed as four questions about any feature of an organism: What is it for? How did it develop during the lifetime of the individual? How did it evolve over the history of the species? And, how does it work?"*

That is, he raises functional, developmental, evolutionary, and mechanistic issues that all have to be answered in order to give a full explanation of existence, structure, and behaviour of an organism.

## 2.6 Multiple Realisability

A key point is that multiple realisability plays a fundamental role in strong emergence [Menzies 2003]. Any particular higher level state can be realised in a multiplicity of ways in terms of lower level states. In engineering or biological cases, a high level need determines the high level function and thus a high level structure that fulfills it. This higher structure is realised by suitable lower level structures, but there are billions of ways this can happen. It does not matter which of the equivalence class of lower level realisations is used to fulfill the higher level need, as long as it is indeed fulfilled. Consequently you cannot even express the dynamics driving what is happening in a sensible way at a lower level.

Consider for example the statements *The piston is moving because hot gas on one side is driving it* and *A mouse is growing because the cells that make up its body are dividing.* They cannot sensibly be described at any lower level not just because of the billions of lower level particles involved in each case, but because *there are so many billions of different ways this could happen at the lower level*, this cannot be expressed sensibly at the proton and electron level. The point is the huge number of different combinations of lower level entities can represent a single higher level variable. Any one of the entire equivalence class at the lower level will do. Thus it is not the individual variables at the lower level that are the key to what is going on: it is the equivalence class to which they belong. But that whole equivalence class can be describer by a single variable at the macro level, so that is the real effective variable in the dynamics. This is a kind of interlevel duality:

$$\{v_{\mathbf{L}} \in \mathbf{L}\} \Leftrightarrow \{v_{\mathbf{i}} : v_{\mathbf{i}} \in E_{\mathbf{L\text{-}1}}(v_{\mathbf{L\text{-}1}}) \in \mathbf{(L\text{-}1)}\} \tag{4}$$

where $E_{\mathbf{L\text{-}1}}(v_{\mathbf{L\text{-}1}})$ is the equivalence class of variables $v_{\mathbf{L\text{-}1}}$ at Level **L-1** corresponding to the one variable $v_{\mathbf{L}}$ at Level **L.** The effective law $\mathbf{EF_L}$ at Level **L** for the variables $v_{\mathbf{L}}$ at that level is equivalent to a law for an entire equivalence class $E_{\mathbf{L\text{-}1}}(v_{\mathbf{L\text{-}1}})$ of variables at Level **L-1**. It does not translate into an Effective Law for natural variables $v_{\mathbf{L\text{-}1}}$ *per se* at Level **L-1**.

The importance of multiple realisability is discussed in [Menzies 2003] [Ellis 2019] and [Bishop and Ellis 2020].

**Essentially higher level variables and dynamics** *The higher level concepts are indispensible when multiple realisability occurs, firstly because they define the space of data* $d_L$ *relevant at Level* $\boldsymbol{L}$*, and secondly because of (4), variables in this space cannot be represented as natural kinds at the lower level. Effective Laws* $\boldsymbol{EF_L}$ *at level* $\boldsymbol{L}$ *can only be expressed at level* $\boldsymbol{L\text{-}1}$ *in terms of an entire equivalence class at that level. One can only define that equivalence class by using concepts defined at level* $\boldsymbol{L}$*.*

To inject reality into this fact, remember that the equivalence class at the lower level is typically characterised by Avagadro's number.

## 2.7 Higher Level Organising Principles

A key issue in the discussion is the degree to which higher level dynamics depends on the lower level dynamics. As can be seen from the previous subsections, the nature of biological causation is quite unlike the nature of causation at the underlying physical levels. What determines these outcomes then?

**Higher Level Organising Principles** The key idea here is that higher level biological Organising Principles exist that are independent of the underlying lower level dynamics, and shape higher level outcomes. The specific lower level realisation is immaterial, as long as it is in the right equivalence class (Section 2.6). Generically they form attractors that shape higher level outcomes [Juarrero 2002]152-162; the lower level components come along for the ride, with many biological oscillators being examples ([Noble 2016]:76-86,179).

**Protectorates** This is parallel to the claim by [Laughlin and Pines 2000] of existence of classical and quantum protectorates, governed by dynamical rules that characterise emergent systems as such. They state

> "*There are higher organising principles in physics, such as localization and the principle of continuous symmetry breaking, that cannot be deduced from microscopics even in principle. ... The crystalline state is the simplest known example of a quantum protectorate, a stable state of matter whose generic low-energy properties are determined by a higher organizing principle and nothing else... they are transcendent in that they would continue to be true and lead to exact results even if the underlying Theory of Everything was changed.*

As an example, [Haken 1996] states that profound analogies between different systems become apparent at the order parameter level, and suggest that the occurrence of order parameters in open systems is a general law of nature. He characterizes this in terms of a *slaving principle*[8] [Haken and Wunderlin 1988]. [Green and Batterman 2020] develop this further, citing the universality of critical phenomena as a physics case. The Renormalisation Group explanation extracts structural features that stabilize macroscopic phenomena irrespective of changes in microscopic details

**Biology** In biology, such organising principles can be claimed to govern microbiology, physiology, and neuroscience (Sections 2.1 and 4). The idea is that once life exists and evolutionary processes have started, they are what shape outcomes, rather than the underlying physical laws, because they express essential biological needs [Kauffman 1995].

[8] I thank Karl Friston for this comment.

Physical laws of course *allow* the outcomes to occur: they lie within the *Possibility Space* $\Omega_L$ of outcomes allowed by the physical laws $L$, for instance the proteins enabling all this to occur are characterised by a possibility space of huge dimension, as are the metabolic networks and gene regulatory networks that lead to specific outcomes [Wagner 2014]. But as emergence takes place through developmental processes repeated many many times over evolutionary timescales, it is these principles that determine biological success. Hence [Ginsburg and Jablonka 2019] it is they that determine evolutionary outcomes in an ongoing Evo-Devo process [Carroll 2005] [Carroll 2008]. They act as attractors for both evolution and for ongoing brain dynamics.

This proposal is supported in multiple ways.

**In functional terms,** homeostasis is a central organising principle in all physiology at multiple scales: "*It is important to note that homeostatic regulation is not merely the product of a single negative feedback cycle but reflects the complex interaction of multiple feedback systems that can be modified by higher control centers*" [Billman 2020]. Also physiological functions acting as dynamical systems have attractors that organise outcomes. For example, this happens in the neural dynamics of cell assemblies ([Scott 2002]:244-248):

> "*In Hopfield's formulation, each attractor is viewed as a* pattern *stored nonlocally by the net. Each such pattern will have a basin of attraction into which the system can be forced by sensory inputs.*"

Thus cell assemblies form attractors ([Scott 2002]:287). Also Hopfield neural networks converge to attractors in an energy landscape [Churchland and Sejnowski 2016]:88-89) and attractor networks are implemented by recurrent collaterals ([Rolls 2016]:75-98).

**In developmental terms** it can be expressed in terms of Waddington's epigenetic landscape [Gilbert 1991] ([Noble 2016]:169,259) which presents much the same idea via cell fate bifurcation diagrams. This is how developmental processes converge on outcomes based in the same higher level organising principles.

**In evolutionary terms,** it can be expressed in terms of the adaptive landscape of Sewell Wright [Wright 1932l] [McGhee 2006], showing how evolution converges to adaptive peaks where these principles are supported to a greater or lesser degree. This viewpoint is supported by much evidence for convergent evolution [McGhee 2011].

**Neuroscience** There is a huge amount written about neuroscience and biological psychology, with a vast amount of detail: [Scott 2002] [Purves *et al* 2008] [Kandel *et al* 2013] [Churchland and Sejnowski 2016] [Clark 2016] [Gray and Bjorklund 2018] for example.

The issue is, Can one extract higher level organising principles for the brain from them? I believe one can, examples being hierarchical predictive coding [Clark 2013] and the Free Energy Principle [Friston 2010]. I collect them together in the following three sections, looking at how the brain handles the constant influx of new data (Section 3), the issue of constantly adjusting to the environment (Section 4), and how the brain uses micro level stochasticity to allow macro level agency (Section 5). I suggest that it is these principles at the macro level that are the real determinants of what happens, solving the puzzle of how ordered outcomes can emerge in the context of an open system, where the microdynamic states of an individual brain cannot in principle determine future outcomes because they do not have the data necessary to do so.

If that is correct, these principles reach down to determine micro level outcomes via the various mechanisms outlined in Section 2.5. Furthermore they are themselves attractors in evolutionary space: they will tend to come into existence because they enhance prospects of reproductive success [Ginsburg and Jablonka 2019].

# 3 The Predictive Brain: Brains as open systems

Each human body, and each brain, is an open system. This is where the difference between synchronic and diachronic emergence is crucial. It has two aspects: our brains are not made of the same matter as time progresses (Section 3.1), and new information is coming in all the time and altering our brain states (Section 3.2). The way this is interpreted depends on the fact that our brain is an emotional brain (Section 3.3) and a social brain (Section 3.4). Language and symbolism enables abstract and social variables to affect outcomes (Section 3.5). Consequently the microphysical state of a specific person's brain is unable as a matter of principle to predict their future brain states (Section 3.6) Predictive brains that can handle this situation are attractor states for brain evolutionary development.

## 3.1 Matter and Metabolism: We are not the same molecules

Because we are open systems [Peacocke 1989], the human body at time $t_2 > t_1$ is not made of the same material particles as it was at time $t_1$. Thus what happens in life is like the case of a candle ([Scott 2002]:303):

> "*As a simple example of an open system, consider the flame of a candle. .. Because the flame is an open system, a relation* $P_1 \to P_2$ *cannot be written - even "in principle"- for the physical substrate. This follows from the fact that the physical substrate is* continually changing. *The molecules of air and wax vapour comprising the flame at time* $t_2$ *are entirely different from those at time* $t_1$*. Thus the detailed speeds and positions of the molecules present at time* $t_2$ *are unrelated to those at time* $t_1$*. What remains constant is the flame itself - a* process."

**Body maintenance**: A balance between protein synthesis and protein degradation is required for good health and normal protein metabolism. Protein turnover is the replacement of older proteins as they are broken down within the cell, so the atoms and elementary particles making up the cell change too. Over time, the human body is not even made up of the same particles: they turn over completely on a timescale of 7 years [Eden *et al* 2011] [Toyama and Hetzer 2013]

**The brain** Neuroscientist Terence Sejnowski states:[9]

> '*'Patterns of neural activity can indeed modify a lot of molecular machinery inside a neuron. I have been puzzled by my ability to remember my childhood, despite the fact that most of the molecules in my body today are not the same ones I had as a child; in particular, the molecules that make up my brain are constantly turning over, being replaced with newly minted molecules. "*

Metabolic networks ensure the needed replacements take place on a continuous basis, despite stochasticity at the molecular level (Section 5). This is where multiple realisability plays a key role (Section 2.6).

**Conclusion** ***Initial data for the specific set of particles making up a specific brain at time*** $t_1$ ***cannot determine emergent outcomes uniquely for that brain over time, for it is not made of the same set of particles at time*** $t_2 \gg t_1$**.**

[9] https://www.edge.org/response-detail/10451

## 3.2 Dealing with New Information: The Predictive Brain

That effect of course takes time. The very significant immediate ongoing effect of being an open system is that incoming sensory information conveys masses of new data on an ongoing basis. This new data may contain surprises, for example a ball smashes a window. The brain has to have mechanisms to deal with such unpredictability: the previously stored data at the microphysics level cannot do so, as it does not take this event into account.

**Hierarchical predictive coding** Indeed, the brain has developed mechanisms to make sense of the unpredictable inflow of data and best way react to it [Clark 2013] [Clark 2016] [Hohwy 2013] [Hohwy 2016] [Szafron 2019]. Andy Clark explains [Clark 2013]:

> "*Brains, it has recently been argued, are essentially prediction machines. They are bundles of cells that support perception and action by constantly attempting to match incoming sensory inputs with top-down expectations or predictions. This is achieved using a hierarchical generative model that aims to minimize prediction error within a bidirectional cascade of cortical processing. Such accounts offer a unifying model of perception and action, illuminate the functional role of attention, and may neatly capture the special contribution of cortical processing to adaptive success. This 'hierarchical prediction machine' approach offers the best clue yet to the shape of a unified science of mind and action."*

In brief, following up Ross Ashby's notion that "*the whole function of the brain is summed up in error correction,*" the following takes place in an ongoing cycle:

**PB1 Hierarchical generative model** *The cortex uses a hierarchical model to generate predictions of internal and external conditions at time $t_2$ on the basis of data available at time $t_1$.*

**PB2 Prediction error and attention** *During the interval $[t_1, t_2]$ sensory systems (vision, hearing, somatosensory) receive new information on external conditions and internal states At time $t_2$, nuclei in the thalamus compare the predictions with the incoming data. If it exceeds a threshold, an error signal ('surprisal') is sent to the cortex to update its model of the internal and external situation (Bayesian updating), and focus attention on the discrepancy.*

**PB3 Action and outcomes** *The updated model is used to plan and implement action. The impact of that action on the external world provides new data that can be used to further update the model of the external world (active intervention).*

This is an interlevel information exchange as described by ([Rao and Ballard 1999]:80):

> "*Prediction and error-correction cycles occur concurrently throughout the hierarchy, so top-down information influences lower-level estimates, and bottomup information influences higher-level estimates of the input signal".*

The outcome [Hohwy 2007] is (as quoted in [Clark 2013]),

> "*The generative model providing the "top-down"predictions is here doing much of the more traditionally "perceptual" work, with the bottom up driving signals really providing a kind of ongoing feedback on their activity (by fitting, or failing to fit, the cascade of downward-flowing predictions). This procedure*

*combines "top-down" and "bottom-up" influences in an especially delicate and potent fashion, and it leads to the development of neurons that exhibit a "selectivity that is not intrinsic to the area but depends on interactions across levels of a processing hierarchy"* ([Friston 2003], p.1349). *Hierarchical predictive coding delivers, that is to say, a processing regime in which context-sensitivity is fundamental and pervasive*".

**Perception** Consequently, perception is a predictive affair [Purves 2010]. Helmholz's inverse problem (how to uniquely determine a 3-d world from a 2-d projection) is solved by filling in missing information on the basis of our expectations. ([Kandel 2012]:202-204) gives a overview of how this understanding originated with Helmholz, who called this top-down process of hypothesis testing *unconscious inference*, and was developed by Gombrich in his book *Art and Illusion* [Gombrich 1961]. [Kandel 2012] (pages 304-321) emphasizes the top-down aspect of this process, and its relation to memory. [Purves 2010](pp.120-124) describes how he came to the same understanding (see also page 221).

**Action** The relation to action is given by [Friston 2003] [Friston *et al* 2009] [Clark 2013]. It is described thus by ([Hawkins and Blakeslee 2007]:158)

> "*As strange as it sounds, when your own behaviour is involved, your predictions not only precede sensation, they determine sensation. Thinking of going to the next pattern in a sequence causes a cascading prediction of what you should experience next. As the cascading prediction unfolds, it generates the motor commands necessary to fulfill the prediction. Thinking, predicting, and doing are all part of the same unfolding of sequences moving down the cortical hierarchy."*

[Seth 2013] summarised the whole interaction thus:

> "*The concept of Predictive Coding (PC) overturns classical notions of perception as a largely 'bottom-up' process of evidence accumulation or feature detection, proposing instead that perceptual content is specified by top-down predictive signals that emerge from hierarchically organized generative models of the causes of sensory signals. According to PC, the brain is continuously attempting to minimize the discrepancy or 'prediction error' between its inputs and its emerging models of the causes of these inputs via neural computations approximating Bayesian inference. Prediction errors can be minimized either by updating generative models (perceptual inference and learning; changing the model to fit the world) or by performing actions to bring about sensory states in line with predictions (active inference; changing the world to fit the model"*

This is a very brief sketch of a very complex program, summarised in Andy Clark's book *Surfing Uncertainty* [Clark 2016] and in [Miller and Clark 2018]. Nothing here contradicts the mechanisms discussed in depth in texts such as [Purves *et al* 2008] [Kandel *et al* 2013] [Churchland and Sejnowski 2016]. Those texts set the foundations for what is proposed above, but do not develop these aspects in depth. For example [Kandel *et al* 2013] has just one relevant section: "Visual perception is a creative process" (page 492).

Thus the viewpoint put here accepts the mechanisms discussed in those books (and the underlying physics), and puts them in a larger context that emphasizes overall organising features that are crucial in enabling the brain to function in the face of uncertainty.

However there are three further important aspects to be taken into account.

## 3.3 The emotional brain

A first further crucial aspect of our brains is that they are **emotional brains**. The understandings and actions enabled by the predictive mechanisms mentioned above are crucially affected and shaped by affective (emotional) states.

The cognitive science paradigm of purely rational choice is not the way the real brain works. Emotion has key effects on cognition [Damasio 2006] and behaviour [Panksepp 2009] [Purves *et al* 2008] [Panksepp and Biven 2012] [Gray and Bjorklund 2018].

**EB1 The emotional brain** *Both primary (innate) and secondary (social) emotions play a key role in guiding cognition and focusing attention.*

The predictive coding paradigm can be extended ([Clark 2016]:231-237) to include this case. [Seth 2013] says the following

> "*The concept of the brain as a prediction machine has enjoyed a resurgence in the context of the Bayesian brain and predictive coding approaches within cognitive science. To date, this perspective has been applied primarily to exteroceptive perception (e.g., vision, audition), and action. Here, I describe a predictive, inferential perspective on interoception: 'interoceptive inference' conceives of subjective feeling states (emotions) as arising from actively-inferred generative (predictive) models of the causes of interoceptive afferents. The model generalizes 'appraisal' theories that view emotions as emerging from cognitive evaluations of physiological changes ... interoceptive inference involves hierarchically cascading top-down interoceptive predictions that counterflow with bottom-up interoceptive prediction errors. Subjective feeling states - experienced emotions - are hypothesized to depend on the integrated content of these predictive representations across multiple levels* "

[Miller and Clark 2018] develop this crucial emotional relationship to cortical activity in depth, using the predictive coding framework:

> "*But how, if at all, do emotions and sub-cortical contributions fit into this emerging picture? The fit, we shall argue, is both profound and potentially transformative. In the picture we develop, online cognitive function cannot be assigned to either the cortical or the sub-cortical component, but instead emerges from their tight co-ordination. This tight co-ordination involves processes of continuous reciprocal causation that weave together bodily information and 'top-down' predictions, generating a unified sense of what's out there and why it matters. The upshot is a more truly 'embodied' vision of the predictive brain in action.*"

As well as influencing immediate functioning of the brain, affect relates crucially to brain plasticity and so to changes in brain micro structure (Section 5.4).

## 3.4 The social brain

Second, a crucial aspect of our brains is that they are **social brains**: we are evolved to live in a social context, which has key influences on our lives and minds as the brain receives data and responds to the situation around. Sociality appears to be a main driver for human brain evolution [Dunbar 1998] [Dunbar 2003] and results in social cognition ([Purves *et al* 2008]:359-392) and cognitive neuroscience [Cacioppo *et al* 2002]. This again crucially affects how we handle the incoming information.

**The advantage of social brains** Living in cooperative groups greatly enhanced our ancestors survival prospects [Harari 2014] enabling the rise of cooperative farming, culture, and technology, which then was the key to the emergence of civilisation that enabled our dominance over the planet [Bronowski 2011]. A social brain was needed for social cohesion to emerge: the cognitive demands of living in complexly bonded social groups selected increasing executive brain (neocortical) size [Dunbar 1998a] [Dunbar 2014].

**The nature of the social brain: Theory of Mind** It is not just a matter of being cooperative and able to communicate: central to the social brain is the ability known as "theory of mind" (ToM) [Dunbar 1998a]. It is very important that we can read other peoples minds (understanding their intentions) - which we do on an ongoing basis [Frith 2013]. We all have a theory of mind [Frith and Frith 2005]. Its cortical basis is discussed by [Frith 2007], but additionally it has a key precortical base related to the primary emotional systems identified by [Panksepp 2009], namely the very strong emotional need to belong to a group [Panksepp and Biven 2012] [Ellis and Toronchuk 2013] [Stevens and Price 2015]

Its evolutionary basis is discussed by [Donald 1991] [Tomasello 2009] [Dunbar 2014]. It is summed up by ([Donald 2001]:86-87) as follows:

> "*Our normal focus is social, and social awareness is highly conscious, that is, it heavily engages our conscious capacity... Conscious updating is vital to social life ... One might even make the case that consciousness- especially our lightning fast, up-to-date, socially attuned human consciousness - is the evolutionary requirement for both constructing and navigating human culture. It remains the basis, the sine qua non, for all complex human interactions*".

Michael Tomasello agrees, as is evident in the title of his book *The Cultural Origin of Human Cognition* [Tomasello 2009].

**Relation to predictive coding** The description of the social brain in terms of the predictive processing paradigm is presented by [Constant *et al* 2019] through the concept of the extended mind:[10]

> "*Cognitive niche construction is construed as a form of instrumental intelligence, whereby organisms create and maintain cause-effect models of their niche as guides for fitness influencing behavior. Extended mind theory claims that cognitive processes extend beyond the brain to include predictable states of the world that function as cognitive extensions to support the performance of certain cognitive tasks. Predictive processing in cognitive science assumes that organisms (and their brains) embody predictive models of the world that are leveraged to guide adaptive behavior. On that view, standard cognitive functions - such as action, perception and learning - are geared towards the optimization of the organism's predictive (i.e., generative) models of the world. Recent developments in predictive processing - known as active inference - suggest that niche construction is an emergent strategy for optimizing generative models.*

Those models include models of social context and of other minds, characterised via cultural affordances [Ramstead *et al* 2016]. [Veissiére *et al* 2020] state

[10] See also [Kirchhoff *et al* 2018] and [Hesp *et al* 2019].

> "*We argue that human agents learn the shared habits, norms, and expectations of their culture through immersive participation in patterned cultural practices that selectively pattern attention and behaviour. We call this process "Thinking Through Other Mind" (TTOM) - in effect, the process of inferring other agents' expectations about the world and how to behave in social context.* "

Then downward causation from the social environment changes the brain:

> "*The brain only has direct access to the way its sensory states fluctuate (i.e., sensory input), and not the causes of those inputs, which it must learn to guide adaptive action - where 'adaptive' action solicits familiar, unsurprising (interoceptive and exteroceptive) sensations from the world. The brain overcomes this problematic seclusion by matching the statistical organization of its states to the statistical structure of causal regularities in the world. To do so, the brain needs to re-shape itself, self-organizing so as to expect, and be ready to respond with effective action to patterned changes in its sensory states that correspond to adaptively relevant changes 'out there' in the world"*

The sociology of this all is discussed by [Berger 1963] and [Berger and Luckmann 1991]. Overall, one can summarise as follows:

**SB1 The social brain** *Because we live in a social world we are very socially aware. We have a social brain which shapes our responses to incoming data in crucial ways on the basis of social understandings, which are continually changing over time.*

Theory of mind is based in prediction, and is a routine part of everyday life [Frith 2013].

## 3.5 The symbolic brain

Third, a key feature of the social brain is its ability to engage in spoken and written language, and more generally to engage in symbolism. This adds in a whole new category of incoming information that the brain has to take into account and respond to.

**Language** A key step in evolution of mind is developing language. [Dunbar 1998a] suggests its prime function is to enable exchange of information regarding bonding in the social group. It is a product of a mind-culture symbiosis ([Donald 2001]:11,202) and forms the basis of culture ([Donald 2001]:274), symbolic technologies ([Donald 2001]:305), as well as cultural learning ([Tomasello 2009]:6) and inheritance ([Tomasello 2009]:13). [Ginsburg and Jablonka 2019] Language enables sharing ideas and information over time and distance, and enables the social and psychological power of stories [Gottschall 2012].

**Abstract and social variables** In evolutionary terms, the transition to the symbolic species [Deacon 1997] enabled abstract causation [Ellis and Kopel 2019] to occur, which *inter alia* involves social interactions and abstract concepts such as the amount of money in my bank account and the concept of a closed corporation [Harari 2014]. Thus not all the relevant variables are physical variables; some are abstract variables resulting from social interactions [Berger 1963] [Berger and Luckmann 1991] which are causally effective.

**Higher order predictability** Symbolism and abstract reasoning greatly increases our power of prediction: we can simulate situations offline, rather than having to enact them to see what the consequences are. It also greatly increases the complexity of our responses to incoming social data, which are interpreted in the light of the social context [Berger 1963] [Berger and Luckmann 1991] [Donald 2001] [Frith 2013]

**SB2 The symbolic brain** *Human social interaction is based in language, in turn based in our symbolic ability. This ability transforms the way our minds interpret much incoming data, as well as allowing internal cognitive processes that are a major causal factor in our individual and social lives.*

This is the fundamental mechanism by which the brain operates at a macro level, for which there is much evidence. Again one can claim that this is the way the brain operates as a matter of fact, it is not just the way we think it operates. Causation at this level is real: the whole of society depends on it.

This will play an important role in Section 6 because it relates to the interaction of the brain to the outside world.

## 3.6 The Dynamics of the Open Brain

An individual brain considered as an entity on its own is an open system, and has been adapted to handle the problems this represents in successfully navigating the world. This rather than initial brain micro data determines it specific outcomes as time progresses.

**Microphysics data for brain states** Consider a specific individual brain at time $t_1$. During a time interval $[t_1, t_2]$, the initial brain microphysics data $\mathcal{D}(t_1)$ is added to by new data $\mathcal{D}_{ext}(t)$ coming from the environment after $t_1$. The data $\mathcal{D}(t_2)$ at a later time $t_2 > t_1$ is not predictable even in principle from $\mathcal{D}(t_1)$. Hence the microphysics evolution is undetermined by data $\mathcal{D}(t_1)$, even in principle. You may for example see a car crash at a time $t_3 > t_1$ that alters all the future brain states; but your brain did not know that was going to happen.

Thus the brain as an open system receives unexpected information and handles it in a predictive way. The initial state of the brain obviously cannot determine these outcomes as it has no control over what the incoming data will be. This is the key outcome of the difference between synchronic and diachronic emergence.

> **The brain is an open system** ***Initial micro data of a brain state at one moment cannot possibly determine what it will do at a later time, not just because new matter comes in and replaces old, but also because new information comes in from outside and alters outcomes. The initial data at time $t_1$ cannot know what the initial data at time $t_2$ will be and hence cannot determine specific later brain outcomes. The brain handles this uncertainty via the predictive brain mechanisms PB1-PB3, EB1, SB1-SB2 outlined above.***

The physicalist gambit is to say ah yes, but microphysics determines uniquely the evolution of all the other systems the brain is interacting, so the system as a whole is determined by the microphysics dynamics alone. I respond to that proposal in Section 6.

**Predictive Brain Mechanisms as Attractor states** Evolutionary processes will hone in on these predictive brain mechanisms as attractor states. This occurs via the mechanism of exploration and selective stabilisation recognised independently by Changeaux and by Edelman ([Ginsburg and Jablonka 2019]:119-123,247-248).

Thus these mechanisms can be claimed to be Higher Order Principles (see Section 2.7) for brain structure and function. It is their remarkable properties that shape brain structure, and its functioning in the face of the unpredictable flow of incoming data.

# 4 The Learning Brain: Plasticity and Adaptation

In carrying out these responses to incoming information, remembering and learning takes place; indeed this is a pre-requisite for functioning of predictive brain mechanisms. This adds a new dimension to the effects just discussed: not only is the new data unpredictable, but also brain structure is changed in ways affected by that inflow of new data. Thus the context for microphysics outcomes - the specific set of constraints determining electron and ion flow possibilities - is also different at the later time.

Plasticity at the macro level as the brain adapts to its environment, remembers, and learns [Gray and Bjorklund 2018] is enabled by corresponding changes at the micro level as neural networks weights change [Kandel *et al* 2013] [Churchland and Sejnowski 2016]. Thus changes take place at the micro level (Section 4.1) driven by incoming data at the macro level, and resulting in plasticity at the macro level (Section 4.2). Because brain neural nets are changing all the time, the context for outcomes of the underlying physics is also changing all the time (Section 4.3) and is not predictable from the initial brain physical microstate.

## 4.1 Plasticity at the micro level

Learning takes place by change of connectivity and weights in neural networks at the neuronal level [Kandel *et al* 2013] [Churchland and Sejnowski 2016], taking place via gene regulation at the cellular level [Kandel 2001]. This alters the context within which electron and ion flows take place in neural networks and in particular at synapses, thereby shaping outcomes of the underlying universal physical laws.

**Developmental processes** This plasticity occurs particularly when brain development is taking place. Random initial connections are refined ([Wolpert *et al* 2002]: §11) and new experiences can modify the original set of neuronal connections ([Gilbert 1991]:642) while the brain is responding to the surrounding environment ([Purves 2010]:§2-§5, 229).

**Learning Processes** Erik Kandel explored the mechanism of learning in depth. He identified gene regulatory process related to learning [Kandel 2001]

> *"Serotonin acts on specific receptors in the presynaptic terminals of the sensory neuron to enhance transmitter release. ... during long-term memory storage, a tightly controlled cascade of gene activation is switched on, with memory-suppressor genes providing a threshold or checkpoint for memory storage ... With both implicit and explicit memory there are stages in memory that are encoded as changes in synaptic strength and that correlate with the behavioral phases of short- and long-term memory"*

**The relation to physics** These changes alter the context within which the underlying physics operates. Changing constraints at the microphysics level is the mechanism of downward causation to that level [Ellis and Kopel 2019]. This determines what dynamics actually takes place at the ion/electron level, which of course the fundamental laws by themselves cannot do. The outcomes are determined by biological context in this way.

**LB1 The Developing and Learning Brain** *The brain is plastic at the micro level, as development and learning takes place. Neural network connections and weights are altered via gene regulatory processes.*

Thus neural network learning [Churchland and Sejnowski 2016] - a real causal process at each network level - alters electron outcomes and so later psychological level dynamics.

## 4.2 Plasticity at the macro level

Eric Kandel states "*One of the most remarkable aspects of an animal's behavior is the ability to modify that behavior by learning*" [Kandel 2001], and emphasizes that social factors affect this learning. [Kandel 1998] gives five principles for psychotherapy that make this clear. For those who are skeptical of psychotherapy, replace that word with 'teaching' or 'coaching' in the following, and its crucial meaning still comes through.

**Kandel Principle 1** All mental processes, even the most complex psychological processes, derive from operations of the brain. The central tenet of this view is that what we commonly call mind is a range of functions carried out by the brain. The actions of the brain underlie not only relatively simple motor behaviors, such as walking and eating, but all of the complex cognitive actions, conscious and unconscious, that we associate with specifically human behavior, such as thinking, speaking, and creating works of literature, music, and art.

**Kandel Principle 2** Genes and their protein products are important determinants of the pattern of interconnections between neurons in the brain and the details of their functioning. Genes, and specifically combinations of genes, therefore exert a significant control over behavior. ... the transcriptional function of a gene - the ability of a given gene to direct the manufacture of specific proteins in any given cell - is, in fact, highly regulated, and this regulation is responsive to environmental factors ... the regulation of gene expression by social factors makes all bodily functions, including all functions of the brain, susceptible to social influences.

**Kandel Principle 3** Behavior itself can also modify gene expression. Altered genes do not, by themselves, explain all of the variance of a given major mental illness. Social or developmental factors also contribute very importantly. Just as combinations of genes contribute to behavior, including social behavior, so can behavior and social factors exert actions on the brain by feeding back upon it to modify the expression of genes and thus the function of nerve cells. Learning ...produces alterations in gene expression.

**Kandel Principles 4/5** How does altered gene expression lead to the stable alterations of a mental process? Alterations in gene expression induced by learning give rise to changes in patterns of neuronal connections. These changes not only contribute to the biological basis of individuality but strengthen the effectiveness of existing patterns of connections, also changing cortical connections to accommodate new patterns of actions.... resulting in long-lasting effect on the the anatomical pattern of interconnections between nerve cells of the brain.

**The hierarchical predictive coding view** The way this all fits into the predictive coding viewpoint discussed in the last section is explained by [Rao and Ballard 1999].

Overall the outcomes can be summarised thus:

**LB2 The Learning Brain** *The brain is plastic at the macro level as learning takes place, supported by plasticity at the micro level. Learning at the macrolevel responds to social and psychological variables.*

### 4.3 The ever adapting brain

The previous section emphasized, in the case of a single brain, that because of incoming data, the microstate at time $t_2$ cannot be predicted from initial data at time $t_1 < t_2$ because it does not include this incoming data. This section emphasizes that in addition, the micro level constraints are changed because neural network wiring or weights will have changed as the brain adapts at both macro and micro levels to ongoing environmental events and changes. So not only is the data different than expected because the brain is an open system, but the dynamical context for the underlying physics is different too.

> **The brain is an adaptive system.** ***Individual brain structure changes in response to incoming data. As new information comes in, neural network weights are continually changed via gene regulation. This change of context alters constraints in the underlying Lagrangian, and so changes the context for future physical interactions. None of this can be determined by the initial brain micro data at time*** $t_1$***, as these changes are shaped by data that has come in since then.***

This is a further reason why diachronic emergence is crucially different from synchronic.

**Adapting and learning brains as attractor states** Evolutionary processes will hone also in on these learning brain mechanisms as attractor states, via the mechanism of exploration and selective stabilisation recognised independently by Changeaux and by Edelman ([Ginsburg and Jablonka 2019]:119-123,247-248).

## 5 The Stochastic Brain and Agency

A key feature undermining physicalist determinism of brain states is the stochasticity that occurs in biology at the molecular level, which uncouples biology from detailed Laplacian determinism. Section 5.1 discusses this stochasticity, and Section 5.2 how this opens up the way for selecting desired low level outcomes that will fulfill higher level purposes - one of the key forms of downward causation. Section 5.3 discusses how this applies specifically to the brain. A key way that randomness is used in shaping the brain is Neural Darwinism (Section 5.4). The issue of how agency is possible arises, and this is essentially via multi level causal closure that takes advantage of this selective process (Section 5.5).

This shows how biological stochasticity opens up the way to higher level biological needs acting as attractors that shape brain dynamics, rather than brain outcomes being the result purely of deterministic or statistical lower level physical dynamics. All of this can again be traced out at the underlying physics level, but it is the biology that is the essential causal factor through setting the context for physical outcomes.

## 5.1 Biology and stochasticity

There is massive stochasticity at the molecular level in biology. This undoes Laplacian determinism at the micro level: it decouples molecular outcomes from details of initial data at the molecular level. How then does order emerge? By biology harnessing this stochasticity to produce desirable higher level results, as happens for example in the case of molecular machines [Hoffmann 2012] and the adaptive immune system [Noble and Noble 2018].

**Stochasticity and molecular machines** As described in Hoffman's book *Life's ratchet: how molecular machines extract order from chaos* [Hoffmann 2012] biomolecules live in a cell where a molecular storm occurs. Every molecular machine in our cells is hit by a fast-moving water molecule about every $10^{-13}$ seconds. He states

> "*At the nanoscale, not only is the molecular storm an overwhelming force, but it is also completely random.*"

The details of the initial data molecular positions and momenta) are simply lost. To extract order from this chaos, "*one must make a machine that could 'harvest' favorable pushes from the random hits it receives.*" That is how biology works at this level.

**Stochasticity in gene expression** Variation occurs in the expression levels of proteins [Chang *et al* 2008]. This is a property of the population as a whole not of single cells, so the distribution curve showing the number of cells displaying various levels of expression is an attractor created by the population ([Noble 2016]:175-176). Promoter architecture is an ancient mechanism to control expression variation [Sigalova *et al* 2020]. Thus one must use stochastic modeling of biochemical and genetic networks ([Ingalls 2013]:280-295) when determining their outcomes. This affects the detailed physical outcomes of memory processes based in gene regulation [Kandel 2001].

**Stochasticity in genetic variation** The processes of genetic variation before selection are mutation and recombination [Alberts 2007], drift [Masel 2011], and migration. They all are subject to stochastic fluctuations. Mutations arise spontaneously at low frequency owing to the chemical instability of purine and pyrimidine bases and to errors during DNA replication [Lodish *et al* 2000]. Because evolution is a random walk in a state space with dimension given by the number of the different strategies present [Geritz *et al* 1997] this shapes evolutionary outcomes in a way that is unpredictable on the basis of microphysics data. This has key present day outcomes in terms of ongoing mutations of microbes and viruses on the one hand, and of immune system responses on the other, both of which are based in taking advantage of stochasticity [Noble and Noble 2018]. The statistics of outcomes however can be studied in terms of evolution over a rugged fitness landscape [Gillespie 1984] [Kauffman and Levin 1987] [Felsenstein 1988] [Orr 2005]. Small fluctuations can end up in a different attractor basin or adaptive peak.

**The microbiome** A key factor in physiology is that our bodies contain many billions of microbes that affect bodily functioning and health. This is being studied in depth by the Human Microbiome Project [Peterson *etal* 2009] and the Integrative Human Microbiome Project [Integrative H. M. P. 2014] [Integrative H. M. P. 2019].

Thousands of microbial species, possessing millions of genes, live within humans: in the gastrointestinal and urogenital tracts, in the mouth: $10^{18}$, in the nose, in the lungs:

$10^9$/ml, on the skin: $10^{12}$. This leads to infectious diseases (rheumatic fever, hepatitis, measles, mumps, TB, AIDS, inflammatory bowel disease) and allergic/autoimmune diseases (Asthma, diabetes, multiple sclerosis, Croon's disease).

Because of the stochasticity in gene mutation, recombination, and horizontal gene transfer, and the huge numbers involved, together with the impossibility of setting data to infinite precision (§6.1), evolution of specific outcomes is unpredictable in principle, but has major macro level outcomes for individuals.

## 5.2 Stochasticity and selection in biology

This level of stochasticity raises a real problem: how could reliable higher levels of biological order, such as functioning of metabolic and gene regulatory networks and consequent reliable development of an embryo [Wolpert *et al* 2002], emerge from this layer of chaos?

The answer is that evolutionary processes have selected for biological structures that can successfully extract order from the chaos. These structures in turn use the same mechanism: they select for biologically desirable outcomes from an ensemble of physical possibilities presented by this underlying randomness. Higher level biological needs may be satisfied this way. As stated by [Noble and Noble 2018]:

> "*Organisms can harness stochasticity through which they can generate many possible solutions to environmental challenges. They must then employ a comparator to find the solution that fits the challenge. What therefore is unpredictable in prospect can become comprehensible in retrospect. Harnessing stochastic and/or chaotic processes is essential to the ability of organisms to have agency and to make choices.*

This is the opposite of the Laplacian dream of the physical interactions of the underlying particles leading to emergent outcomes purely on the basis of the nature of those interactions. It is the detailed structure of molecular machines, together with the lock and key molecular recognition mechanism used in molecular signalling [Berridge 2014], that enables the logic of biological processes to emerge as effective theories governing dynamics at the molecular level. They exist in the form they do because of the higher level organising principles that take over. The emergent levels of order appear because they are based in higher level organising principles characterising emergent protectorates as described by [Laughlin and Pines 2000] (see Section 2.7). For example Friston's Free Energy Principle [Friston 2010] [Friston 2012] is such a higher level organising principle. It does not follow from the microphysical laws. In fact all the higher level Effective Theories $\mathbf{ET_L}$ (Section 2.4) characterise such Higher Level Organising principles.

**STB1 Variation leads to a variety of states, from which outcomes are selected; *States that fulfill biological functions are attractor states for function and hence for evolution and development.***

This agrees with ([Ginsburg and Jablonka 2019]:245)

> "*Biological attractors are usually functional - the mechanisms enabling them to be reached reliably, in spite of different starting conditions, evolved by natural selection*".

This is the process exploration and selective stabilisation mechanism that is described in ([Ginsburg and Jablonka 2019]:119-123,247-248). The driving of the process by biological needs is the reason that convergent evolution occurs [McGhee 2011].

## 5.3 The brain and stochasticity

There are various kinds of stochasticity in brain function, apart from the fact that it involves necessarily the stochasticity in molecular dynamics just discussed.

**Stochasticity in Neural Activity** The neural code is spike chains [Rieke *et al* 1999] where [Shadlen and Newsoms 1994] the timing of successive action potentials is highly irregular. Also fluctuations in cortical excitability occur [Stephani *et al* 2020]. This results in stochasticity in neural outcomes [Glimcher 2005] in contrast to deterministic dynamics, suggesting an organising principle [Stephani *et al* 2020]:

> "*Brain responses vary considerably from moment to moment, even to identical sensory stimuli. This has been attributed to changes in instantaneous neuronal states determining the system's excitability. Yet the spatio-temporal organization of these dynamics remains poorly understood. .... criticality may represent a parsimonious organizing principle of variability in stimulus-related brain processes on a cortical level, possibly reflecting a delicate equilibrium between robustness and flexibility of neural responses to external stimuli.*"

This stochasticity allows higher level organising principles such as attractors to shape neural outcomes in decision making in the brain [Rolls and Deco 2010] [Deco *at al* 2009]. The higher level structure of attractor networks [Rolls 2016]:95-134) determines outcomes. A particular case where a randomisation and selection process is used is Boltzmann machines and annealing ([Churchland and Sejnowski 2016]:89-91). This demonstrates the principle that *stochasticity greatly enhances efficiency in reaching attractor states* [Palmer 2020].

**Creativity** A key feature of mental life is creativity, which has transformed human life both through inventiveness in science (Maxwell, Turing, Bardeen, Townes, Cormack, and so on) and in commerce (Gates, Jobs, Zuckerberg, Bezos, and so on). It has been proposed ([Rolls 2016]:137) that the possibility of creativity is an outcome of stochasticity due to random spiking of neurons, resulting in a brain state being able to switch from one basin of attraction to another.

**The gut-brain axis** The body microbiome (Section 5.1) has a key influence on the brain. Effects are as follows [Cryan *et al* 2019]

> "*The microbiota and the brain communicate with each other via various routes including the immune system, tryptophan metabolism, the vagus nerve and the enteric nervous system, involving microbial metabolites such as short-chain fatty acids, branched chain amino acids, and peptidoglycans. Many factors can influence microbiota composition in early life, including infection, mode of birth delivery, use of antibiotic medications, the nature of nutritional provision, environmental factors, and host genetics. At the other extreme of life, microbial diversity diminishes with aging. Stress, in particular, can significantly impact the microbiota-gut-brain axis at all stages of life. Much recent work has implicated the gut microbiota in many conditions including autism, anxiety, obesity, schizophrenia, Parkinson's disease, and Alzheimer's disease.*"

It is also involved in neurodegenerative disease [Rosario Iet al 2020]. Because of the unpredictability of how the microbiome will develop, both due to the stochasticity of its genetic mutation and the randomness of the microbes imported from the environment,

the specific outcomes of these interactions are unpredictable from initial micro biological data in an individual body, and hence *a fortiori* from knowledge of the details of the underlying physical level. One can however study the statistics of molecular evolution over the mutational landscape [Gillespie 1984] [Kauffman and Levin 1987].

Note that not all the key factors determining outcomes are purely microbiological or physiological: stress, a mental state, is a key factor in its dynamics.

## 5.4 Neural Plasticity and Neural Darwinism

As well as changing neural network weights [Churchland and Sejnowski 2016] via gene regulatory networks[Kandel 2001], neural plasticity during development involves pruning connections that were initially made randomly [Wolpert *et al* 2002] as learning takes place.

An important way variation and selection happens is via **Neural Darwinism** (or Neuronal Group Selection) [Edelman 1987] [Edelman 1993]. This is a process where neural connections are altered by neuromodulators such as dopamine and serotonin that are diffusely spread from precortical nuclei to cortical areas via 'ascending systems'. They then modify weights of all neurons that are active at that time, thus at one shot strengthening or weakening an entire pattern of activation - a vary powerful mechanism. This mechanism ([Ginsburg and Jablonka 2019]:119-123,247-248) was also discovered by [Changeux and Danchin 1976]. [Seth and Baars 2005] describe these processes thus:

> "*In the brain, selectionism applies both to neural development and to moment-to-moment functioning. Edelman postulates two overlapping phases of developmental and experiential variation and selection. The first is the formation during development of a primary repertoire of many neuronal groups by cell division, migration, selective cell death, and the growth of axons and dendrites. This primary repertoire of neurons is epigenetically constructed through a suite of genetic and environmental influences, and generates a high level of diversity in the nascent nervous system. The second, experiential, phase involves the dynamic formation from this primary repertoire of a secondary repertoire of functional neuronal groups, by the strengthening and weakening of synapses through experience and behavior. This phase involves the selective amplification of functional connectivities among the neurons produced in the first phase, with which it overlaps. In this manner, an enormous diversity of anatomical and functional circuits is produced*"

This provides a key mechanism for experientially based selection of connectivity patterns.

**Primary Emotions** An important feature of Edelman's theory is that the subcortical nuclei involved, as well as the neuromodulators, are precisely the same as are involved in Jaak Panksepp's primary emotional systems [Panksepp 2009] (Section 3.3). Hence the theory is in fact a theory of ***Affective Neural Darwinism*** [Ellis and Toronchuk 2005] [Ellis and Toronchuk 2013], making clear the importance of affect (emotion) for brain plasticity and learning.

**STB2 Stochasticity and Neural Darwinism** ***Brain plasticity is affected by neuromodulators diffusely projected to the cortex from nuclei in subcortical arousal system via ascending systems, selecting neuronal groups for strengthening or weakening. In this way emotions affect neural plasticity.***

This is a key way that interactions at the social level reach down to alter brain connections and hence the context of physical interactions in the brain. It is a specific example of the 'vary and select' topdown process (Section 2.5: **TD3B**) that plays a key role in all biology.

## 5.5 Agency, Self-Causation, and Causal Closure

Agency clearly takes place at the psychological level. People plan and, with greater or lesser success, carry out those plans [Gray and Bjorklund 2018], thus altering features of the physical world. In this way, technological developments such as farming and metallurgy and abstract ideas such as the design of an aircraft or a digital computer have causal power [Ellis 2016] and alter history [Bronowski 2011].

The emergent psychological dynamics of the brain demonstrably has real causal powers. So how does such agency occur?

**Self-causing systems** Agency is centrally related to the idea of a *self-causing system*. The idea of a *system* is crucial, "an integration of parts into an orderly whole that functions as an organic unity" ([Juarrero 2002]:108-111). This enables self-causation ([Juarrero 2002]:252):

> "*Complex adaptive systems exhibit true self-cause: parts interact to produce novel, emergent wholes; in turn these distributed wholes as wholes regulate and constrain the parts that make them up".*

([Murphy and Brown 2007]:85-104) develop the theme further, emphasizing firstly how a complex adaptive system represents the emergence of a system with a capacity to control itself. Secondly, agency is related to the variation and selection process emphasized here: the dynamical organisation of a complex adaptive system functions as an internal selection process, established by the system itself, that operates top-down to preserve and enhance itself. This process is an example of the interlevel causal closure that is central to biology [Mossio 2013] [Ellis 2020b]. It leads to the circularity of the embodied mind [Fuchs 2020]:

> "*From an embodied and enactive point of view, the mind-body problem has been reformulated as the relation between the lived or subject body on the one hand and the physiological or object body on the other. The aim of the paper is to explore the concept of circularity as a means of explaining the relation between the phenomenology of lived experience and the dynamics of organism-environment interactions. .. It will be developed in a threefold way:*
>
> (1) *As the circular structure of embodiment, which manifests itself (a) in the homeostatic cycles between the brain and body and (b) in the sensorimotor cycles between the brain, body, and environment. This includes the interdependence of an organism's dispositions of sense-making and the affordances of the environment.*
>
> (2) *As the circular causality, which characterizes the relation between parts and whole within the living organism as well as within the organism-environment system.*
>
> (3) *As the circularity of process and structure in development and learning. Here, it will be argued that subjective experience constitutes a process of sense-making that implies (neuro-)physiological processes so as to form modified neuronal structures, which in turn enable altered future interactions.*

*On this basis, embodied experience may ultimately be conceived as the integration of brain-body and body-environment interactions, which has a top-down, formative, or ordering effect on physiological processes."*

This is also related to the Information Closure theory of consciousness [Chang *et al* 2020].: *"We hypothesize that conscious processes are processes which form non-trivial informational closure (NTIC) with respect to the environment at certain coarse-grained scales. This hypothesis implies that conscious experience is confined due to informational closure from conscious processing to other coarse-grained scales."*

**The predictive coding view** Intentional action is a process of agent selection from various possibilities. This possibility of agency is congruent with the predictive coding view, as three examples will demonstrate. Firstly [Seth *et al* 2012] state,

"*We describe a theoretical model of the neurocognitive mechanisms underlying conscious presence and its disturbances. The model is based on interoceptive prediction error and is informed by predictive models of agency, general models of hierarchical predictive coding and dopamine signaling in cortex ...The model associates presence with successful suppression by top-down predictions of informative interoceptive signals evoked by autonomic control signals and, indirectly, by visceral responses to afferent sensory signals. The model connects presence to agency by allowing that predicted interoceptive signals will depend on whether afferent sensory signals are determined, by a parallel predictive-coding mechanism, to be self-generated or externally caused."*

Secondly [Negru 2018] puts it this way:

"*The aim of this paper is to extend the discussion on the free-energy principle (FEP), from the predictive coding theory, which is an explanatory theory of the brain, to the problem of autonomy of self-organizing living systems. From the point of view of self-organization of living systems, FEP implies that biological organisms, due to the systemic coupling with the world, are characterized by an ongoing flow of exchanging information and energy with the environment, which has to be controlled in order to maintain the integrity of the organism. In terms of dynamical system theory, this means that living systems have a dynamic state space, which can be configured by the way they control the free-energy. In the process of controlling their free-energy and modeling of the state space, an important role is played by the anticipatory structures of the organisms, which would reduce the external surprises and adjust the behavior of the organism by anticipating the changes in the environment. In this way, in the dynamic state space of a living system new behavioral patterns emerge enabling new degrees of freedom at the level of the whole."*

Finally [Szafron 2019] characterizes it thus:

*" Using the Free Energy Principle and Active Inference framework, I describe a particular mechanism for intentional action selection via consciously imagined goal realization, where contrasts between desired and present states influence ongoing neural activity/policy selection via predictive coding mechanisms and backward-chained imaginings (as self-realizing predictions). A radically embodied developmental legacy suggests that these imaginings may be intentionally*

*shaped by (internalized) partially-expressed motor predictions and mental simulations, so providing a means for agentic control of attention, working memory, and behavior."*

The overall result is a final higher level organising principle that acts as an attractor state during evolution:

**STB3 Stochasticity and Agency** *Stochasticity at the micro level allows macro level dynamics to select preferred micro outcomes, thus freeing higher levels from the tyranny of domination by lower levels. By this mechanism, downward selection of preferred micro outcomes enables self-causation and agency.*

The big picture is that randomness is rife in biology. Evolutionary processes have adapted biological systems to take advantage of this [Hoffmann 2012], with higher level processes selecting preferred outcomes from a variety of possibilities at the lower levels, thereby enabling the higher level organising principles characterised in the previous two sections to shape physical outcomes [Noble and Noble 2018]. The underlying physics enables this, but does not by itself determine the particular outcomes that occur, for they are contextually determined via time dependent dynamical constraints ([Juarrero 2002]:131-162).

# 6 The Whole Universe Gambit and Causal Closure

The hardcore reductionist responds to the previous sections by saying yes the brain is an open system, but the universe as a whole is not. Extend your micro data to include all the particles in the universe - well, in the region of the universe that is causally connected to us (i.e inside the particle horizon [Hawking and Ellis 1973]) - and it is then causally closed. All the data incoming to the brain is determined by causally complete microphysical processes in this extended domain, hence brain outcomes are determined by them too.

The response, denying that this can work, has many levels. Please note that as stated before I am concerned with the possibility of physics determining specific outcomes, such as the words in Carlo's emails, not just statistical outcomes. His emails did not contain a statistical jumble of letters or words: they contained rational arguments stated coherently. This is what has to be explained. The question is how the underlying physics relates to such specific rational outcomes.

In order of increasing practical importance the issues are as follows.

First, Section 6.1 denies that the micro physical level is in fact causally complete, because of irreducible quantum indeterminism. While this can indeed have an effect on the brain, its primary importance is to deny that physics at the micro level is in principle causally complete.

Second, Section 6.2 makes the case that even if the incoming data was determined uniquely by microphysics everywhere in the surroundings, they would not determine a unique brain micro state in any individual because of the multiple realisability of macro states by microstates (§2.6).

Third, Section 6.3 makes the case that important aspects of macro physics are in practice indeterminate because of the combination of chaotic dynamics and the impossibility of specifying initial data to infinite precision. This has neural outcomes *inter alia* because it applies to weather patterns and forest fires.

Fourth, Section 6.4 points out that there is considerable randomness in the external world biology that the mind interacts with at both micro and macro levels. These biological outcomes are not precisely predictable from their micro physics initial data. It has key impacts on the mind related in particular to the relations between humans and viruses.

Fifth, Section 6.5 points out that because the brain is a social brain (§3.4), its macro level responses to incoming data are not purely mechanical: they are highly sophisticated responses at the psychological level to social interactions. These are affected by unpredictable effects such as weather and pandemics. By the mechanisms discussed in Section 4, these understandings reach down to structure the neural context of brain microphysics.

Finally Section 6.6 makes the case that the larger environment interacts with the brain by providing the setting for interlevel circular causation. This can be claimed to be the real nature of causal closure. It is what is involved in order to have the data, constraints, and boundary conditions needed to determine specific outcomes in real world contexts, enabling self-causation. This is the opposite of being determined by microphysics alone.

## 6.1 Is micro physics causally complete?

Carlo's argument is that micro data dependence of all outcomes undermines the possibility of strong emergence. To summarise, suppose I am given the initial positions $\mathbf{r}_i$ and momenta $\mathbf{p}_i$ of all particles in the set $\mathcal{P}$ everywhere,[11] where

$$\mathcal{P} := (\text{protons, neutron, electrons}) \tag{5}$$

at a foundational level **L1**. At a higher level **L2** this constitute an emergent structure **S**, such as a neural network. The details of **S** are determined by the microdata, even though its nature cannot be recognised or described at level **L1**. The forces between the particles at level **L1** completely determine the dynamics at level **L1**. Hence the emergent outcomes at level **L2** are fully determined by the data at level **L1**, so the emergence of dynamical properties and outcomes at level **L2** must be weak emergence and be predictable, at least in principle, from the state (5) of level **L1**, even if carrying out the relevant computations is not possible in practice. This would apply equally to physical, engineering, and biological emergent systems. It is in effect a restatement of the argument from supervenience.

There are problems with the argument just stated as regards both microphysics and macrophysics.

**Quantum physics uncertainty relation** The Heisenberg uncertainty relations undermine this Laplacian dream because initial data cannot be specified with arbitrary accuracy [Heisenberg 1949]. The standard deviations of position $\sigma_x$ and momentum $\sigma_p$ obeys

$$\sigma_x \sigma_p \geq \hbar/2 \tag{6}$$

[Kennard 1927], so one cannot in principle set initial data precisely at level **L1**. Consequently, outcomes based on standard Lagrangians dependent on $x$ and $p$ are uncertain in principle.

Essentially the same issue arise in the case of classical physics [Del Santo and Gisin 2019] because data cannot be prescribed to infinite accuracy [Ellis *et al* 2018]. Further in the quasi-classical approximation, it will be subject to the uncertainty (6), reinforcing that conclusion [Del Santo 2020]. This affects outcomes of chaotic dynamics (§6.3).

**Irreducible uncertainty in quantum outcomes** There is irreducible uncertainty of quantum outcomes when wave function collapse to an eigenstate takes place, with outcomes only predictable statistically via the Born rule [Ghirardi 2007]. One cannot for

[11] "Everywhere" means within the particle horizon [Hawking and Ellis 1973].

example predict when an excited atom will emit a high energy photon, one can only predict the probability of such an event This unpredictability has consequences that can get amplified to macrolevels, for example causing errors in computer memories due to cosmic rays [Ziegler and Lanford 1979] [Gorman *et al* 1996]. The specific errors that occur are not determined by physics, because quantum physics is foundationally causally incomplete.

**Biological damage due to cosmic rays** Cosmic rays can alter genes significantly enough to cause cancer. In particular, galactic cosmic rays lead to significant fatality risks for long-term space missions. This is discussed in [Cucinotta and Cacao 2017] [Cucinotta *et al* 2017] [Cekanaviciute *et al* 2018] This shows both the contextual dependence of local outcomes in this case, and their unpredictability in principle.

**Unpredictable brain effects** This obviously can affect the mental processes of those undertaking space travel. The brain can be affected crucially by distant events that are in principle unpredictable because they result from quantum decay of excited atoms.

The statistics of outcomes is strictly predicted by quantum theory. But in terms of causal completeness of biological events, we wish to know which specific person gets affected at what specific time, thereby changing individual thought patterns. Detailed microphysical initial data everywhere cannot tell us that.

This is a situation that only affects a small number of people, but it is important because it establishes that *in principle* the physicalist whole world gambit does not work (after all, that argument is an in principle argument: no one argues that it can work in practice in terms of allowing actual predictions of unique biological outcomes).

## 6.2 Mental states and multiple realisability

Incoming sensory data in a real world context affects brain macrostates which then shape micro level connections via learning. But they do not do so in a unique way: incoming sensory data does determine unique brain microstates because of the multiple realisability of higher level states at the physical level.

**Mental states and multiple realisability** A given set of incoming data does not result in a unique brain physical microstate because of multiple realisability of the higher level state at the lower level (Section 2.6). This is a key property of brain function. [Silberstein and McGeever 1999] state

> *"Functionalists (and others) claim that mental states are radically multi-realizable, i.e., that mental states like believing that p and desiring that p can be multiply realized within individual across time, across individuals of the same species, across species themselves, and across physical kinds such as robots. If this is true, it raises crucial question: why do these states always have the same behavioural effects? In general we expect physically similar states to have similar effects and different ones to have different effects. So some explanation is required of why physically disparate systems produce the same behavioural effects. If there is nothing physically common to the 'realizations' of a given mental state, then there is no possibility of any uniform naturalistic explanation of why the states give rise to the same mental and physical outcomes."*

The brain responds to incoming data via the predictive processing mechanisms discussed in Section 3.2 , with updating of the relevant expectations taking place all the time on the basis of experience. The macro psychological processes that occur in this way reach down to shape neural network connections and weights [Kandel 2001] [Kandel 1998] in ways that are not unique. These then change constraints at the electron/ion level, realising any one of the billions of possible changes at that level that are in the right equivalence class.

**Unpredictable brain effects** Unique micro level physical conditions in the brain (the specific details of constraints in the electron/ion Lagrangian that will determine the ongoing brain dynamics) cannot in principle be determined by incoming data from the external world because of multiple realisability. Ordered outcomes appear at the brain macro level according to the predictive coding logic outlined in Section 5.3, which then activates any one of the microstates in the corresponding equivalence class at the micro level (Section 2.6). All of this can of course be traced at the microphysical level, both internal to the brain and externally. But what is driving it is psychological level understandings.

## 6.3 Is macro physics causally complete?

The atmosphere is an open system dynamically driven by the Sun's radiation, and with vary complex interactions taking place between the atmosphere, subject to winds and convection, water (the seas and lakes and clouds and ice), and land [Ghil and Lucarini 2020]. These are unpredictable in detail because of chaotic dynamics.

**Convection patterns** Consider a higher physical level **L3** in the context of a fluid where convection patterns take place. Because of the associated chaotic dynamics together with the impossibility (6) of setting initial data to infinite precision (Section 6.1), macroscopic outcomes are unpredictable in principle from micro data (5) . Convection patterns are an example [Bishop 2008]: an extremely small perturbation in a fluid trapped between two levels where a heat differential is maintained can influence the particular kind of convection that arises. [Anderson 2001] puts it this way:

> "*A fluid dynamicist when studying the chaotic outcome of convection in a Benard cell knows to a gnat's eyelash the equations of motion of this fluid but also knows, through the operation of those equations of motion, that the details of the outcome are fundamentally unpredictable, although he hopes to get to understand the gross behaviour. This aspect is an example of a very general reality: the existence of universal law does not, in general, produce deterministic, cause-and-effect behaviour.*"

The outcome is an emergent layer of unpredictability at both local scales (thunderstorms, tornados, typhoons, and so on) and globally (large scale weather outcomes). The latter are famously characterised by strange attractors [Lorenz 1963], involving instability and fractals, but much more importantly interactions between different length scales that make prediction impossible in principle [Lorenz 1969], as discussed in depth by [Palmer *et al* 2014]. Downward constraints then entrains lower level dynamics to follow, as stated by [Bishop 2012]: "*Large-scale structures arise out of fluid molecules, but they also dynamically condition or constrain the contributions the fluid molecules can make, namely by modifying or selecting which states of motion are accessible to the fluid molecules".*

**The Butterfly Effect** Lorenz intended the phrase 'the butterfly effect' to describe the existence of an absolute finite-time predictability barrier in certain multi-scale fluid systems, implying a breakdown of continuous dependence on initial conditions for large enough forecast lead times [Palmer *et al* 2014]. [Lorenz 1969] states

> "*It is proposed that certain formally deterministic fluid systems which possess many scales of motion are observationally indistinguishable from indeterministic systems. Specifically, that two states of the system differing initially by a small observational error will evolve into two states differing as greatly as randomly chosen states of the system within a finite time interval, which cannot be lengthened by reducing the amplitude of the initial error.*"

This happens because of the interactions between the different length scales involved. Palmer's illuminating paper [Palmer *et al* 2014] concludes that this real butterfly effect[12] does indeed occur - but only for some particular sets of initial data. Nevertheless occurring from time to time denies causal closure of physics on this scale in practice. That clearly means the underlying physics at the particle level cannot have been causally closed either. The multiscale weather dynamics studied by [Lorenz 1969] reaches down to influence atomic and electron motions (think thunderstorms) at the lower physics level. But this is the crucial point: you cannot predict when those cases will occur.

**Forest Fires** are an example of self-organised critical behaviour [Malamud *et al* 1998] affected by local atmospheric convection activity in that firstly many forest fires are cased by lightning[13], and secondly the spread of the fire is determined by local winds which are changed by local convection effects due to the fire. The detailed dynamics of the fire are unpredictable because of these links; even probabilities are tricky [Mata *et al* 2010].

**Unpredictable brain effects** In terms of the effect on the brain, these random outcomes shape decisions from whether to open an umbrella on a trip to the shops, to farmers' decisions as to when to harvest crops, aircraft pilots decisions about *en route* flight planning, and homeowners decisions about whether or not to flee a forest fire. It causes an essential unpredictability in mental outcomes. This is a first reason the external world has an ongoing unpredictable key effect on individual brains.

## 6.4 Biological Randomness: the Microbiome

Biological dynamics in the external world are subject to unavoidable uncertainty because of the random nature of molecular level events, already alluded to in Section 5.1.

**Interacting microbiomes and viruses** The immense complexity of each individual person's microbiome (Section 5.1) interacts, through social events, with other people's microbiomes, as do their viruses. Genetic variability is central to the mutation of microbes and viruses in the external world. Detailed physical microstates everywhere determine the statistics of such variations, but not the specific ones that actually occur, which are due *inter alia* to mutation and recombination and horizontal gene transfer in the case of microbes, mistakes by RNA or DNA polymerases, radiation- or chemical- or host cell defences-induced mutation, and re-assortment in the case of viruses. Predicting mutations

[12] See https://www.youtube.com/watch?v=vkQEqXAz44I for an enlightening lecture on the common and real butterfly effects. The implication is that you need ensemble forecasts.

[13] Dry Thunderstorms Could Accelerate the California Wildfires,

is essentially impossible, even for viruses with $10,000$ bases like HIV. All you CAN say is that the known mutation rate for that organism predicts that every single copy of the HIV genome (for example) will have at least one mutation ($10^{-4}$ rate).[14]

**Unpredictable brain effects** This has crucial effects in our brains that are completely unpredictable because firstly of the randomness of the genetic mutations leading to these specific microbes and viruses, and secondly of the details of the events that lead to their spread through animal and human populations; this can all be expressed in terms of rugged adaptive landscapes [Orr 2005]. This firstly directly affects human health and brain dynamics in each of the set of interacting brains (§6.5) via the gut brain axis [Cryan *et al* 2019], and then plays a key role in individual associated mental events such as individual planning of what do to about anxiety, obesity, schizophrenia, Parkinson's disease, and Alzheimer's disease, or flu, AIDS or COVID-19. This is a second reason the external world has an important unpredictable key effect on individual brains.

## 6.5 Social understandings and individual brains

Our brain is a social brain (Section 3.4). Information from the external world affects mental states via ongoing complex social interactions, which have real causal powers. They structure our mental activities in everyday life.

**Social understandings** There is an intricate relation between the individual and society [Berger 1963] [Longres 1990] [Berger and Luckmann 1991] [Donald 2001] and between individuals and institutions in a society [Elder-Vass 2010] [Elder-Vass 2012]. The downward effect of the social context on an individual brain is mediated by social interactions and understandings (Section 3.4). In this social context, a complex interaction takes place involving mind reading, prediction, filling in of missing data, taking social context into account [Longres 1990] [Donald 2001] [Frith 2009]. This nature of the interactions of a many brains, each a self causing open system (Section 5.5), is the main practical day to day reason that microphysics everywhere cannot determine unique outcomes in each of the brains involved. Downward causation from the social level interactions to individual brains to the underlying molecular biology and thence physical levels is the causal chain.

**Abstract variables have causal powers** This is all enabled by our symbolic ability [Deacon 1997], resulting in our use of spoken and written language, which is the key factor enabling this to happen [Ginsburg and Jablonka 2019] (Section 3.5). This affects our individual brain operations as we consider the continually changing detailed implications of money, closed corporations, laws, passports, and so on in our lives.

**Policy decisions have causal powers** Given this context, social variables have causal power [Longres 1990] [Harari 2014] and affect brain states; in particular, this applies to policy decisions. The interaction outcomes are shaped at the social level, which is where the real causal power resides, and then affect individual brain states in a downward way. Complex interpretative processes take place shaping psychological level reactions, which then shape neural network and synapse level changes in a contextual way on an ongoing basis as studied by social neuroscience [Cacioppo *et al* 2002].

[14] I thank Ed Rybicki for this comment.

**Unpredictable brain effects** Policy decisions are sometimes based in unpredictable events such as cyclones or forest fires (Section 6.3) or a global pandemic or local infectious outbreak (Section 6.4). Mandatory evacuating of towns in the face of a cyclone or wild fire,[15] going into shelters in the case of a tornado, or policy decisions such as lockdowns in the face of a pandemic will all be unpredictable because their cause is unpredictable, and so will cause unpredictable outcomes in individual brains at macro and micro levels. The causal chain is an unpredictable trigger event, followed by a social policy choice that then changes outcomes in individual brains. Detailed physical data everywhere enables this to happen by providing the basis for stochastic outcomes that cannot be determined uniquely from that data be because of the real butterfly effect in the case of weather, and its analogue in the case of microbe and viral mutations. Carlo's view that microphysics determines all brain dynamics in this extended context could hold if it were not for the random nature of the trigger events.

## 6.6 Real Causal Closure

Carlo's move of bringing into focus the larger context is certainly correct in the following sense: the way that causal closure takes place in reality involves the whole environment. But that means it is an interlevel affair, for the environment involves all scales.

**The real nature of causal closure**

- From my viewpoint, what is meant by the phrase "causal closure" as used by Carlo and other physicists is in fact that one is talking *about existence of a well-posed effective theory* $\mathbf{EF_L}$ that holds at some emergent level $\mathbf{L}$. This means data $d_{\mathbf{L}}$ for variables $v_{\mathbf{L}}$ at that level $\mathbf{L}$ specifies unique outcomes, or the statistics of such outcomes, at that level.

- However existence of such a theory does by itself not determine any specific physical outcomes. It implies that *if* the right data and boundary conditions are present, *and* all constraints that hold are specified, *then* a unique or statistical outcome is predicted by the physics at that level.

- It does not attempt to say where that data, boundary conditions, and constraints come from. But without them you do not have causal closure in what should be taken to be the real meaning of the term: *sufficient conditions are present to causally determine real world outcomes that happen.* For example social dynamics are active causal factors that reach down to affect physics outcomes, as is abundantly clear in the COVID-19 crisis: policies about face masks affect physical outcomes.

- My use of the term, as developed in full in [Ellis 2020b], regards causal closure as interlevel affair, such as is vital to biological emergence [Mossio 2013]. The conjunction of upward and downward effects must self-consistently determine the boundary conditions, constraints, and initial data at a sufficient set of levels that unique or statistical outcomes are in fact determined by the interlocking whole.

- When that happens you can of course trace what is happening at whatever physics level you choose as a base level **L0**. But over time, the later initial data, boundary conditions, and constraints at that level are dynamically affected by the downward mechanisms outlined in Section 2.5. Because causation is equally real at each level,

[15] For a typical context see https://www.nytimes.com/2020/08/20/us/ca-fires.html.

the higher levels are just as much key factors in the causal nexus as is level **L0**, as time proceeds. Higher Level Organising Principles, independent of the lower level physics, shape physical outcomes. The state of variables at level **L0** at time $t_0$ uniquely determines the higher levels at that time, but not at a later time $t_1 > t_0$.

- The freedom for higher levels to select preferred lower level outcomes exists because of the stochastic nature of biological processes at the cellular level (Section 5.2).
- The illusion of the effective theory at a physical level **L0** being causally complete is because physicists neglect to take into account their own role in the experiments that establish the validity of the effective theory that holds at that level. When you take that role into account, those experiments involve causal closure of all levels from **L0** to the psychological level **L6** where experiments are planned and the social level **L7** which enables the experimental apparatus to come into being.

Another term used for causal closure in this sense is *operational closure*: the organisational form of the processes that enable autopoietic self-production and conservation of system boundaries [Di Paolo and Thompson 2014] [Ramstead *et al* 2019].

**The predictive coding/free energy viewpoint** My view agrees with the growing predictive coding consensus, as presented in previous sections. Karl Friston (private communication) says the following:

> "*I imagine that downward causation is an integral part of the free energy formalism; particularly, its predication on Markov blankets. I say this in the sense that I have grown up with a commitment to the circular causality implicit in synergetics and the slaving principle (c.f., the Centre Manifold Theorem in dynamical systems). As such, it would be difficult to articulate any mechanics without the downward causation which completes the requisite circular causality. Practically, this becomes explicit when deriving a renormalisation group for Markov blankets. We use exactly the same formalism that Herman Haken uses in his treatments of the slaving principle [Haken 1996] [Haken and Wunderlin 1988] to show that Markov blankets of Markov blankets constitute a renormalisation group. If existence entails a Markov blanket, then downward causation (in the sense of the slaving principle) must be an existential imperative.*"

The final conclusion of this section is the following

> **Unique causal outcomes in individual microphysical brain states do not occur when one includes causal effects of the external world.** ***This does not work (i) because microphysics is not in fact causally closed due to quantum wave function collapse, (ii) external information cannot uniquely determine microphysical states in the brain - multiple realisability makes this impossible, (iii) unpredictable macro level chaotic dynamics occurs, (iv) microbiome dynamics that affect brain states is unpredictable, and (v) the way external states influence brain states is strongly socially determined and can include events that are in principle unpredictable.***

However it certainly is true that such downward causal effects on individual brains occur. They just do not do so in a way that is uniquely determined by physical effects alone.

# 7 Microphysics Enables but Does Not Determine

In this section I summarise my response (Section 7.1), and comment on the relation of all the above to the issue of free will (Section 7.2) and to the possibility spaces that are the deep structure of the cosmos (Section 7.3).

## 7.1 The basic response

There are a series of key issues that shape my response.

- I am concerned with what determines the specific outcomes that occur in real world contexts, not just with statistical prediction of outcomes. How does physics underlie the existence of a Ming dynasty teapot? Of the particular digital computer on which I am typing this response? Of Einstein's publication of his paper on General Relativity? Of the election of Donald Trump as President of the United States of America?

- Consider a specific individual brain at a particular time. The difference between synchronic and diachronic emergence is key. Carlo's view can be defended in the synchronic emergence case, but cannot be correct in the diachronic case, because individual brains are open systems. The initial microphysical state of the brain simply does not include all the data that determine its outcomes at later times.

This is what is discussed in depth in Sections 3 - 5.

**The whole universe context** Claiming that this problem is solved by going to a larger scale where causal closure does indeed hold (the cosmological scale), which therefore implies that the specific evolution of all subsystems such as individual brains is also uniquely determined, does not work for a series of reasons. I list them now with the theoretically most important issues first. This is the inverse of the order that matters in terms of determining outcomes in practical terms. As far as that is concerned, the most important issues are the later ones.

It does not work because of,

1. Irreducible quantum uncertainty at the micro level which affects macro outcomes; this demonstrates that the claim is wrong in principle. It can indeed have macro effects on the brain, but this is not so important at present times because of the shielding effect of the earth's atmosphere. However it has played important role in evolutionary history [Percival 1991], as discussed by [Todd 1994] [Scalo and Wheeler 2002]

2. The fact that downward effects from that larger context to the brain, which certainly occur via neural plasticity and learning, cannot in principle determine a unique brain microstate, because of multiple realisability of those detailed physical states when this occurs. Unique brain microstates cannot occur in this way.

3. Uncertainty in principle at the ecosystem level due to chaotic dynamics and the real butterfly effect plus the inability to set initial molecular conditions precisely. This has major unpredictable effects on individual brains due to forest fires, thunderstorms, tornadoes, and tropical cyclones.

4. Microbiome dynamics that is in principle unpredictable because of the molecular storm and huge number of molecules involved, plus the inability to set initial molecular conditions precisely. This affects individual and social outcomes as evolution

takes place on a rugged adaptive landscape that keeps changing as all the interacting species evolve. This crucially affects brain dynamics through the gut-brain axis.

5. Social understandings that shape how external signals are interpreted by the brain, when social level policies and choices (which may involve unpredictable events such as thunderstorm details or pandemic outbreaks) chain down to influence flows of electrons in axons. It simply is not a purely physics interaction.

Carlo's vision of the external world as a whole evolving uniquely and thereby determining unique brain states because they are a part of the whole, may work in some contexts where irreducible uncertainty 1. and effective uncertainty 3. and 4. do not occur. It cannot however work when any of these effects come into play, which certainly happens in the real world. This demonstrates that as a matter of principle, it is the higher level effects - psychological and social variables - that are sometimes calling the tune. But that means they are always effectively doing so in the social context which is the habitat of minds.

**Causal closure** Real world causal closure is an interlevel affair, with microphysical outcomes determined by features ranging from global warming and tropical cyclones to COVID-19 policy decisions. It simply cannot occur at the microphysics level alone. Some of the effective variables which have changed human history are abstract concepts such as the invention of arithmetic, the concept of money, and the idea of a closed corporation. These have all crucially affected microphysical outcomes, as have abstract theories such as the theory of the laser and the concepts of algorithms, compilers, and the internet.

Overall, as stated by [Bishop 2012], the situation is that

> *"Whatever necessity the laws of physics carry, it is always conditioned by the context into which the laws come to expression."*

So in response to Carlo's final email (Section 1.1):

**CR** Today the burden of the proof is not on this side. Is on the opposite side. Because:

**(i)** There is no single phenomenon in the world where microphysics has been proven wrong (in its domain of validity of velocity, energy, size...).

**GE** The view I put respects the microphysics completely. Of course it underlies all emergent phenomena, without exception. Microphysics certainly is not wrong.

**CR** **(ii)** By induction and Occam's razor, is a good assumption that in its domain validity it holds.

**GE** Yes it holds in its domain of validity, which is at the microscale. The question at stake is, How much larger is its domain of validity? I think the comment is meant to say that its domain of validity includes biology and the brain, in the sense that, by itself it fully determines all biological and brain outcomes.

However completely new kinds of behaviour emerge in the biological domain. The kind of causation that emerges is simply different than the kind of statistically deterministic relation between data and outcomes that holds at the microphysical level. The microphysics allows this emergence: it lies within the space of possibilities determined by that physics. In that sense the higher level outcomes lie within the domain of validity of the microphysics. But the microphysics by itself does not determine the macro level outcomes (see the listed points above). Occam's razor does not work.

**CR (iii)** There are phenomena too complex to calculate explicitly with microphysics. These provide no evidence against (ii), they only testify to our limited tools.

**GE** Carlo only considers efficient causation, because that is what physicists study. As Aristotle pointed out [Bodnar 2018], that is only one of the four kinds of causation that occur in the real world (Section 2.5). In the real world the other kinds of causation play a key role in determining outcomes. All four are needed to determine specific outcomes.

I accept the need to provide the burden of proof. I have done so in the preceding sections.

## 7.2 What about Free Will?

The implication of Carlo's argument is that the causal power of microphysics prevents the existence of free will. This touches on a vast and complex debate. My arguments above deny that the underlying physics can disprove existence of free will in a meaningful sense. But that does not dispose of the debate. Does neurobiology/neuroscience deny free will?

**Incomplete reductionism: neurobiology** Francis Crick gives a neuroscience based reductionist argument regarding the brain in *The astonishing hypothesis* [Crick 1994]:

> *"You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. "*

Now the interesting point is that this a denial of Carlo's arguments. Crick is assuming that the real level of causation is at the cellular and molecular biology levels: that is where the action is, it is at that level that physical outcomes are determined. The implication is that this is what determines what specific dynamical outcomes take place at the underlying physical level - which is my position [Ellis and Kopel 2019].

**Free will and neurobiology** As to free will itself, does neurobiology and neuroscience deny its existence? That is a long and fraught debate related to intentionality and agency. Amongst the deeply considered books that argue for meaningful free will are [Donald 2001] [Dupré 2001] [Murphy and Brown 2007] [Murphy *et al* 2009] [Baggini 2015]. [Frith 2013] has a nuanced discussion on agency and free will. [Murphy and Brown 2007] conclude (page 305) that "*free will should be understood as being the primary cause of one's own actions; this is a holistic capacity of mature, self-reflective human organisms acting within suitable social contexts*". This is essentially the consensus of the authors just named, Libet's experiments notwithstanding. Chris Frith [16] expresses it this way [Frith 2009]:

> " *I suggest that the physiological basis of free will, the spontaneous and intrinsic selection of one action rather than another, might be identified with mechanisms of top-down control. Top-down control is needed when, rather than responding to the most salient stimulus, we concentrate on the stimuli and actions relevant to the task we have chosen to perform. Top-down control is particularly relevant when we make our own decisions rather than following the instructions of an experimenter. Cognitive neuroscientists have studied top-down control extensively and have demonstrated an important role for dorsolateral prefrontal cortex and anterior cingulate cortex. If we consider the individual in isolation,*

[16] One of the topmost cited neuroscientists in the world: he had 203,843 citations on 2020/08/01.

*then these regions are the likely location of will in the brain. However, individuals do not typically operate in isolation. The demonstration of will even in the simplest laboratory task depends upon an implicit agreement between the subject of the experiment and the experimenter. The top of top-down control is not to be found in the individual brain, but in the culture that is the human brain's environmental niche*"

This is a good description of both topdown effects in the brain [Ellis 2018] and interlevel causal closure [Ellis 2020b]. It is also expressed well by [Baggini 2015].

**Free will denialists don't really believe it** The physicist Anton Zeilinger told me the following story. He was once being harassed by someone who strongly argued that we do not have free will. Anton eventually in frustration reached out and slapped him in the face. He indignantly shouted, "*Why did you do that?*", to which Anton responded "*Why do you ask me that question? You have just been explaining to me at length that I am not responsible for my actions. According to you, it's not a legitimate question.*"

If you have an academic theory about the nature of causation and free will, it must apply in real life too, not just when you are engaged in academic argumentation. If not, there is no reason whatever for anyone else to take it seriously - for you yourself do not.

**Free will and the possibility of science** The ultimate point is that if we don't have meaningful free will, in the sense of the possibility of making rational choices between different possible understandings on the basis of coherence and evidence, then science as an enterprise is impossible. You then cannot in a meaningful way be responsible for assessing theories anyone proposes. The theory that free will does not exist causes the demise of any process of scientific investigation that is alleged to lead to that theory. We had better find a better theory - such as those in the books cited above.

**Denial of consciousness or qualia** Finally one should note that many pursuing the view that free will does not exists also deny that consciousness and/or qualia exist and play any role in brain function. But neuroscience simply does not know how to solve the hard problem of consciousness [Chalmers 1995]. As stated in [Tallis 2016], neuroscience helps define the necessary conditions for the existence of human consciousness and behaviour, but not the sufficient conditions. The self-defeating philosophical move of denying that consciousness and/or qualia exist ("what one canot explain does not exist" [Tallis 2016]) does not succeed in explanatory terms: how can you deny something if you have no consciousness? In that case, you do not satisfy the necessary conditions to deny anything: you do not exist in any meaningful sense [Donald 2001]. For more on this see also [Gabriel 2017] [Dennett and Strawson 2018].

But in any event this is a different debate than my debate with Carlo, which in the end is about physics denying free will.

## 7.3 Possibility Spaces

All the biological effects discussed here are allowed by the underlying physical levels: they do not violate or alter those generic equations, which apply to all physical interactions without exception.

A useful way to characterise this is in terms of *possibility spaces*. In the case of physics, these include phase spaces [Arnold 1989] (classical physics) and Hilbert spaces (quantum

physics) [Isham 2001]. These are determined by the laws of physics, and are indeed equivalent to them: the laws allow the possibilities described by the possibility spaces, and the possibility spaces characterise the nature of the underlying laws.

Now the interesting point is that there are also biological possibility spaces. At the microbiology level they include a space of all possible proteins allowed by the laws of physics [Wagner 2014], of which only some have been realised on Earth by evolutionary processes [Petsko and Ringe 2009]. Similarly there are sets of possible genotype-phenotype maps for metabolism and for gene regulation [Wagner 2014]. There are limitations on physiological possibilities due to the nature of physics [Vogel 2000] and consequent scaling laws for biology [West *et al* 1997]. These are all immutable biological possibilities that, just like the laws of physics, are the same everywhere in the universe at all times, and allow life as we know it to come into existence in suitable habitats.

The claim one can make then is that Higher Level Organising principles such as those that are identified in this paper for biology in general and for the brain in particular are also of this nature: they are timeless and eternal principles that can be expected to apply to life everywhere, because this is the only way it can work in principle. Thus these would include the possibility of metabolic networks, genes and gene regulatory networks [Wagner 2014]; of physiological systems [Rhoades and Pflanzer 1989] [Hall 2016] and developmental processes [Wolpert *et al* 2002]; of Evo-Devo type evolution [Carroll 2008]; and of social brains [Dunbar 1998] operating on hierarchical predictive coding principles [Clark 2013], that are capable of analytical thought based in a symbolic capacity [Deacon 1997].

These are all possible, as they do indeed exist, so the possibility of their existence is written into the nature of things. This is allowed by the underlying physics, but they represent higher order principles allowing life to come into existence and flourish, as discussed in depth by Stuart Kauffman inhos book *At Home in the Universe* [Kauffman 1995].

**Acknowledgments** I thank Carlo Rovelli for his patient dialogue with me regarding this issue. It is a pleasant contrast with the arrogant dismissive comments and *ad hominem* contemptuous personal attacks that are common in some reductionist circles and writings.

I thank Karl Friston, Tim Palmer, and Ed Rybicki for helpful comments.

# References

[Alberts 2007] B Alberts, A Johnson, J Lewis, M Raff, K Roberts, and P Walter (2007) *Molecular Biology of the Cell* (Garland Science)

[Alon 2006] Alon, Uri (2006). *An introduction to systems biology: design principles of biological circuits* (CRC Press).

[Anderson 2001] Anderson, Philip W (2001) "Science: A 'dappled world' or a 'seamless web'?" *Stud. Hist. Phil. Mod. Phys.* **32**:487-494.

[Arnold 1989] V I Arnol'd (1989) *Mathematical Methods of Classical Mechanics* (Springer)

[Baggini 2015] Baggini, Julian (2015). *Freedom regained: The possibility of free will.* (University of Chicago Press, 2015.

[Bateson and Laland 2013] Bateson, Patrick, and Kevin N. Laland (2013)
"Tinbergen's four questions: an appreciation and an update." *Trends in ecology and evolution* **28**:712-718.

[Berger 1963] Berger, Peter L (1963)
*Invitation to Sociology: A Humanistic Perspective* (Anchor Books)

[Berger and Luckmann 1991] Berger, Peter L., and Thomas Luckmann (1991)
*The social construction of reality: A treatise in the sociology of knowledge.* (Penguin Uk).

[Berridge 2014] M Berridge (2014)
*Cell Signalling Biology* Portland Press, doi:10.1042/csb0001001
http://www.cellsignallingbiology.co.uk/csb/

[Billman 2020] Billman, George E (2020)
"Homeostasis: The Underappreciated and Far Too Often Ignored Central Organizing Principle of Physiology." *Frontiers in Physiology* **11**:200.

[Bishop 2005] Bishop, Robert C. (2005)
"Patching physics and chemistry together." *Philosophy of Science* **72**:710-722.

[Bishop 2008] Bishop, Robert C (2008).
"Downward causation in fluid convection." *Synthese* **160**: 229-248.

[Bishop 2012] Bishop, Robert C. (2012)
"Fluid convection, constraint and causation." *Interface Focus* **2**:4-12. C

[Bishop and Ellis 2020] Bishop, Robert,and George Ellis (2020)
"Contextual Emergence of Physical Properties" *Foundations of Physics*: 1-30.

[Bodnar 2018] Bodnar, Istvan (2918)
"Aristotle's Natural Philosophy", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.),
¡https://plato.stanford.edu/archives/spr2018/entries/aristotle-natphil/¿.

[Booch 2006] Booch, Grady (2006)
*Object oriented analysis and design with application* (2nd Edition). (Addison Wesley).

[Bronowski 2011] Bronowski, Jacob (2011).
*The ascent of man*. (Random House).

[Buchanan *et al* 2010] Buchanan, Mark, Guido Caldarelli, Paolo De Los Rios, Francesco Rao, and Michele Vendruscolo, eds. (2010)
*Networks in cell biology* (Cambridge University Press).

[Cacioppo *et al* 2002] Cacioppo, John T., Gary G. Berntson, Ralph Adolphs, C. Sue Carter, Martha K. McClintock, Michael J. Meaney, Daniel L. Schacter, Esther M. Sternberg, Steve Suomi, and Shelley E. Taylor, eds. (2002)
*Foundations in social neuroscience.* (MIT press).

[Campbell 1974] Campbell, Donald T. (1974).
"Downward causation in hierarchically organised biological systems" in *Studies in the Philosophy of Biology: Reduction and Related Problems*, eds F. J. Ayala and T. Dobhzansky (Berkeley, CA: University of California Press), 179-186.

[Campbell and Reece 2008] Campbell, Neil A and Jane B Reece (2008)
*Biology* (San Francisco: Benjamin Cummings).

[Carroll 2005] Carroll, Sean B (2005).
*Endless forms most beautiful*. (WW Norton & Company).

[Carroll 2008] Carroll, Sean. B. (2008).
Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1), 25-36.

[Castellani 2002] Castellani, Elena (2002)
"Reductionism, emergence, and effective field theories." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* **33**: 251-267.

[Cekanaviciute *et al* 2018] Cekanaviciute, Egle, Susanna Rosi, and Sylvain V. Costes (2918).
"Central nervous system responses to simulated galactic cosmic rays." *International journal of molecular sciences* **19**:3669.

[Chalmers 1995] Chalmers, David J (1995). "Facing up to the problem of consciousness." *Journal of consciousness studies* **2**:200-219.

[Chalmers 2000] Chalmers, David J. (2006)
"Strong and weak emergence", In P. Davies and P. Clayton, Eds, *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion* (Oxford University Press).

[Chang *et al* 2020] Chang, Acer YC, Martin Biehl, Yen Yu, and Ryota Kanai (2020)
"Information closure theory of consciousness." *Frontiers in Psychology* **11** (2020).

[Chang *et al* 2008] Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. and Huang, S., (2008)
*Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature* **453**:544-547.

[Changeux and Danchin 1976] Changeux, Jean-Pierre, and Antoine Danchin (1976)
"Selective stabilisation of developing synapses as a mechanism for the specification of neuronal networks." *Nature* **264**:705-712.

[Churchland and Sejnowski 2016] Churchland, Patricia S., and Terrence J. Sejnowski (2016). *The computational brain* (MIT Press).

[Clark 2013] Clark, Andy (2016)
"Whatever next? Predictive brains, situated agents, and the future of cognitive science" *Behavioral And Brain Sciences* (2013) **36**: 181-253

[Clark 2016] Clark, Andy (2016)
*Surfing uncertainty: Prediction, action, and the embodied mind.* (New York: Oxford University Press).

[Constant *et al* 2019] Constant, A., Clark, A., Kirchhoff, M., and Friston, K. J. (2019).
"Extended active inference: constructing predictive cognition beyond skulls." *Mind and Language*.

[Crick 1994] Crick, Francis (1994).
*The astonishing hypothesis: the scientific search for the soul* (New York, Touchstone).

[Cryan *et al* 2019] Cryan, John F *et al* (2019)
"The microbiota-gut-brain axis." *Physiological reviews* **99**:1877-2013.

[Cucinotta and Cacao 2017] Cucinotta, Francis A., and Eliedonna Cacao (2017).
"Non-targeted effects models predict significantly higher mars mission cancer risk than targeted effects models" *Scientific reports* **7**:1-11.

[Cucinotta *et al* 2017] Cucinotta, Francis A., Khiet To, and Eliedonna Cacao (2017)
"Predictions of space radiation fatality risk for exploration missions." *Life sciences in space research* **13**:1-11.

[Damasio 2006] Damasio, Antonio R (2006)
*Descartes' error.* (Random House).

[Davies 2019] Davies, Paul (2019).
*The Demon in the Machine*: *How Hidden Webs of Information are Solving the Mystery of Life.* (University of Chicago Press).

[Deacon 1997] Deacon, Terence (1097)
*The Symbolic Species: The Co-Evolution of Language and the Brain* (New York: WW Norton).

[Deco *at al* 2009] Deco, Gustavo, Edmund T. Rolls, and Ranulfo Romo (2009).
"Stochastic dynamics as a principle of brain function." *Progress in neurobiology* **88**:1-16.

[Del Santo 2020] Del Santo, Flavio. "Indeterminism, causality and information: Has physics ever been deterministic?." arXiv preprint arXiv:2003.07411 (2020).

[Del Santo and Gisin 2019] Del Santo, Flavio, and Nicolas Gisin (2019)
"Physics without determinism: Alternative interpretations of classical physics."
*Physical Review* **A 100**:6 062107.

[Dennett and Strawson 2018] Daniel C. Dennett, Galen Strawson (2018)
"Magic, Illusions, and Zombies": An Exchange *New York Review of Books,* April 3, 2018.

[Di Paolo and Thompson 2014] Di Paolo, Ezequiel Alejandro, and Evan Thompson (2014)
"The enactive approach." *The Routledge handbook of embodied cognition*:68-78.

[Donald 1991] Donald, Merlin (1991)
*Origins of the modern mind: Three stages in the evolution of culture and cognition.* (Harvard University Press).

[Donald 2001] Donald, Merlin (2001)
*A mind so rare: The evolution of human consciousness.* (WW Norton and Company).

[Dunbar 1998] Dunbar, Robin I M (1998)
"The social brain hypothesis." *Evolutionary Anthropology* **6**:178-190.

[Dunbar 1998a] Dunbar, Robin (1998)
"Theory of mind and the evolution of language." *Approaches to the Evolution of Language*:92-110.

[Dunbar 2003] Dunbar, Robin I M (2003)
"The social brain: mind, language, and society in evolutionary perspective." *Annual review of anthropology* **32**:163-181.

[Dunbar 2014] Dunbar, Robin (2014).
*Human evolution: A Pelican introduction*. (Penguin UK).

[Dupré 2001] Dupré, John (2001).
*Human nature and the limits of science*. (Clarendon Press, Oxford).

[Edelman 1987] Edelman, Gerald M (1987)
*Neural Darwinism: The theory of neuronal group selection.* (Basic books).

[Edelman 1993] Edelman, Gerald M (1993)
"Neural Darwinism: selection and reentrant signaling in higher brain function." *Neuron* **10**: 115-125.

[Eden *et al* 2011] Eran Eden *et al* (2011)
"Proteome Half-Life Dynamics in Living Human Cells" *Science* **331**:764-768

[Elder-Vass 2010] Elder-Vass, Dave (2010)
*The Causal Power of Social Structures* (Cambridge University Press).

[Elder-Vass 2012] Elder-Vass, Dave (2012)
"Top-down causation and social structures." *Interface focus* **2**:82-90.

[Ellis 2005] Ellis, George (2005)
"Physics, Complexity, and Causality" *Nature* **435**:743.

[Ellis 2012] Ellis, George F. R. (2012) "Top-down causation and emergence: some comments on mechanisms. *Interface Focus* **2**:126 -140.

[Ellis 2016] Ellis, George (2016)
*How can physics underlie the mind? Top-down causation in the human context* (Springer-Verlag, Heidelberg.)

[Ellis 2018] Ellis, George (2018)
"Top-down effects in the brain" *Physics of life reviews* **31**: 1-30.

[Ellis 2019] Ellis, George (2019)
"Why reductionism does not work" Essay for the Kurt Gödel prize and references.

[Ellis 2020a] Ellis, George F R (2020)
"Emergence in Solid State Physics and Biology'
*Foundations of Physics* DOI 10.1007/s10701-020-00367-z 2020
http://arxiv.org/abs/2004.13591.

[Ellis 2020b] Ellis, George F R (2020)
"The Causal Closure of Physics in Real World Contexts"
*Foundations of Physics* DOI 10.1007/s10701-020-00366-0
https://arxiv.org/abs/2006.00972

[Ellis and Drossel 2019] Ellis, George, and Barbara Drossel (2019)
"How Downwards Causation Occurs in Digital Computers." *Foundations of Physics* **49**: 1253-1277. https://arxiv.org/pdf/1908.10186

[Ellis and Kopel 2019] Ellis, George and Jonathan Kopel (2019)
"The dynamical emergence of biology from physics" *Frontiers in physiology* **9**: 1966.

[Ellis *et al* 2018] Ellis, George FR, Krzysztof A. Meissner, and Hermann Nicolai (2018).
"The physics of infinity." *Nature Physics* **14**: 770-772.

[Ellis and Toronchuk 2005] Ellis, George FR, and Iudith A. Toronchuk (2005)
"Neural development Affective and immune system in?uences." In *Consciousness and Emotion: Agency, conscious choice, and selective perception* Ed, R D Ellis and N Newton (John Benjamins):81.

[Ellis and Toronchuk 2013] Ellis, George FR, and Judith A. Toronchuk (2013).
"Affective neuronal selection: the nature of the primordial emotion systems." *Frontiers in psychology* **3**: 589.

[Felsenstein 1988] Felsenstein, Joseph (1988). "Phylogenies from molecular sequences: inference and reliability." *Annual review of genetics* **22**:521-565.

[Fink and Noble 2008] Fink, Martin and Denis Noble (2008)
"Noble Model" Scholarpedia, 3(2):1803

[Friston 2003] Friston, Karl (2003)
"Learning and inference in the brain". *Neural Networks* **16**:1325-1352.

[Friston 2010] Friston, Karl (2010).
"The free-energy principle: a unified brain theory?." *Nature reviews neuroscience* **11**:127-138.

[Friston 2012] Friston, Karl (2012).
"A free energy principle for biological systems." *Entropy* **14**:2100-2121.

[Friston *et al* 2009] Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. (2009)
"Reinforcement learning or active inference?" PLoS One **4**:e6421.

[Frith 2007] Frith, Chris D (2007).
"The social brain?." *Philosophical Transactions of the Royal Society* **B362**: 671-678.

[Frith 2009] Frith, Chris D (2009)
"Free will top-down control in the brain." in [Murphy *et al* 2009]:199-209.

[Frith 2013] Frith, Chris (2013)
*Making up the mind: How the brain creates our mental world.* (John Wiley and Sons).

[Frith and Frith 2005] Frith, Chris, and Uta Frith (2005).
"Theory of mind." *Current biology*

[Fuchs 2020] Fuchs, Thomas (2020)
'The circularity of the embodied mind.". *Front. Psychol.* **15**:R644-R645.

[Gabriel 2017] Gabriel, Markus (2017).
*I am Not a Brain: Philosophy of Mind for the 21st Century* (John Wiley and Sons).

[Geritz *et al* 1997] Geritz, Stefan AH, Johan AJ Metz, Éva Kisdi, and Géza Meszéna (1997)
"Dynamics of adaptation and evolutionary branching." *Physical Review Letters* **78**:2024.

[Ghil and Lucarini 2020] Ghil, Michael, and Valerio Lucarini (2020)
"The physics of climate variability and climate change." *Rev. Mod. Phys* **92**:035002 (arXiv:1910.00583).

[Ghirardi 2007] Ghirardi, G. (2007).
*Sneaking a Look at God's Cards: Unraveling the Mysteries of Quantum Mechanics* (Princeton University Press).

[Gibb *et al* 2019] Gibb, S., Hendry, R. F., and Lancaster, T. (Eds.) (2019).
*The Routledge Handbook of Emergence*. (Routledge).

[Gilbert 1991] Gilbert, Scott F. (1991)
"Epigenetic landscaping: Waddington's use of cell fate bifurcation diagrams." *Biology and Philosophy* **6**:135-154.

[Gilbert and Epel 2009] Gilbert, Scott F., and David Epel (2009)
*Ecological developmental biology: integrating epigenetics, medicine, and evolution.* (Sinauer Associates)

[Gillespie 1984] Gillespie, John H. (1984)
"Molecular evolution over the mutational landscape." *Evolution*: 1116-1129.

[Gillett 2019] Gillett, Carl (2019).
"Emergence, Downward Causation and its Alternatives: Critically surveying a foundational issue." *The Routledge Handbook of Emergence*. (Routledge), 99-110.

[Ginsburg and Jablonka 2019] Ginsburg, Simona, and Eva Jablonka (2019)
*The evolution of the sensitive soul: Learning and the origins of consciousness*. (MIT Press).

[Glimcher 2005] Glimcher, Paul W (2005)
"Indeterminacy in brain and behavior" *Annual review of psychology* **56**:25-60.

[Gombrich 1961] Gombrich, Ernst Hans (1961).
*Art and illusion* (New York: Pantheon Books).

[Gorman *et al* 1996] O'Gorman, T. J.*et al* (1996).
"Field testing for cosmic ray soft errors in semiconductor memories". *IBM Journal of Research and Development* **40**:41-50.

[Gottschall 2012] Gottschall, Jonathan (2012)
*The storytelling animal: How stories make us human.* (Houghton Mifflin Harcourt).

[Gray and Bjorklund 2018] Gray, Peter and David Bjorklund (2018)
*Psychology* (8th Edition: Macmillan Learning).

[Green and Batterman 2020] Green, Sara, and Robert Batterman (2020).
"Causal slack and top-down causation: Universality and functional equivalence in physics and biology". To appear in *Top-Down Causation and Emergence*, Ed Otávio Bueno (Springer Synthese Library Book Series).

[Haken 1996] Haken, Hermann (1996) "Slaving principle revisited." *Physica D: Nonlinear Phenomena* **97**:95-103.

[Haken and Wunderlin 1988] Haken, H., and A. Wunderlin (1988).
"The Slaving Principle of Synergetics?An Outline." In *Order and Chaos in Nonlinear Physical Systems*, pp. 457-463. (Springer, Boston, MA).

[Hall 2016] Hall, John E (2016)
*Guyton and Hall Textbook of Medical Physiology*. (Elsevier).

[Harari 2014] Harari, Yuval Noah (2014)
*Sapiens: A brief history of humankind*. (Random House).

[Hartwell *et al* 1999] Hartwell, Leland H., John J. Hopfield, Stanislas Leibler, Andrew W. Murray (1999)
"From molecular to modular cell biology." *Nature* **402**: 6761: C47-C52.

[Hawking and Ellis 1973] Hawking, Stephen W, and George F R Ellis (1973)
*The Large Scale Structure of Spacetime* (Cambridge University Press).

[Hawkins and Blakeslee 2007] Hawkins, Jeff, and Sandra Blakeslee (2007)
*On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines* (Macmillan).

[Heisenberg 1949] Heisenberg, Werner (1949)
*The physical principles of the quantum theory* (Courier Corporation).

[Hesp *et al* 2019] Hesp, Casper, Maxwell Ramstead, Axel Constant, Paul Badcock, Michael Kirchhoff, and Karl Friston (2019)
"A multi-scale view of the emergent complexity of life: A free-energy proposal." In *Evolution, Development and Complexity*, pp. 195-227. (Springer, Cham).

[Hodgkin and Huxley 1952] Hodgkin Andrew L, and Andrew Huxley AF (1952)
"A quantitative description of membrane current and its application ...". *The Journal of Physiology* **117**:500-544.

[Hoffmann 2012] Hoffmann, P (2012) *Life's Ratchet: How Molecular Machines Extract Order from Chaos* (Basic Books)

[Hofmeyer 2017] Hofmeyr Jan-Hendrik (2017)
"Basic Biological Anticipation" In R. Poli (ed.), *Handbook of Anticipation* (Springer International Publishing AG)

[Hofmeyer 2018] Hofmeyr, Jan-Hendrik (2018)
"Causation, Constructors and Codes" *Biosystems* **164**:121-127.

[Hohwy 2007] Hohwy, Jacob (2007)
"Functional Integration and the mind." *Synthese* **159**:315-28.

[Hohwy 2013] Hohwy, Jacob(2013)
*The predictive mind*. (Oxford: Oxford University Press).

[Hohwy 2016] Hohwy, Jacob(2013)
"The self-evidencing brain". *Noûs* **50** 259-285.

[Houghton 2009] Houghton, J. (2009).
*Global warming: the complete briefing.* (Cambridge University Press).

[Ingalls 2013] Ingalls, Brian P (2013)
*Mathematical modeling in systems biology: an introduction.* (MIT press).

[Integrative H. M. P. 2014] Integrative HMP Research Network Consortium. (2014)
"The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease". *Cell host and microbe,* **16**:276.

[Integrative H. M. P. 2019] Integrative HMP Research Network Consortium (2019)
"The integrative human microbiome project" *Nature* **569**:641–648

[Isham 2001] Isham, Chris J (2001)
*Lectures on quantum theory Mathematical and structural foundations.* (Allied Publishers).

[Jacob 1974] Jacob, François (1974)
*The Logic of Living Systems* (Albert Lane, London)

[Jacob and Monod 1961] Jacob, François, and Jacques Monod (1961)
"Genetic regulatory mechanisms in the synthesis of proteins." *Journal of molecular biology* **3**:318-356.

[Juarrero 2002] Juarrero, Alicia (2002).
*Dynamics in action: Intentional behavior as a complex system.* (MIT Press).

[Junker & Schreiber 2011] Junker, Björn H., and Falk Schreiber (2011).
*Analysis of biological networks.* (John Wiley and Sons).

[Kandel 1998] Kandel, Eric R (1998).
'A new intellectual framework for psychiatry" *American journal of psychiatry* **155**: 457-469.

[Kandel 2001] E R Kandel (2001) "The molecular biology of memory storage: a dialogue between genes and *Science*:294:1030-1038.

[Kandel 2012] Kandel, Eric R (2012).
*The age of insight: The quest to understand the unconscious in art, mind, and brain, from Vienna 1900 to the present.* (Penguin Random House).

[Kandel *et al* 2013] E Kandel, J H Schwartz, T M Jessell, S A Siegelbaum, and A J Hudspeth (2013)
*Principles of Neural Science* (McGraw Hill Professional).

[Kauffman 1995] Kauffman, Stuart (1995)
*At home in the universe: The search for the laws of self-organization and complexity* (Penguin)

[Kauffman and Levin 1987] Kauffman, Stuart, and Simon Levin (1987)
"Towards a general theory of adaptive walks on rugged landscapes." *Journal of theoretical Biology* **128**:11-45.

[Kennard 1927] Kennard, E. H. (1927)
"Zur Quantenmechanik einfacher Bewegungstypen", *Zeitschrift für Physik* **44**:326-352

[Kirchhoff *et al* 2018] Kirchhoff, Michael, Thomas Parr, Ensor Palacios, Karl Friston, and Julian Kiverstein (2018)
"The Markov blankets of life: autonomy, active inference and the free energy principle." *Journal of The royal society interface* **15**: 20170792.

[Laughlin and Pines 2000] Laughlin, Robert and David Pines (2000)
"The Theory of Everything". *Proc Nat Acad Sci* **97**:28-31.

[Lodish *et al* 2000] Lodish Harvey, Berk A, Zipursky SL, et al (2000). *Molecular Cell Biology* (New York: W. H. Freeman).

[Longres 1990] John E Longres (1990)
*Human behaviour in the social environment* (F E Peacock)

[Lorenz 1963] Lorenz, Edward N (1963)
"Deterministic nonperiodic flow." *Journal of the atmospheric sciences* **20**:130-141.

[Lorenz 1969] Lorenz, Edward N. (1969)
"The predictability of a flow which possesses many scales of motion." *Tellus* **21**:289-307.

[Malamud *et al* 1998] Malamud, Bruce D., Gleb Morein, and Donald L. Turcotte (1998)
"Forest fires: an example of self-organized critical behavior." *Science* **281**:1840-1842.

[Marr 2010] Marr DC: Vision (2010)
*A Computation Investigation into the Human Representational System and Processing of Visual Information.* (San Francisco: Freeman)

[Masel 2011] Masel, Joanna (2011)
"Genetic drift." *Current Biology* **21**: R837-R838.

[Mata *et al* 2010] Mata, Aitor, Bruno Baruque, Belén Pérez-Lancho, Emilio Corchado, and Juan M. Corchado (2010). "Forest fire evolution prediction using a hybrid intelligent system." In *International Conference on Information Technology for Balanced Automation Systems*, pp. 64-71. (Springer, Berlin, Heidelberg).

[Mayr 2001] Mayr, Ernst (2001).
*What evolution is.* (Basic books).

[McGhee 2006] McGhee, George R (2006)
*The geometry of evolution: adaptive landscapes and theoretical morphospaces.* (Cambridge University Press).

[McGhee 2011] McGhee, George R (2011).
*Convergent evolution: limited forms most beautiful.* (MIT Press).

[Menzies 2003] Menzies, Peter (2003)
"The causal efficacy of mental states." *Physicalism and Mental Causation* (Exeter: Imprint Academic): 195-224.

[Miller and Clark 2018] Miller, Mark, and Andy Clark (2018)
"Happily entangled: prediction, emotion, and the embodied mind." *Synthese* **195**:2559-2575.

[Monod 1971] Monod, Jacques (1971)
*Chance and Necessity: Essay on the Natural Philosophy of Modern Biology* (New York, Alfred A. Knopf)

[Mossio 2013] Mossio, Matteo (2013).
*Closure, causal.* In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, H. Yokota.(Eds) *Encyclopedia of Systems Biology* (Springer) pp.415-418.

[Mossio and Moreno 2010] M Mossio and A Moreno (2010)
"Organisational Closure in Biological Organisms" *Hist. Phil. Life Sci.* **32**:269-288.

[Murphy and Brown 2007] Murphy, Nancey, and Warren S. Brown(2007)
*Did my neurons make me do it?: Philosophical and neurobiological perspectives on moral responsibility and free will.* (Oxford University Press).

[Murphy *et al* 2009] Murphy, Nancey, George Ellis, and Tim O'Connor (2009)
*Downward causation and the neurobiology of free will* (Springer, New York).

[Negru 2018] Negru, Teodor (2018)
*Self-organization, Autopoiesis, Free-energy Principle and Autonomy. Organon* **F** **25**:215-243.

[Noble 2008] Noble Dennis (2008)
*The music of life: biology beyond genes*. (Oxford: Oxford University Press).

[Noble 2012] Noble, Denis (2012)
"A theory of biological relativity: no privileged level of causation." *Interface focus* **2**: 55-64.

[Noble 2016] Noble, Denis (2016).
*Dance to the tune of life: Biological relativity.* (Cambridge University Press).

[Noble and Noble 2018] Noble, Raymond, and Denis Noblem(2018).
"Harnessing stochasticity: How do organisms make choices?." *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28**:106309.

[Noble and Noble 2019] Noble, Raymond, and Denis Noble (2019).
"A-Mergence of Biological Systems."in [Gibb *et al* 2019], 387-399.

[Noble *et al* 2019] Noble, Ray, Kazuyo Tasaki, Penelope Noble, and Denis Noble. (2019).
"Biological Relativity requires circular causality but not symmetry of causation: so, where, what and when are the boundaries?" *Frontiers in physiology* **10**:827.

[Nurse 2008] Nurse, Paul (2008).
"Life, logic and information." *Nature* **454**:424-426.

[Orr 2005] Orr, H. Allen (2005)
"The genetic theory of adaptation: a brief history." *Nature Reviews Genetics* **6**:119-127.

[Oyama et al 2001] Oyama, Susan, Paul E Griffiths, and Russell D Gray (2001)
*Cycles of Contingency: Developmental Systems and Evolution* (Cambridge Mass: MIT Press)

[Palmer 2020] Palmer, Tim (2020).
"Human Creativity and Consciousness: Unintended Consequences of the Brain?s Extraordinary Energy Efficiency"? *Entropy* **22**:281.

[Palmer *et al* 2014] Palmer, T. N., A. Döring, and G. Seregin (2014)
"The real butterfly effect." *Nonlinearity* **27**:R123.

[Panksepp 2009] Panksepp, Jaak (2009)
*Affective neuroscience: The foundations of human and animal emotions*. (Oxford university press).

[Panksepp and Biven 2012] Panksepp, Jaak, and Lucy Biven (2012).
*The archaeology of mind: neuroevolutionary origins of human emotions* (WW Norton and Company).

[Peacocke 1989] Peacocke, Arthur R (1989)
*An introduction to the physical chemistry of biological organization* (Oxford University Press, USA).

[Pearl 2009] Pearl, Judea (2009).
*Causality: Models, Reasoning, and Inference*. Cambridge university press, 2009.

[Pearl and Mackenzie 2018] Pearl, Judea, and Dana Mackenzie (2018).
*The book of why: the new science of cause and effect*. (Basic Books).

[Percival 1991] Percival, Ian (1991)
"Schrödinger's quantum cat". *Nature*: **351**:357.

[Peterson *etal* 2009] Peterson, Jane *et al* (2009)
''The NIH human microbiome project." *Genome research* **19**: 2317-2323.

[Petsko and Ringe 2009] Petsko, Gregory A. and Dagmar Ringe (2009)
*Protein structure and function.* (Oxford University Press).

[Piccinini and Shagrir 2014] Piccinini, Gualtiero and Oron Shagrir (2014)
"Foundations of computational neuroscience" *Current Opinion in Neurobiology* **25**:25-30

[Purves 2010] Purves, Dale (2010)
*Brains: how they seem to work*. (Ft Press).

[Purves *et al* 2008] Purves, Dale, Roberto Cabeza, Scott A. Huettel, Kevin S. LaBar, Michael L. Platt, Marty G. Woldorff, and Elizabeth M. Brannon (2008).
*Cognitive neuroscience*. (Sunderland: Sinauer Associates).

[Ramstead *et al* 2019] Ramstead, M. J., Kirchhoff, M. D., Constant, A., and Friston, K. J. (2019)
"Multiscale integration: beyond internalism and externalism" *Synthese* (2019):1-30.

[Ramstead *et al* 2016] Ramstead, Maxwell JD, Samuel PL Veissière, and Laurence J. Kirmayer (2016)
"Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention." *Frontiers in psychology* **7**:1090.

[Rao and Ballard 1999] Rao, Rajesh PN, and Dana H. Ballard (1999).
"Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects." *Nature neuroscience* **2**:79-87.

[Rhoades and Pflanzer 1989] R Rhoades and R Pflanzer (1989)
*Human physiology* (Fort Worth: Saunders College Publishing)

[Rieke *et al* 1999] Rieke, F., Warland, D., Van Steveninck, R., and Bialek, W. S. (1999)
*Spikes: exploring the neural code.* (Cambridge, Mass: MIT press).

[Rolls 2016] Rolls, Edmund T (2016)
*Cerebral cortex: principles of operation.* (Oxford University Press).

[Rolls and Deco 2010] Rolls, Edmund T., and Gustavo Deco (2010).
*The noisy brain: stochastic dynamics as a principle of brain function.*(Oxford: Oxford university press)

[Rosario Iet al 2020] Rosario, D., Boren, J., Uhlen, M., Proctor, G., Aarsland, D., Mardinoglu, A., and Shoaie, S. (2020)
"Systems Biology Approaches to Understand the Host-Microbiome Interactions in Neurodegenerative Diseases". *Frontiers in Neuroscience*, **14**:716.

[Scalo and Wheeler 2002] Scalo, John, and J. Craig Wheeler (2002).
"Astrophysical and astrobiological implications of gamma-ray burst properties." *The Astrophysical Journal* **566**:723.

[Schweber 1993] Schweber, Silvan S (1993)
"Physics, community, and the crisis in physical theory." *Physics Today* **46**: 34-34.

[Scott 2002] Scott, Alwyn (2002).
*Neuroscience: A mathematical primer* (Springer)

[Seth 2013] Seth, Anil K (2013).
"Interoceptive inference, emotion, and the embodied self." *Trends in cognitive sciences* **17**:565-573.

[Seth and Baars 2005] Seth, Anil K., and Bernard J. Baars (2005)
"Neural Darwinism and consciousness." *Consciousness and Cognition* **14**:140-168.

[Seth *et al* 2012] Seth, Anil K., Keisuke Suzuki, and Hugo D. Critchley (2012)
"An interoceptive predictive coding model of conscious presence." *Frontiers in psychology* **2**: 395.

[Shadlen and Newsoms 1994] Shadlen, Michael N., and William T. Newsome (1994)
"Noise, neural codes and cortical organization." *Current opinion in neurobiology* **4**: 569-579.

[Sigalova *et al* 2020] Sigalova, Olga M., Amirreza Shaeiri, Mattia Forneris, Eileen EM Furlong, and Judith B. Zaugg (2020)
"Predictive features of gene expression variation reveal mechanistic link with differential expression." *Molecular Systems Biology* **16**: e9539.

[Silberstein and McGeever 1999] Silberstein, Michael, and John McGeever (1999)
""The search for ontological emergence." *The Philosophical Quarterly* **49**:201-214.

[Simon 2019] Simon, Herbert A (2019)
*The architecture of complexity. Sciences of the artificial.* (MIT press).

[Stephani *et al* 2020] T. Stephani, G. Waterstraat, S. Haufe, G. Curio, A. Villringer and V. V. Nikulin (2020)
"Temporal signatures of criticality in human cortical excitability as probed by early somatosensory responses" *Journal of Neuroscience* JN-RM-0241-20.

bioRxiv (2020): 809285.

[Stevens and Price 2015] Stevens, Anthony, and John Price (2015).
*Evolutionary psychiatry: A new beginning* (Routledge).

[Szafron 2019] Safron, Adam (2019).
"The radically embodied conscious cybernetic Bayesian brain: Towards explaining the emergence of agency". https://psyarxiv.com/udc42/download?format=pdf

[Tallis 2016] Tallis, Raymond (2016)
*The explicit animal: a defence of human consciousness.* (Springer).

[Todd 1994] Todd, P (1994).
"Cosmic radiation and evolution of life on earth: Roles of environment, adaptation and selection." *Advances in Space Research* **14**:305-313.

[Tomasello 2009] Tomasello, Michael (2009).
*The cultural origins of human cognition.* (Harvard university press).

[Toyama and Hetzer 2013] Toyama, B. H. and Hetzer, M. W. (2013).
"Protein homeostasis: Live long, won't prosper". *Nature Reviews Molecular Cell Biology* **14**:55-61.

[Veissiére *et al* 2020] Veissiére, S. P., Constant, A., Ramstead, M. J., Friston, K. J., and Kirmayer, L. J. (2020)
"TTOM in action: Refining the variational approach to cognition and culture". *Behavioral and Brain Sciences* 43.

[Vogel 2000] Vogel, Steven (2000)
*Cats' paws and catapults: Mechanical worlds of nature and people.* (WW Norton and Company).

[Wagner 2014] Wagner, Andreas (2014)
*Arrival of the fittest: solving evolution's greatest puzzle.* (Penguin).

[West *et al* 1997] West, Geoffrey B., James H. Brown, and Brian J. Enquist (1997)
"A general model for the origin of allometric scaling laws in biology." *Science* **276**: 122-126.

[Wiener 1948] Wiener, Norbert (1948).
*Cybernetics or Control and Communication in the Animal and the Machine.* (MIT Press).

[Wolpert *et al* 2002] Wolpert, Lewis, *et al* (2002)
*Principles of development* (Oxford University Press).

[Woodward 2018] Woodward, James. F. (2018)
"Explanation in Neurobiology: An Interventionist Perspective." In *Integrating Psychology and Neuroscience: Prospects and Problems*, Ed. D M Kaplan.

[Wright 1932l] Wright, Sewall (1932)
"The roles of mutation, inbreeding, crossbreeding, and selection in evolution" *Proceedings of the Sixth International Congress on Genetics*. **1** (8): 355-66.

[Ziegler and Lanford 1979] Ziegler, J. F., and Lanford, W. A. (1979).
"Effect of cosmic rays on computer memories." *Science* **206**:776-788.

Version 2020/09/06

# Making sense of top-down causation: Universality and functional equivalence in physics and biology

**Article** · January 2021

CITATIONS
2

READS
157

**2 authors:**

Sara Green
University of Copenhagen
**53** PUBLICATIONS **611** CITATIONS
SEE PROFILE

Robert Batterman
University of Pittsburgh
**49** PUBLICATIONS **2,087** CITATIONS
SEE PROFILE

Book chapter to appear in Voosholz, J. & Gabriel, M. (2021, eds.). *Top-Down Causation and Emergence*. MIT Press.

# Making sense of top-down causation:

# Universality and functional equivalence in physics and biology

Sara Green[1] & Robert W. Batterman[2]

**Abstract**

Top-down causation is often taken to be a metaphysically suspicious type of causation that is found in a few complex systems, such as in human mind-body relations. However, as Ellis and others have shown, top-down causation is ubiquitous in physics as well as in biology. Top-down causation occurs whenever specific dynamic behaviors are realized or selected among a broader set of possible lower-level states. Thus understood, the occurrence of dynamic and structural patterns in physical and biological systems presents a problem for reductionist positions. We illustrate with examples of *universality* (a term primarily used in physics) and *functional equivalence classes* (a term primarily used in engineering and biology) how higher-level behaviors can be multiple realized by distinct lower-level systems or states. Multiple realizability in both contexts entails what Ellis calls "causal slack" between levels, or what others understand as *relative explanatory autonomy*. To clarify these notions further, we examine procedures for upscaling in multi-scale modeling. We argue that simple averaging strategies for upscaling only work for simplistic homogenous systems (such as an ideal gas), because of the scale-dependency of characteristic behaviors in multi-scale systems. We suggest that this interpretation has implications for what Ellis calls *mechanical top-down causation,* as it presents a stronger challenge to reductionism than typically assumed.

**Keywords**: Functional equivalence class; Multiple realizability; Reductionism; Top-down causation; Universality; Constraint

[1] Section for History and Philosophy of Science, Department of Science Education, University of Copenhagen, Niels Bohr Bygningen, Rådmandsgade 64, 2200 Copenhagen, DK. Email: sara.green@ind.ku.dk.

[2] Department of Philosophy, University of Pittsburgh, 1029-H Cathedral of Learning Pittsburgh, PA 15260, USA.

## 1. Introduction

The problem of top-down causation refers to how and whether changes of higher-level variables can have causal effects on lower-level behaviors (Campbell 1974; Ellis 2005; 2008). Top-down causation remains a contested issue in science and philosophy of science alike (Auletta, Ellis and Jaeger 2008). A common assumption is that if macroscale systems consist of "no more than" physical-chemical components, it should be possible to describe higher-level phenomena bottom-up from more fundamental lower-level descriptions (e.g., Oppenheim and Putnam 1958; Crick and Clark 1994; Bedau 1997). This chapter comments on and adds to important insights from George Ellis' work that challenge this assumption. Ellis has been one of the key figures emphasizing that topics such as reductionism and top-down causation are not only philosophically interesting but have important practical implications for science and medicine. We examine further examples in support of this view, by stressing an even stronger interpretation of what Ellis (2012) calls *mechanical top-down causation*.

An important precondition for the existence of top-down causation is that explanations of phenomena at higher scales or levels are (relatively) autonomous of explanations at lower levels. If higher-level explanations and parameters were fully reducible to or derivable from more "fundamental" ones, appeals to top-down causation would be unnecessary or even misleading. In arguing against bottom-up determination of higher-level properties, Ellis (2008, 2012) appeals to the existence of multiple realizability, illustrated through the existence of *equivalence classes* in different scientific domains. In the following, we therefore examine the connections between multiple realizability, equivalence classes, and top-down causation.

Multiple realizability means that a higher-level state or property is realized by different heterogenous states or properties at a lower level. The term is often introduced in discussions about the ontological or explanatory autonomy of higher-level phenomena and models. For instance, Putnam argues against physical reduction of mental states by highlighting that mental kinds are multiple realized by distinct physical kinds (Putnam 1980). Others have appealed to multiple realizability in discussions about explanatory unification (Fodor 1974; Sober 1999; see also Brigandt and Love 2017). However, for the purpose of the discussion of top-down causation, the most important aspect of multiple realizability is that it supports the explanatory autonomy of more general higher-level models that capture similarity in behaviors of heterogenous systems (see also

Batterman 2000; 2018). Ellis connects the issues of multiple realizability and top-down causation as follows:

> Top-down causation takes place owing to the crucial role of context in determining the outcomes of lower level causation. Higher levels of organization constrain and channel lower level interactions, paradoxically thereby increasing higher level possibilities. A key feature here is *multiple realizability* of higher level functions, and consequent existence of *equivalence classes* of lower level variables as far as higher level actions are concerned. An equivalence class identifies all lower level states that correspond to the same higher level state. (Ellis 2012, p. 128, emphasis added)

We unpack with further examples the claims that i) multiple realizability supports explanatory autonomy of higher-level features, and ii) top-down causation can be interpreted as the effects of higher-level constraining relations that determine outcomes of lower-level causation.

Equivalence classes are also sometimes called *universality classes* in physics. Both concepts highlight how systems with distinct microstructures often display general or universal patterns of behavior. Describing these behaviors does not require reference to microscale details – in fact, generic models and explanations are often identified through procedures that abstract from or selectively leave out irrelevant details (Batterman 2000; Green and Jones 2016). We illustrate the relation between multiple realizability and universality through the example of thermodynamics near critical points in Section 2.1. We then examine what Ellis terms *functional equivalence classes* in biology, exemplified through feedback control (Section 2.2). Functional equivalence can be interpreted as an instance of universality that applies only to engineered and living systems, since equivalent behaviors here are characterized in functional terms such as information, robustness, homeostasis, control, etc. (Ellis 2008; 2012). Functional equivalence classes typically are more context-dependent than classical examples of universality in physics, a point we shall elaborate on further below.

For multiple realizability (or conditional independence) to be possible, Ellis holds, there must be *causal slack* between lower and higher levels (Ellis 2012). The notion of causal slack usefully highlights how the explanatory autonomy of higher levels is justified by empirical demonstrations of *conditional independence* of upper-level behavior on many lower-level details (see also Woodward, forthcoming and this volume). The term implies that the autonomy is relative to certain conditions that hold for a given equivalence class. Just like a sail can be slack within

certain length limits of the sheets, so do relations of conditional independence hold within certain boundaries or parameter spaces. Hence, the autonomy of macro-level explanations is not absolute, but relative (see also Batterman 2018). Critics may argue that this threatens the explanatory autonomy of higher-level explanations, and hence the possibility of top-down causation. Instead, we believe that the notion of relative autonomy avoids many problems afflicting strong accounts of top-down causation (such as difficulties in understanding how different levels are connected), and therefore also offers a better description of how scientists develop multi-scale models in practice.

The emphasis on *relative* explanatory autonomy parallels Ellis' distinction between his account of top-down causation and a "stronger" interpretation in which top-down causation is described as efficient causation operating across levels. A strong account has been criticized for giving rise to the problem of causal overdetermination or to mysterious cause-effect relations, which would violate the lower-level laws of physics (Kim 1998; 2000). However, if we interpret top-down effects as higher-level *constraining-relations* on the possible lower-level states of a given system, top-down causation becomes a matter of understanding how higher levels define the *boundary conditions* of lower-level dynamics (see e.g., Ellis 2008; 2012; 2016; Green 2018; Moreno and Mossio 2015; Mossio and Moreno 2010). Constraints are here understood as physical conditions that limit the degree of freedom of a dynamic process, thus enabling only selected system states (Christiansen 2000; Hooker 2013; Juarrero 1998; 1999). Constraints are typically regarded as being at a higher spatial scale than the entities and operations of the constrained processes. In biology, constraints are often further defined according to functional levels in a hierarchically organized system (Pattee 1971; 1973; Salthe 1985; Wimsatt 2007). For instance, the shape and size of blood vessels enable efficient circulation by limiting the degrees of freedom of liquid motion, or blood flow. The interpretation of top-down causation as the ability of upper-level variables to set the context for lower-level ones may be seen as a "weaker" form of top-down causation (Emmeche et al. 2000). Yet, it allows for an understanding of how constraints productively can channel system states that are not possible to reach for an unconstrained system.

The productive aspect of constraints can be illustrated by how an open respiratory system would not be able to provide sufficient gas exchange for a large organism. Similarly, the constraints provided by a sail on wind flow enable a sailboat to move. When one increases the drag by trimming the sail, one does not (effectively) change the operating cause (the wind). Rather, one modifies the structural constraints that channel a pressure difference across the windward and leeward side of the sail. Constraints thus have causal power by delimiting the space of possibilities

for lower-level causes. Without appeals to top-down constraints, we would not be able to explain why specific states are realized among multiple possible lower-level states and through such selections give rise to emergent properties.

Ellis distinguishes between several types of top-down causation. The most basic form of top-down causation is *mechanical* top-down causation, exemplified by how the rigid boundaries of a gas container constrains the degree of freedom of the lower-scale movement of gas molecules (see also Christiansen 2000). Elsewhere, Ellis also refers to this form as *algorithmic* top-down causation, because top-down causation can be understood mathematically as the effects of boundary conditions on the solution to equations describing lower-level dynamics (Ellis 2012). A similar account is defended by the systems biologist Denis Noble in the context of multi-scale cardiac modeling (Noble 2012; 2017; see also Emmeche et al. 2000; Green 2018).

Whereas physical or chemical systems can exhibit *mechanical* or *algorithmic* top-down causation, living systems display multiple additional types such as non-adaptive information control, adaptive selection, and intelligent top-down causation (Ellis 2008; 2012). These are often considered as stronger forms of top-down causation, because biological functions must be understood through goals of whole organisms and species, which again depends on higher-level features such as the environmental and evolutionary background (Ellis and Kopel 2017). These "stronger" types of top-down causation specific to biological systems are not the focus of our chapter.[3] Rather, our aim is to show that the ideal of "bottom-up reductionism" (Gross and Green 2017) can also be challenged with examples of *mechanical top-down causation*. Our chapter responds to a common view or concern that in the contexts of physics and chemistry, "it is not always clear whether traditional reductionist point of view is actually overcome, since these [high-level variables] can again be understood as a complicated effect of more elementary processes" (Auletta, Ellis and Jaeger 2008, p. 1162). We draw on examples from both physics and biology to argue that higher-level variables used to model many multi-scale systems are, in fact, not reducible in the sense often assumed. Hence, we think that mechanical top-down causation presents a stronger challenge to reductionism than typically assumed.

The analysis is structured as follows. After clarifying the concept of universality in physics and relating it to Ellis' notion of equivalence class, we further elaborate on Ellis' suggestion that

[3] Readers interested in these types of downward causation, as well as debates on the metaphysical implications of downward causation, may find Paoletti and Orilia's (2017) comprehensive anthology on downward causation interesting. For examples of downward causation in ecology, see also (Allen and Star 1982; Ulanowicz 1986; 1997).

feedback control and network motifs are examples of functional equivalence classes (Section 2). We then compare top-down causation based on information control to mechanical or algorithmic top-down causation (Section 3). Drawing on examples of multi-scale modeling in physics and biology, we argue that high level variables used to describe heterogeneous systems cannot be derived from micro-scale details through coarse-graining. We end the chapter with reflections on the practical implications of the autonomy of scales and top-down causation for science and medicine (Sections 4 and 5).

## 2. Universality and functional equivalence

One way to expresses the challenge to the reductionist is to ask: "how can systems that are heterogenous at the microscale exhibit the same pattern of behavior at the macro-scale?" (Batterman 2018, p. 861). We can answer this question by demonstrating that details of the heterogenous realizers are to a large extent explanatorily irrelevant, and thus that we are justified in idealizing (and thereby effacing) many lower-scale details when our explanatory target is at higher scales. Similarly, Ellis highlights that if an upper-level behavior is multiply realized, we do not have to appeal to micro-level details but can explain higher-level patterns through the generic characteristics of the equivalence class. For Ellis, the features explaining the characteristics of equivalence classes are higher-level constraining relations that *channel similar outputs in heterogenous systems*. Such effects can be interpreted as instances of top-down causation (Ellis 2012; Section 3).

### *2.1. Universality and multiply realizability in physics*

A paradigmatic example of universality is found in thermodynamic behavior near critical points. Various fluids consisting of different chemical elements will have different critical temperatures and pressures. That is to say, they will undergo so-called continuous phase transitions at pressures and temperatures that depend upon the micro details of the molecules. However, the behavior of many different fluids at the critical point of phase transitions are identical and can be characterized by the same critical exponents. During a phase transition, e.g., when water boils in a pot, the densities of the liquid water and the vapor (steam) will differ. And, in fact, there will be regions of

liquid and regions of vapor that coexist in the pot. If one plots the difference in the densities of the liquid and the vapor one notices that as the temperature approaches a critical value, this difference exhibits power law scaling behavior. Remarkably, when one plots this behavior for a certain class of fluids in dimensionless (reduced) coordinates $(\frac{\rho}{\rho_c}, \frac{T}{T_c})$, one can show that they all exhibit the same scaling behavior. See Figure 1 for a dramatic display of this behavior. Thus, there is a universal property (shared behavior at macroscales by systems distinct at microscales) for systems near their respective critical points. More remarkable still is the fact that systems like magnets near criticality also display very similar coexistence curves. (For magnets the order parameter is the net magnetization[4] but the scaling exponent is identical.) Part of the reason for this universal behavior is the fact, noted explicitly by Kadanoff (1971) that the closer the system is to criticality, the less the macroscopic/continuum properties depend on the dynamical details of the system.

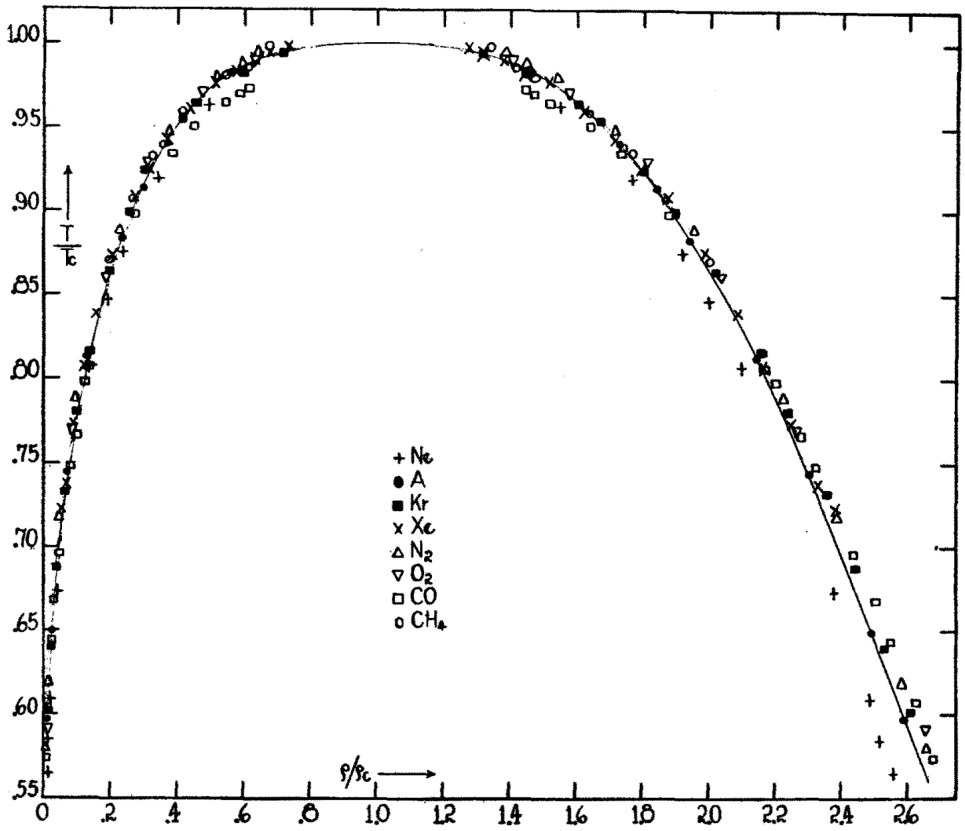


Figure 1. Vapor-liquid coexistence curve for various fluids. The figure shows the difference in densities at temperatures in reduced coordinates. At $(\frac{\rho}{\rho_c}, \frac{T}{T_c})$ = (1,1) liquid phase (left) and vapor phase (right) have the same density, thus their

[4] The net magnetization can also be understood as a difference in densities. The densities of up-spins vs. down-spins. This difference vanishes at the critical temperature, as the high temperature randomizes the directions of the spins.

difference vanishes. Reprinted from Guggenheim, Edward A. 1945. "The principle of corresponding states." *The Journal of Chemical Physics* 13, 7, 253-261, with the permission of AIP Publishing

If we wish to understand why many different systems exhibit such similar behaviors, we are unlikely to find the answer in reductionist appeals to fundamental force laws for each chemical component. Rather, the answer is given by the so-called renormalization group explanation of the universality of critical phenomena (henceforth RG explanation). This explanation takes advantage of the fact that near criticality systems exhibit self-similar, fractal like, behavior. Thus, one can introduce a transformation on the space of Hamiltonians that throws away details via a kind of coarse-graining. Repeated application of this transformation eliminates details that genuinely distinguish the different systems from one another. The hope is to find a fixed point of the transformation from which one can determine the value for the scaling exponent. All systems that evolve under this transformation to the fixed point for the universality class of systems exhibit the same macroscopic scaling behavior. The RG explanation thus extracts structural features that stabilize macroscopic phenomena irrespective of changes in or perturbations of microscopic details. This example from physics not only gives a concrete interpretation of multiple realizability, but also allows for a better understanding of why some very simple models (Ising models, for example) can be used in quite varied contexts to help explain the behaviors of real systems (fluids and magnets) (Batterman 2000; 2018).

Phase transitions illustrate the emergence of new macroscopic features through the different characteristics of liquid water, steam, and crystalline ice. They involve discontinuous alterations in higher-level behaviors through the influences of higher-level variables (such as temperature and pressure) on lower-level interactions (Ellis 2016, 139). In Ellis' view, environmental variables triggering phase transitions should be interpreted as coarse-grained higher-level variables, because temperature and pressure cannot be attributed to isolated molecules. Like order parameters, these point to collective properties arising in a constrained system, such as a gas container (see also Christiansen 2000). Yet, pressure and temperature are interpreted as coarse-grained because they can be identified through averaging of lower-level details (see Section 3). Accordingly, phase transitions of this type represent the most basic (or weak) form of top-down causation (Ellis 2016, pp. 224-225). We shall return to this point in Section 3, after examining some examples from biology for comparison.

### *2.2. Functional equivalence and information control in biology*

Ellis uses the term *functional equivalence class* when referring to models in the life sciences that identify correspondence (or equivalence) of lower-level states or systems with respect to the corresponding higher-level variables and behavior (Ellis 2008; 2012). Examples of these are recurring network motifs and feedback control, which exemplify multiple realizability in biology. We further unpack examples from systems biology that support this view.

A hallmark of living systems is *homeostasis*, i.e., the ability of organisms to maintain a relatively stable internal environment despite external perturbations (Bernard 1927/1957; Cannon 1929). The robustness of functional steady states in organisms is typically explained with reference to feedback control, a concept imported to biology from engineering in the 1920s and later formalized by the mathematician Norbert Wiener (Wiener 1948). Wiener's book was ground-breaking in suggesting that the same mathematical models can be used to describe feedback control in very different oscillatory systems, from electrical circuits to metabolic regulation in different organisms. The hope for generic systems principles was highlighted also in Rachevsky's mathematical biology and Bertalanffy's general system theory (Green and Wolkenhauer 2013). In recent years, systems biology has further strengthened and elaborated on this view by using generic models from control theory and graph theory to describe so-called *organizing* or *design principles* in living systems (Alon 2007; Green 2015).

An example of a design principle is integral feedback control, which is used to explain robust perfect adaptation in bacterial chemotaxis. The example is described in further detail in other publications (Green and Jones 2016; Serban and Green, 2020), and we shall here focus only on why integral feedback control can be seen as an instance of multiple realizability. An important question in biology is how various functions are maintained despite environmental perturbations. For instance, biologists are interested in understanding how motile bacteria can detect changes in the concentrations of nutrients or toxins in their environment and optimize their movements according to these. Remarkably, chemotactic bacteria have receptor systems that can detect and respond to concentration changes in their environments with the same precision before and after stimulus. In engineering terms, the receptor system is said to display *robust perfect adaptation* (RPA), i.e., the system will return to its pre-stimulus value and regain sensitivity over a large range of parameter values (Alon et al. 1999). Achieving this kind of robustness is a hard problem in engineering. Engineers are often interested in designing systems that asymptotically track a fixed steady-state

value, so as to maintain system function despite noisy input signals (or changes in initial conditions).

In engineering, robust adaptation to pre-stimulus steady-state values can be achieved through a design principle called integral feedback control (IFC). IFC refers to a quantifiable feedback relation in which the difference between the desired output (steady-state activity) and the actual output is fed back to the system as the integral of the system error. Strikingly, the mathematical description of bacterial receptor systems has been found to be equivalent to formal models of integral feedback control in engineering. Systems biologist John Doyle and colleagues derived the principle through a mathematical analysis in which they reduced a mechanistically detailed model of the receptor system to a generic description involving only relations sufficient for the higher-level property of robust perfect adaptation (Yi et al. 2000).[5] As a result, IFC was proposed as a *design principle* that generically constrains functional behaviors and enables robust perfect adaptation, regardless of the causal details of the heterogenous systems realizing this capacity. In Ellis' terms, we can say that there is sufficient "causal slack" between higher-level behaviors and lower-level realizers to allow for the *same principle* to apply to systems as different as bacteria and engineered thermostats. We shall comment further on this example below.

Similar examples of multiple realizability in biology are so-called network motifs in gene regulatory networks (Ellis 2012; Fang 2020). Network motifs are small sub-circuits of regulatory connections that have been found to be frequent in biological regulatory networks and have been hypothesized to display characteristic generic functions (Alon 2007).[6] For instance, a so-called coherent feedforward loop, cFFL, has been shown to implement a sign-sensitive delay of outputs in response to input signals. This function was first demonstrated mathematically through a simple Boolean input function (see Figure 2), and the hypothesized function was subsequently confirmed experimentally in living bacteria (Mangan et al. 2003). In biological systems, it can function as a persistence detector that can filter out noisy input signals, such as brief fluctuations in the concentration of nutrients available in a bacteria's environment. This ensures that protein synthesis of metabolizing enzymes is only activated when the activating signal (availability of sugars) is persistent (for further details, see Alon 2007).

[5] This aspect is analyzed in further detail in (Green and Jones 2016).

[6] Investigations of the stabilizing aspects of global constraints in networks have been explored much earlier, e.g., by Stuart Kauffman's demonstrations of how the structure of Boolean networks constrains the possible network states (Kauffman 1969; 1993).

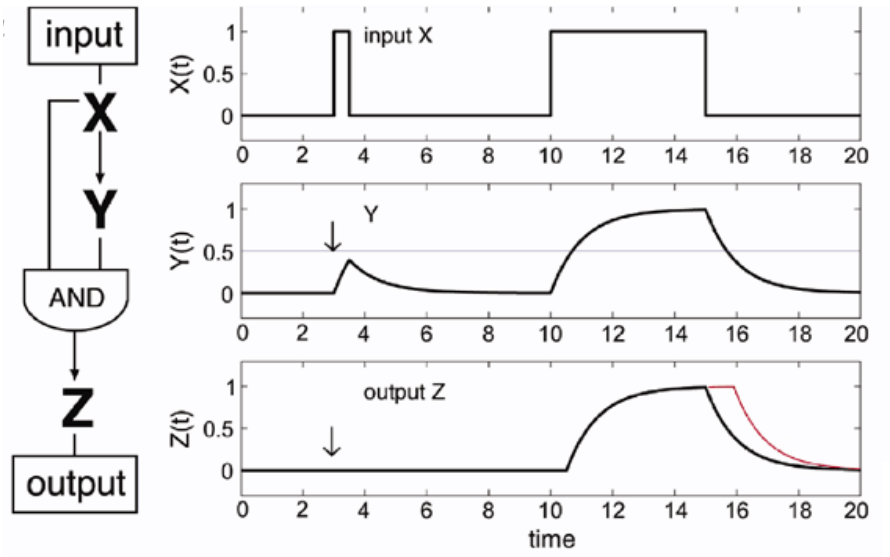


**Figure 2.** The design principle of the coherent FFL describes how the rate of a protein, Z, is a function of the activity of two transcription factors, X and Y, which are activated by two independent signal inputs ($S_x$ and $S_y$). Detection of signal persistence arises as a result of the time difference between direct and indirect activation routes (X to Y, and X to Z via Y), and the requirement for persistent stimuli from both X and $S_y$ before Y is activated (compare graphs for activation of X, Y, and Z). Hence, short pulses do not lead to activation of Z, but persistent activation of X and Y will. Reprinted by permission from Macmillan Publishers Ltd: Shen-Orr et al. Network motifs in the transcriptional regulation network of *Escherichia coli*, *Nature Genetics*, 31, 64-68, Copyright (2002).

Systems biologists more generally use the term *design principles* to highlight that generic network structures instantiate general dynamic patterns that i) are independent of specific realizations in different causal systems, and ii) serve functional or goal-oriented roles in engineering and biology (Green 2015). In other words, the characteristics of functional equivalence classes are explained with reference to *how network structures constrain dynamic outputs* to enable generic types of functions such as sustained oscillations, noise filtering, robust perfect adaptation, signal amplification, bi-stable switching, etc. (Doyle and Stelling 2006; Tyson et al. 2003; Tyson and Novák 2010).

The quest for design principles highlights the hope in systems biology that any network circuit with a specific structure, regardless of the specific details of its causal constituents, will belong to a more general functional equivalence class. If so, this would allow gene regulatory functions to be predicted and explained independently of detailed knowledge about the lower-level genetic and molecular details of specific systems. Generic functions of network motifs have been demonstrated in various contexts (Alon 2007). In the neighboring field of synthetic biology, multiple realizability through network motifs is exploited as a design heuristic for the synthesis of synthetic circuits with pre-defined functions (Koskinen 2017; 2019). Similarly, systems biologists

have recently explored the global properties of gene regulatory networks, following up on Kauffman's (1969; 1993) insight that even complex networks often converge to a limited set of stable states. Using the framework of dynamical systems theory, systems biologists have demonstrated that many different molecular mechanisms can lead to the same attractor states (representing biological functions or states of cell differentiation), thus moving the focus from the details of causal pathways to system trajectories (Huang 2011). This approach can potentially explain biological robustness (via top-down causation), because it can show how stable functional states are largely independent of the specific states (or initial conditions) of specific network nodes.

Functional equivalence classes in biology are, however, more contested than the classical examples of universality in physics (Auletta, Ellis and Jaeger 2008). Analyses of global network topologies and network motifs have shown that functions of genetic circuits are dependent on the contexts of the gene regulatory network, the environment, and organisms as a whole (Huang 2011; DiFrisco and Jaeger 2019). It has therefore been debated to what extent the structure of network motifs determines gene regulatory functions (Isalan et al. 2008; Jiménez et al. 2017). Similarly, systems biologists have debated whether biological systems exhibiting robust perfect adaptation necessarily realize integral feedback control (cf., Yi et al. 2000; Briat et al. 2014). Importantly, however, conditional independence is compatible with some degree of context-dependence within defined boundaries. We further clarify this below.

The discovered complexity has sparked an interest in understanding how wider system contexts can influence the characteristic functioning of specific network motifs. This can for instance be done through simulations where parameter spaces for the strengths of inputs and weighting of regulatory connections are varied (Tyson and Novák 2010). Hence, an aim here is to explore the conditions under which generic functions can be inferred from structural network types. Systems biologists have also used computer simulations to explore the possibility spaces for network structures that can realize specific functions of interest. For instance, they explore how many network topologies fall within a functional equivalence class and which structural features characterize the class. As an example, Ma et al. (2009) conducted a computational search for networks capable of performing robust perfect adaptation and investigated their regulatory wiring patterns. From a starting point of 16,038 possible network topologies, they found that only 395 were capable of performing RPA, and that they all fell into two generic structural classes (one is a negative feedback loop with a buffer node, the other is an incoherent feedforward loop). Interestingly, all known biological examples of RPA are instantiations of the negative feedback

control type, as described by Yi et al. (2000). The example thus highlights how structural constraints may realize functions that are multiply realized in distinct systems and thus "unify the organization of diverse circuits across all organisms" (Ma et al. 2009, 760).[7]

The complexity and diversity of biological systems presents a major challenge to provide an analysis in this context similar to the RG explanation in physics. But despite the limitations for universal laws in biology, generic models have proven useful for explaining why characteristic dynamic patterns arise in causally different systems. An ideal in systems biology is to shift the focus from inherent properties of specific genes or proteins to how those are interconnected through stabilizing regulatory structures that give rise to similar higher-level behaviors. In physics and chemistry as well as biology, an important part of scientific analysis is thus to determine "how many values of hidden variables can underlie the same higher-level description" (Ellis 2016, p. 120). In the following, we examine further how the causal slack of "hidden variables" supports a relative explanatory autonomy of higher-level models.

### 2.3. Causal slack and explanatory autonomy

The notion of hidden variables can be understood as a domain of lower-level causal details that would not change the output of a higher-level function, e.g., because the system trajectory would converge to the same fixed point or attractor in an abstract phase space.[8] As Ellis highlights, for a given equivalence class "it does not matter which particular lower level state occurs, as long as the corresponding higher-level variables are in the desired range" (Ellis 2008, p. 74). This has important implications for the way natural phenomena are represented and explained.

A notable feature of systems biology textbooks, compared to those of molecular biology, is that molecular details are almost absent in the figures and diagrams (cf., Alon 2007; Lodish et al. 2008). The use of highly abstract illustrations not only highlights how the functional descriptions are (relatively) independent of molecular details, but also that functional equivalence classes are identified through procedures of what Ellis (2012) calls *information hiding*. Akin to how we arrive at explanations for universal behaviors in physics by the use of RG explanations, systems biologists

[7] This necessarily involves abstraction from lower-level details. In the words of Ma et al.: "Here, instead of focusing on one specific signaling system that shows adaptation, we ask a more general question: what are all network topologies that are capable of robust adaptation?" They further state that the aim to "construct a unified function-topology mapping […] may otherwise be obscured by the details of any specific pathway and organism". (Ma et al. 2009)

[8] It goes (almost) without saying that this notion of "hidden variables" is not quantum mechanical.

must necessarily abstract from molecular details in order to make the generic patterns of network organization visible (Levy and Bechtel 2013; Green 2015).

Some would argue that information hiding strips the abstract network models of their explanatory power, and that these are explanatory only in so far as details are added back to explain the workings of specific systems. For instance, Matthiessen (2016) argues that design principles such as IFC are not explanatory if they do not allow us to distinguish between different species of bacteria or between organisms and thermostats. But while fine-graining by adding details would serve the explanatory purpose of how specific systems work (what we can call a type-I question), it would not address the (type-II) question of what these systems have in common or *why the same abstract model or principle applies to causally diverse systems* (Batterman 2002, p. 23). The reductionist would hence have to explain how higher-level descriptions can be relatively autonomous from changes of lower-scale details (see also Wimsatt 2007).

To respond to the type-II question, it is important to highlight two implications of what Ellis calls causal slack, and others understand through *scale separation* of multi-level systems. Scale separation refers to the manifestation of different dominant behaviors at different length scales, which accordingly must be described through different types of mathematical models (Batterman 2012). If, for instance, one is interested in modeling the bending properties of a material such as steel, it would not be useful for take the lowest (atomic) scale as the starting point. Instead, one typically starts with variables at the mesoscale and upscales to higher-level variables (such as elastic material parameters) that are used in continuum models. In fact, a detailed microscale models would often be irrelevant if the aim is to describe upper-scale continuum behaviors, as upper-level behaviors are literally invisible at the lowest scales. Scale separation explains why meso- and macroscale models often work well, despite ignoring or even misrepresenting many lower-level details. Indeed, mesoscale parameters are often dependent on some microscopic details, and these are accounted for through so-called homogenization strategies (see also Section 3). One can interpret the use of such strategies as a way to determine the degree and kind of causal slack between different spatial scales.

Top-down causation implies that higher-level features are not just relatively autonomous from lower-level description but also influence the latter through constraining relations. An objection to this view may be that since higher-level features primarily select among possible lower-level states, higher levels are not really autonomous after all. If lower levels define the possibilities, and if emergent features are always realized through materials at lower levels, what

does autonomy really consist in?[9] Our response is as follows. First, causal slack implies that higher levels are not fully controlled or determined by lower-level features but also by the structuring of the system. Second, the requirement of higher-level boundary conditions to constrain the space of causal possibilities implies that some lower-level states cannot be accounted for without the boundary conditions imposed by higher-level structures. For instance, Noble (2012) highlights how a phenomenon such as the heart rhythm is not possible without top-down causation, because the constraining relations of a membrane (understood mathematically as cell voltage) are required to produce oscillations of ionic currents.

Noble's example illustrates how constraints, as indicated in the introduction, must be understood in an *enabling* as well as *limiting* sense, as they allow for stable system behaviors or robust functions that would be impossible to reach in an unconstrained system (Pattee 1971; 1973). Another example illustrating this point is how our rigid skeleton enables upright movement on land by delimiting the possible directions of muscle contraction (Hooker 2013). It would be misleading to say, in this context, that the top-down constraints of the bone primarily provide a (non-explanatory) background for the lower-level states or operations of molecular muscle cells. Rather, as in the case of cardiac cells, certain emergent properties become possible only when constraining relations are applied on lower-level states. The causal power of top-down constraints can also be clarified through "negative" examples, where the constraints are removed and functions as a result become impossible to obtain. Breaking a bone has immediate effects on the causal possibilities for muscle performance, just like a ripped sail immediately changes the speed and control of a sailboat. Similarly, Noble mathematically demonstrates how removal of the top-down constraints of cell voltage causes oscillations to cease in a simulation of the heart rhythm (Noble 2012). Generally, many biological processes would not be possible without inter-level constraining relations (see DiFrisco and Jaeger 2020 for further examples).

[9] This concern was for instance raised after a talk by Ellis entitled "On the Nature of Causality in Complex Systems", at the conference The Causal Universe, Krakow, Poland, May 17-18, 2012. Available online: https://www.youtube.com/watch?v=nEhTkF3eG8Q. In the following we further elaborate on a possible response to this question.

## 3. Top-down causation and high-level variables

For Ellis, a crucial feature of top-down causation is how coherent higher-level actions emerge from top-down constraints on lower-level dynamics (Ellis 2012, p. 128). Such constraints are often mathematically described as boundary conditions that delimit the set of lower-level variables (e.g., the set of initial conditions as inputs to the dynamics) within which a given function can be realized. Top-down causation should thus not be understood as causal effects that are completely autonomous from lower-level dynamics. The organizing or design principles defined at higher scales nevertheless have causal effects that constrain the distributions or values of lower-level constituents or states, thus enabling emergent behaviors of lower levels (cf. Emmeche et al. 2000). As mentioned in the introduction, Ellis (2008; 2012) distinguishes between several types of downward causation. Following up on sections 2.1. and 2.2, we focus only on two, namely top-down causation by information control and mechanical or algorithmic top-down causation.

Feedback control loops in biology are instances of what Ellis calls *top-down causation by information control* (Auletta, Ellis and Jaeger 2008; Ellis 2012). The IFC-principles and the proposed generic functions of network motifs highlight a basic condition for equivalence classes, namely that many different input situations or causal "realizers" would give rise to equivalent operational outcomes, as long as basic structural requirements are obeyed. Importantly, a feedback loop is here interpreted as a structural constraint that delimits the space of possibilities for lower-level dynamics. Hence, feedback control consists in *information selection*. For instance, the structure of a feedforward loop motif determines whether a genetic circuit should respond only to persistent input signals (coherent feedforward loop) or immediately to any nutrient detected (incoherent feedforward loop). As illustrated in Figure 2, coherent feedforward loops control transcription by introducing a time difference between direct and indirect transcription activation routes. Similarly, incoherent feedforward loops can minimize the response to disturbance by simultaneously activating the transcription of an output product and an intermediary transcription factor that inhibits output protein production. Hence, structures such as network motifs constrain the dynamic possibilities of gene regulation at a lower scale in a hierarchy (see also Bechtel 2017). Ellis defines top-down causation by information control as follows:

> Top-down causation by information control occurs thanks to the connection between equivalence classes and information control […] In this case, the feedback control circuits

> produce reliable responses to higher level information (Ellis 2006; 2008), allowing equivalences classes of lower level operations that give the same higher level response for a certain goal. (Auletta, Ellis and Jaeger, pp. 1169-1170).

Different lower-level operations are here considered as *controlled by information from above* in the sense that the control circuits are considered as higher-level entities in two senses. First, the functions are implemented by networks that cannot be reduced to the operation of lower-level entities in isolation (Ellis 2008). Second, functional goals are higher-level concepts referring to the properties of a whole system (an organism, metabolic system, or circuit of interacting processes or entities). The second feature requires some clarification. Ellis (2008) argues that: "[t]he goals in biological systems are "intrinsic higher-level properties of the system considered, and determine the outcome (unlike the usual physical case, where the initial state plus boundary conditions determine the outcome). […] The initial state of the system is then irrelevant to its final outcome, provided the system parameters are not exceeded" (Ellis 2008, p. 74).

Ellis stresses that top-down relations in biology include considerations of part-whole relations, which do not necessarily translate to the physical context. While we agree with this characterization, we do not view downward causation primarily as a compositional relation between parts and wholes (see also Woodward, forthcoming). Rather, we view top-down causation as relations between higher-level and lower-level variables. Thus understood, we find it potentially misleading to consider the relative independence from initial conditions as a prime feature that separates physical and biological systems. As we have seen, universality in physics is characterized through insensitivity to lower scale details, and both biological and physical systems can be described through equivalence classes within certain boundaries of system parameters. Accordingly, we suggest, that "bottom-up reductionism" can also be challenged by examples of mechanical or algorithmic top-down causation in both physics and biology. Although typically considered a "weaker form" of top-down causation, compared to top-down causation by information control, cases of multi-scale modeling in both domains highlight the limitations of a bottom-up approach (Batterman and Green 2020). Mechanical or algorithmic top-down causation refers to a ubiquitous form of top-down causation that occurs whenever "high-level variables have causal power over lower level dynamics through system structuring or boundary conditions, so that the outcome depends on these higher level variables" (Ellis 2012, p. 128). With Ellis, we believe that this form of top-down causation is much more common than typically recognized.

*3.1. Revisiting mechanical top-down causation*

As mentioned, Ellis views mechanical top-down causation as a phenomenon occurring also in physical and chemical systems. This was exemplified in the way changes in higher-level variables (such as pressure or temperature) can lead to changes in lower level interactions in gases and fluids, enabling new properties such as gas ignition or phase transitions (Section 2.1).[10] In these examples, Ellis seems to assume the correspondence between lower- and higher-level variables is given by a relatively simple relation between these, i.e., that higher-level variables can be derived from coarse-graining of lower-level ones (Auletta, Ellis and Jaeger 2008). Ellis broadly defines high level variables as follows:

> 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 molecular positions and velocities). (Ellis 2008, p. 70).

Ellis further distinguishes between *coarse-grained* higher-level variables and *irreducible* higher-level variables. The pressure of an ideal gas exemplifies the former, whereas feedback loops or the tertiary structure of protein folding illustrate the latter (see also Brigandt and Love 2017). Coarse-grained variables can be obtained by averaging over a set of lower level variables, and they are therefore in principle possible to derive from lower-level details (although this is often not done for practical reasons). This has implications for the strength of mechanical top-down causation because Ellis views all the high-level variables concerned with this type of top-down causation as coarse-grained. He further writes: "The resulting high level relations are then an inevitable consequence of the low level interaction, given both the high level context and the low level dynamics (based in physics)" (Ellis 2008, p. 72).

As a friendly amendment to this view, we suggest that the scope of mechanical or algorithmic top-down causation be expanded to include cases that go beyond instances of simply coarse-grained higher-level variables. While we agree that the high level variables are coarse-

[10] Other examples from physics are discussed in (Bishop 2012; Christiansen 2000; see also Ellis 2018).

grained in the case of homogeneous systems such as ideal gases, there are many multi-scale systems in physics where meso- and macroscale parameters cannot be obtained via simple averaging procedures (Batterman and Green, 2020). This presents a further challenge to the reductionist point of view.

Indeed, in the case of an ideal gas, we can assume that the system is homogenous, and we can therefore upscale to the higher-level thermodynamic behavior by relatively simple averaging over micro-scale details (such as molecular spatial and velocity distributions). However, whenever multi-scale systems are heterogenous and display more complicated limit behaviors, this approach would fail. For instance, if the task is to develop a multi-scale model of the bending behaviors and relative strength of heterogenous materials such as a steel beam or vertebrate bone, complex homogenizations strategies are typically adopted to account for mesoscale structures that are not observable at lower scales (Batterman and Green 2020). In both contexts, simple averaging over lower-scale variables would not enable scientists to predict macroscale material properties.

For multiscale systems that are heterogeneous (e.g., composites of materials with different conductivities or elastic behaviors), the aim of upscaling is to find effective (continuum scale) parameters (like Young's modulus) that code for microstructural details of the composites.[11] Typically this involves the examination of a representative volume element (RVE) that reflects the nature of heterogeneities at scales (mesoscales) where those structures are deemed to be important. One introduces correlation functions to characterize (primarily) the geometric and topological aspects of the mixture in the RVE (Torquato 2002). The mathematics that enables one to find the effective parameters that characterize the behavior of the composite at the continuum scale is called "homogenization theory." (Batterman and Green 2020 discusses some aspects of this in the context of materials science and biology. See also Batterman, forthcoming.) Note that the relative autonomy of the homogenized system (at the continuum scale) from the atomic lower scale details reflects a kind of emergence. This sense of emergence is weaker than that associated with higher-level variables characterizing the human mind or social phenomena. Yet, the higher-level parameters cannot be reduced to or derived from lower-level details.

Consider Young's modulus, an example of a higher-level parameter of central importance in materials science and biophysics. Young's modulus parameterizes the stiffness of a material and is identified as the slope (or coefficient of proportionality) of a stress-strain curve of a given material. Stiffness is understood as the resistance of a material to deformation in response to applied force

[11] Note that "microstructure" here refers to structures far above the atomic and far below the continuum.

and cannot be understood or defined at atomic scales. More generally, material parameters describe mechanical properties of a larger continuum of structure that cannot be measured or defined at the level of individual "parts". As stressed by developmental biologist (and biophysicist) Lance Davidson and colleagues, material parameters of relevance for modeling of the development of an embryo are inherently higher-level concepts:

> The capacity of the notochord to resist bending as it extends the embryo comes from the structure of the whole notochord. Measurements at the level of the individual collagen fiber or fluid-filled cells that make up the structure would not reveal the mechanical properties of the whole notochord. (Davidson et al. 2009, p. 2157)

We can interpret this as a form of (mechanical) top-down causation because biomechanical features influence the development of vertebrate embryos through constraints on motility and bending of cells (Green and Batterman 2017). As in the case of feedback loops, the higher-level material parameters cannot be achieved through coarse-graining but requires measurements where the whole structure must be intact. Similarly, the action potential in neurons is a mesoscale parameter that cannot be measured or understood at the molecular or sub-cellular level because the property depends on the whole cell structure (Noble 2012). We should therefore not think of higher-level variables in this context as merely "smeared-out versions" of a more fundamental lower-level description. At the same time, the explanatory autonomy should not be overemphasized as there is clearly some connection between microstructure and material parameter values. In both physics and biology, modeling of materials over large spatial scales requires that scales are bridged, e.g., via the identification of RVEs. In biology, the relations between higher and lower-scale variables are often illustrated in diagrams through feedback relations going both up and down (see e.g., Noble 2012; Lesne 2013). Thus, the idea of a scale dependent *relative autonomy* offers an alternative to more extreme positions (reductionism or anti-reductionism), while capturing aspects of how scientists deal with multi-scale systems.

In summary, examples of multiscale systems that are heterogenous (unlike ideal gases) support a stronger interpretation of the causal role of higher-level variables in the context of mechanical top-down causation. Higher-level variables in such contexts are not coarse-grained in the weaker sense that they can be *derived* from lower-level details via simple averaging. Rather, the requirement for homogenization strategies in multiscale modeling highlights how higher-level variables are *relatively autonomous* from lower-level descriptions. The examples thus challenge an

assumption in the definition of mechanical top-down causation (Ellis 2008; 2012), but in doing so they also strengthen Ellis' point about the importance and ubiquity of top-down causation and emergence in science.

## 4. The practical importance of top-down causation

Debates on the possibility or strength of top-down causation are often assumed to primarily be of theoretical interest to philosophers. However, as emphasized by Ellis and others, top-down causation and equivalence classes have important practical implications (Auletta, Ellis and Jaeger 2008; Ellis, Noble and O'Connor 2012; see also Wimsatt 1976; 2007). Top-down causation, for Ellis, is not a mysterious metaphysical concept, but an empirical phenomenon that can be demonstrated through experimental intervention. He defines the following operational criteria for top-down causation:

> To characterize some specific causal effect as a top-down effect, we must demonstrate that a change of higher level conditions alters the sequence of processes at lower levels; we do this by changing higher level conditions and seeing what happens at the lower levels (Ellis 2012).

Thus, top-down causation is here given a concrete interpretation as a relation between system variables at different spatial scales or levels (see also Ellis 2016, p. 16, and Woodward, forthcoming). If intervening on macroscale variables can change the dynamic states of microscale processes, it has important practical implications for the design of experiments, for multi-scale modeling, as well as for discussions about where and how to intervene to control and change future outcomes (such as disease definitions and treatment modalities in medicine, for instance). In the following, we present further examples of this view.

### *4.1. Top-down causation and multi-scale modelling*

Scientific explanations often highlight molecular activities (biology) or the role of laws and initial conditions (physics), but boundary conditions are often equally important for understanding system behaviors. In the context of multi-scale cardiac modeling, Noble (2012) argues forcefully that that the models describing the processes at the lowest scale, i.e., ordinary differential equations

describing ionic currents, cannot be solved without the boundary conditions determined by the cell voltage. It is important to note here that cell voltage is a parameter that cannot be defined at the molecular or subcellular scale. Similarly, models of the action potential at the level of cells depend on inputs from models at the tissue scale (defined via partial differential equations), which describe how biophysical features of the tissue can influence the propagation of electrical currents through the 3D structure of the heart (Qu et al. 2011; Green 2018).

Although parameters such as the cell voltage, or the geometrical and electrical properties of different tissue types are "nothing but" properties of physical structures, they are not reducible to or derivable from lower-scale variables. In fact, one cannot measure or even conceptualize these variables at lower scales. They also cannot be reduced to explanatory background conditions for descriptions of causal efficacy at lower scales, because the boundary conditions and higher-level parameters in general are required to channel the lower-level behaviors in the first place. As highlighted by Noble: "without the downward causation from the cell potential, there is no [heart] rhythm" (Noble 2012, p. 58). A reductionist perspective thus faces great difficulties in terms of showing that a bottom-up analysis is itself sufficient.

In the context of developmental biology and cancer research, the importance of top-down causation is also increasingly acknowledged. New experimental techniques to manipulate higher-level biomechanical cues have revealed that macroscale biomechanical properties (e.g., tissue stiffness) can influence gene expression, molecular signaling pathways, as well as cell differentiation (Miller and Davidson 2013). This has important implications for understanding how biomechanical constraints can buffer genetic "noise", and how there is sufficient causal slack between macroscale biomechanical models and molecular details to allow modelers to efface many lower-scale details. Moreover, it has (negative) implications for the view that genetic or molecular causation has a privileged role in developmental biology (this is further discussed in Green and Batterman 2017). The biases of such perspectives can also have important social implications, as we now clarify.

### *4.2. Research and treatment modalities in medicine*

Ellis (2012) highlights the existence of different treatment modalities in medicine, depending on whether one commits to a reductionist (bottom-up) perspective or to a more holistic view. This is particularly apparent in the case of mental disorders, where the focus can span from genetic susceptibilities and molecular dysfunctions to how states of the mind can impact physical health

(see also Ellis 2016). Philosophical assumptions concerning the possibility and relative importance of top-down causation (vs. bottom-up causation) does, in fact, influence which research questions and treatment regime are seen as most promising. Attention to top-down causation can fruitfully point to the missed opportunities of the prioritized focus on reductionist research programs – and positively highlight the potential benefits of a broader perspective (Fuchs 2018).

In closing, we provide a few examples to illustrate this view. Cancer is often understood as a disease caused by accumulation of somatic mutations. However, increasing evidence suggests that tissue-scale properties can sometimes overrule genetic instructions (Soto et al. 2008; Bissel and Hines 2011). The constraints of the tissue can either promote or reduce cell proliferation and motility, depending on the biomechanical properties of the tumor microenvironment, which is another important example of top-down causation (Green, forthcoming). Hence, the reductionist perspective may create unfortunate blind spots, such as the opportunity to develop treatment strategies that target tissue-scale properties (Stylianopoulos 2017).

Similarly, the criticism of reductionism is highly relevant in the context of preventive medicine. With the promotion of precision medicine, the research focus on genetic factors that increase an individual's susceptibility for developing complex diseases like cancer, depression, or dementia has intensified. But a focus on genetic factors is neither sufficient for understanding and treating such complex diseases, nor is it necessarily more precise. An important concern is that genetic risk profiling at the individual level shifts attention away from structural causes at the population level, such as socio-economic disparities, that may be more efficient to intervene on (Hoeyer 2019; Olstad and McIntyre 2019). Top-down causation is therefore not only of theoretical philosophical interest, but it is an empirical phenomenon with profound scientific and social implications.

## 5. **Concluding remarks**

Explanatory autonomy of levels or scales is often defended with reference to the existence of universality or functional equivalence classes. Universal or functionally equivalent behaviors are described through macroscale models and parameters that cannot be reduced to lower-level models. Examples examined in this chapter include thermodynamics near critical points as well as feedback control in biology. The examples illustrate how models can be explanatory without specifying how a behavior is causally realized in any specific system. In fact, generic models are explanatory *because* they show how many causal details are explanatorily irrelevant as long as stabilizing structures defining an equivalence class are in place. The notion of information hiding highlights

that there is enough "causal slack" at the bottom to make the inclusion of all possible lower-level details irrelevant or even counterproductive for the purpose of explaining higher level or multiscale systems.

A key issue in discussions of reductionism and top-down causation is whether higher-level variables can be derived from lower-level details. Mechanical top-down causation is often seen as a weak form of top-down causation, because the high-level variables are taken to be nothing but coarse-grained correlates of lower-level variables. Thus understood, it is unclear whether reductionism is overcome. We have argued that upscaling of variables via coarse-graining only works for simple homogenous systems, such as an ideal gas. For systems with complex microstructures at the mesoscale (such as steel or bone), more involved upscaling techniques are required. The reason is that physical systems at different scales display distinct physical structures and behaviors, and that higher-level behaviors are dependent on some microstructural details that are best studied at the mesoscale.

Mesoscale parameters (such as material parameters) differ from what Ellis calls coarse-grained high-level variables in being identified via homogenization strategies. The need for such strategies signals a stronger explanatory autonomy of high-level variables also in physical examples than often assumed. Hence, attention to scale-dependency of characteristic behaviors in multi-scale systems offers support to Ellis' account by further extending the scope and significance of mechanical top-down causation. Top-down causation is not a suspicious, rare form of causation, but is ubiquitous in physical and biological systems alike. This has important practical implications not only for scientific modeling and explanation, but also for how we best approach complex socio-scientific problems.

**References**

Allen, Timothy FH, and Thomas B. Starr. 1982. *Hierarchy: perspectives for ecological complexity*. Chicago: University of Chicago Press.

Alon, Uri, Michael G. Surette, Naama Barkai, and Stanislas Leibler. 1999. "Robustness in bacterial chemotaxis." *Nature* 397, no. 6715: 168.

Alon, Uri. 2007. *An Introduction to Systems Biology. Design Principles of Biological Circuits*. Boca Raton: Chapman and Hall/CRC.

Auletta, Gennaro, George F.R. Ellis, and Luc Jaeger. 2008. "Top-down causation by information control: from a philosophical problem to a scientific research programme." *Journal of the Royal Society Interface* 5, no. 27: 1159-1172.

Batterman, Robert W. 2000. "Multiple realizability and universality." *The British Journal for the Philosophy of Science* 51, no. 1: 115-145.

Batterman, R. W. (2002). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence*. New York: Oxford University Press.

Batterman, Robert W. 2012. "The tyranny of scales." In *Oxford Handbook of Philosophy of Physics*, ed. Robert W. Batterman, 225–86. Oxford: Oxford University Press.

Batterman, Robert W. 2018. "Autonomy of theories: An explanatory problem." *Noûs* 52, no. 4: 858-873.

Batterman, Robert W. forthcoming. *A Middle Way: A Non-Fundamental Approach to Many-Body Physics*. Oxford: Oxford University Press.

Batterman, Robert W., and Sara Green. 2020. "Steel and bone. Mesoscale modeling and middle-out strategies in physics and biology." *Synthese*, published online: https://doi.org/10.1007/s11229-020-02769-y.

Bechtel, William. 2017. Top-Down Causation in Biology and Neuroscience. Control Hierarchies. In *Philosophical and scientific perspectives on downward causation* pp. 203-224, eds. Paoletti, Michele Paolini, and Francesco Orilia, New York, NY: Taylor & Francis.

Bedau, Mark A. 1997. "Weak emergence." *Noûs* 31: 375-399.

Bernard, Claude. 1927/1957. *An Introduction to the Study of Experimental Medicine*. New York: Dover Publications.

Bishop, Robert C. 2012. "Fluid convection, constraint and causation." *Interface Focus* 2, no. 1: 4-12.

Bissell, Mina J., and William C. Hines. 2011. "Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression." *Nature medicine* 17, no. 3: 320.

Briat, Corentin, Ankit Gupta, and Mustafa Khammash. 2014. "Integral feedback generically achieves perfect adaptation in stochastic biochemical networks.", *arXiv*:1410.6064v3.

Brigandt, Ingo, and Alan Love. 2017. "Reductionism in biology." In the *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University. https://plato.stanford.edu /entries/reduction-biology/.

Campbell, Donald T. 1974. "Downward causation in hierarchically organised biological systems." In *Studies in the Philosophy of Biology: Reduction and Related Problems*, ed. Francisco Ayala and Theodosius Dobzhansky, 179–86. London: Macmillan.

Cannon, Walter B. 1929. "Organization for physiological homeostasis." *Physiological reviews* 9, no. 3: 399-431.

Christiansen, Peder V. 2000. “Macro and micro-levels in physics.” In *Downward Causation: Minds, Bodies and Matter*, eds. Peter Bøgh Andersen, Claus Emmeche, Niels Ole Finnemann, and Peder Voetmann Christiansen, 51–62. Aarhus: Aarhus University Press.

Crick, Francis, and J. Clark. 1994. "The astonishing hypothesis." *Journal of Consciousness Studies* 1, no. 1: 10-16.

Davidson, Lance, Michelangelo von Dassow, and Jian Zhou. 2009. "Multi-scale mechanics from molecules to morphogenesis." *The International Journal of Biochemistry & Cell Biology* 41, no. 11: 2147-2162.

DiFrisco, J., & Jaeger, J. 2019. “Beyond networks: mechanism and process in evo-devo. *Biology & Philosophy*, 34(6), 1-24.

DiFrisco, J., & Jaeger, J. 2020. “Genetic causation in complex regulatory systems: an integrative dynamic perspective.” *BioEssays*, 42(6), 1900226.

Doyle, Francis J., and Joerg Stelling. 2006. "Systems interface biology." *Journal of the Royal Society Interface* 3, no. 10: 603-616.

Emmeche, Claus, Simo Køppe, and Frederik Stjernfelt. 2000. “Levels, emergence, and three versions of downward causation.” In *Downward Causation: Minds, Bodies and Matter*, eds. Peter Bøgh Andersen, Claus Emmeche, Niels Ole Finnemann, and Peder Voetmann Christiansen, 13–34. Aarhus: Aarhus University Press.

Ellis, George FR. 2005. "Physics, complexity and causality." *Nature* 435, no. 7043: 743.

Ellis, George FR. 2006. "On the nature of emergent reality.” In *The re-emergence of emergence,* Eds. Philip Clayton and Paul Davies, Oxford: Oxford University Press, 79-107.

Ellis, George FR. 2008. "On the nature of causation in complex systems." *Transactions of the Royal Society of South Africa* 63, no. 1: 69-84.

Ellis, George FR. 2012. "Top-down causation and emergence: some comments on mechanisms." *Interface focus* 2, no. 1: 126-140.

Ellis, George FR. 2016. *How can Physics Underlie the Mind? Top-Down Causation in the Human Context*. Berlin: Springer.

Ellis, George FR. 2018. "Top-down causation and quantum physics." *Proceedings of the National Academy of Sciences* 115, no. 46: 11661-11663.

Ellis, George FR, Denis Noble, and Timothy O'Connor. 2012. "Top-down causation: an integrating theme within and across the sciences?" *Interface focus*: 1-3.

Ellis, George FR, and Jonathan Kopel. 2017. "On the difference between physics and biology: Logical branching and biomolecules." https://arxiv.org/abs/1709.00950

Fang, Wei. 2020. "Multiple realization in systems biology." *Philosophy of Science* 87: 663-684.

Fodor, Jerry A. 1974. "Special sciences (or: The disunity of science as a working hypothesis)." *Synthese* 28, no. 2: 97-115.

Fuchs, Thomas. 2018. *Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind*. Oxford, UK: Oxford University Press.

Green, Sara. 2015. "Revisiting generality in the life sciences: Systems biology and the quest for general principles." *Biology and Philosophy* 30: 629-652.

Green, Sara. 2018. "Scale dependency and downward causation in biology." *Philosophy of Science* 85, no. 5: 998-1011.

Green, Sara. Forthcoming "Cancer beyond genetics: On the practical implications of downward causation." In *Biological Levels: Composition, Scale and Evolution in Complex Systems*, Eds. Daniel S. Brooks, James DiFrisco & William C. Wimsatt, MIT Press.

Green, Sara, and Robert Batterman. 2017. "Biology meets physics: Reductionism and multi-scale modeling of morphogenesis." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 61: 20-34.

Green, Sara, and Nicholaos Jones. 2016. "Constraint-based reasoning for search and explanation: Strategies for understanding variation and patterns in biology." *dialectica* 70, no. 3: 343-374.

Green, Sara, and Olaf Wolkenhauer. 2013. "Tracing organizing principles: Learning from the history of systems biology." *History and Philosophy of the Life Sciences* 35: 553-576.

Gross, Fridolin, and Sara Green. 2017. "The sum of the parts: large-scale modeling in systems biology." *Philosophy & Theory in Biology*, 9 no. 10.

Guggenheim, Edward A. 1945. "The principle of corresponding states." *The Journal of Chemical Physics* 13, no. 7: 253-261.

Hooker, Cliff. 2013. "On the import of constraints in complex dynamical systems." *Foundations of Science* 18, no. 4: 757-780.

Hoeyer, Klaus. 2019. "Data as promise: Reconfiguring Danish public health through personalized medicine." *Social Studies of Science* 49, no. 4: 531-555.

Huang, Sui. 2011. "Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways." *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, no. 1575: 2247-2259.

Isalan, Mark, Caroline Lemerle, Konstantinos Michalodimitrakis, Carsten Horn, Pedro Beltrao, Emanuele Raineri, Mireia Garriga-Canut, and Luis Serrano. 2008. "Evolvability and hierarchy in rewired bacterial gene networks." *Nature* 452, no. 7189: 840.

Jiménez, Alba, James Cotterell, Andreea Munteanu, and James Sharpe. 2017. "A spectrum of modularity in multi-functional gene circuits." *Molecular Systems Biology* 13, no. 4.

Juarrero, Alicia. 1998. Causality as constraint. In *Evolutionary Systems: Biological and Epistemological Perspectives on Selection and Self-Organization*, eds. G. Van de Vijver, S. Salthe and M. Delpos (pp. 233-242), Dordrecht: Springer.

Juarrero, Alicia. 1999. *Dynamics in Action: Intentional Behavior as a Complex System*. Cambridge, MA: MIT Press.

Kadanoff, Leo P. 1971. "Critical behavior, universality and scaling". In *Proceedings of the International School of Physics "Enrico Fermi" Course LI,* ed. M. S. Green, 100—117, New York: Academic Press.

Kauffman, Stuart A. 1969. "Metabolic stability and epigenesis in randomly constructed genetic nets." *Journal of Theoretical Biology* 22, no. 3: 437-467.

Kauffman, Stuart A. 1993. "The Origins of Order: Self-organization and Selection in Evolution. Oxford: Oxford University Press.

Kim, Jaegwon. 1998. "Mind in a Physical World." Cambridge, MA: MIT Press.

Kim, Jaegwon. 2000. "Making sense of downward causation." In *Downward Causation: Minds, Bodies and Matter*, eds. Peter Bøgh Andersen, Claus Emmeche, Niels Ole Finnemann, and Peder Voetmann Christiansen, 305–21. Aarhus: Aarhus University Press.

Koskinen, Rami. 2017. "Synthetic biology and the search for alternative genetic systems: Taking how-possibly models seriously." *European Journal for Philosophy of Science* 7, no. 3: 493-506.

Koskinen, Rami. 2019. "Multiple realizability as a design heuristic in biological engineering." *European Journal for Philosophy of Science* 9, no. 1: 15.

Lesne, Annick. 2013. "Multiscale analysis of biological systems." *Acta biotheoretica* 61, no. 1: 3-19.

Levy, Arnon, and William Bechtel. 2013. "Abstraction and the organization of mechanisms." *Philosophy of science* 80, no. 2: 241-261.

Lodish, Harvey, Arnold Berk, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, Anthony Bretscher, Hidde Ploegh, and Paul Matsudaira. 2008. "*Molecular Cell Biology*." Macmillan.

Ma, Wenzhe, Ala Trusina, Hana El-Samad, Wendell A. Lim, and Chao Tang. 2009. "Defining network topologies that can achieve biochemical adaptation." *Cell* 138, no. 4: 760-773.

Mangan, Shmoolik, and Uri Alon. 2003. "Structure and function of the feed-forward loop network motif." *Proceedings of the National Academy of Sciences* 100, no. 21: 11980-11985.

Matthiessen, Dana. 2017. "Mechanistic Explanation in Systems Biology: Cellular Networks." *The British Journal for the Philosophy of Scienc,* 68, no. 1: 1-25.

Miller, Callie Johnson, and Lance A. Davidson. 2013. "The interplay between cell signalling and mechanics in developmental processes." *Nature Reviews Genetics* 14, no. 10: 733.

Moreno, Alvaro and Matteo Mossio. 2015. *Biological Autonomy: A Philosophical and Theoretical Inquiry*. Dordrecht: Springer.

Mossio, Matteo and Alvaro Moreno. 2010. "Organisational closure in biological organisms." *History and Philosophy of the Life Sciences*, 32:269-288.

Noble, Denis. 2012. "A theory of biological relativity: no privileged level of causation." *Interface focus* 2, no. 1: 55-64.

Noble, Denis. 2017. "Systems biology beyond the genome." In *Philosophy of Systems Biology*, ed. Sara Green, 227-235. Cham: Springer.

Olstad, Dana L. and Lynn McIntyre. 2019. "Reconceptualising precision public health." *BMJ open* 9, no. 9: e030279.

Oppenheim, Paul, and Hilary Putnam. 1958. "Unity of science as a working hypothesis." In *Minnesota Studies in the Philosophy of Science*, ed. Herbert Feigl, Grover Maxwell, and Michael Scriven, 3–36. Minneapolis: University of Minnesota Press.

Pattee, Howard H. 1971. "Physical Theories of Biological Co-ordination." *Quarterly Reviews of Biophysics* 4: 255–76.

Pattee, Howard H. ed 1973. *Hierarchy Theory*. New York, NY: Braziller.

Paoletti, Michele Paolini, and Francesco Orilia, eds. 2017, *Philosophical and scientific perspectives on downward causation*. New York, NY: Taylor & Francis.

Putnam, Hillary. 1980. "Philosophy and our mental life". In *Readings in Philosophy of Psychology*, ed. Ned Block, volume 1, pp. 134-143. Cambridge, MA: Harvard University Press, Cambridge.

Qu, Zhilin, Alan Garfinkel, James N. Weiss, and Melissa Nivala. 2011. "Multi-scale modeling in biology: how to bridge the gaps between scales?." *Progress in Biophysics and Molecular Biology* 107, no. 1: 21-31.

Salthe, Stanley S. 1985. *Evolving hierarchical systems. Their structure and representation*. New York, NY: Columbia University Press.

Serban, Maria, and Sara Green. 2020. "Biological robustness: Design, organization, and mechanisms." In *Philosophical Perspectives on the Engineering Approach in Biology: Living Machines?,*141-164, eds. Sune H. Holm & Maria Serban. New York, NY: Taylor and Francis.

Sober, Elliott. 1999. "The multiple realizability argument against reductionism." *Philosophy of Science* 66, no. 4: 542-564.

Soto, Ana M., Carlos Sonnenschein, and Paul-Antoine Miquel. 2008. "On physicalism and downward causation in developmental and cancer biology." *Acta Biotheoretica* 56, no. 4: 257-274.

Stylianopoulos, Triantafyllos. 2017. "The solid mechanics of cancer and strategies for improved therapy." *Journal of Biomechanical Engineering* 139, no. 2: 021004.

Torquato, Salvatore. 2002. "Random Heterogenous Materials: Microstructure and Macroscopic Properties." New York: Springer.

Tyson, John J., Katherine C. Chen, and Bela Novak. 2003. "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell." *Current opinion in cell biology* 15, no. 2: 221-231.

Tyson, John J., and Béla Novák. 2010. "Functional motifs in biochemical reaction networks." *Annual review of physical chemistry,* no. 61: 219-240.

Ulanowicz, Robert E. 1986. *Growth and development. Ecosystems phenomenology*. New York, NY: Springer.

Ulanowicz, Robert E. 1997. *Ecology, the ascendent perspective*. New York, NY: Columbia University Press.

Wiener, Norbert. 1948. "Cybernetics or Control and Communication in the Animal and the Machine". MIT press.

Wimsatt, William C. 1976. "Reductionism, levels of organization, and the mind-body problem", In *Consciousness and the Brai*n, 199-267, ed. G.G. Globus, G. Maxwell, and I. Savodnik. New York: Plenum.

Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings*. Cambridge, MA: Harvard University Press.

Woodward, James. Forthcoming. "Downward causation and levels." In *Biological Levels: Composition, Scale and Evolution in Complex Systems*, eds. Daniel S. Brooks, James DiFrisco & William C. Wimsatt (Eds.), MIT Press.

Yi, Tau-Mu, Yun Huang, Melvin I. Simon, and John Doyle. 2000. "Robust perfect adaptation in bacterial chemotaxis through integral feedback control." *Proceedings of the National Academy of Sciences* 97, no. 9: 4649-4653.

# Summary of Comments on Biology meets physics: Reductionism and multi-scale modeling of morphogenesis

Subject: Highlight  Date: 5/24/2022 9:52:40 AM

The key underlying analytic concept is existence of functional equivalence classes of lower level structures and functions (Auletta et al., 2008; Ellis, 2016) corresponding to a specific emergent structure or function. Equivalence classes at a lower level collect elements whose differences are irrelevant for the emergent target feature at the higher level; it does not matter which one is used to realize the higher level feature. Existence of such functional equivalence classes is an indication of top-down causation (Auletta et al., 2008). An important example is the relation of developmental systems to the genome: a huge number of different genotypes (a genotype network) can result in the same phenotype (Wagner, 2017). Any one of these genotypes can be selected for through evolutionary processes in order to lead to a particular emergent function that promotes survival. As far as the higher level function is concerned, it is irrelevant which specific genotype is selected, so it is membership of the equivalence class at the lower level that is the key to what genotype gets selected when adaptation takes place. The huge size of these equivalence classes is what enables adaptive selection to find the needed biomolecules and interaction networks on geological timescales (Wagner, 2017).

\The characteristics of functional equivalence classes are explained with reference to how network structures constrain dynamic outputs to enable generic types of functions such as sustained oscillations, noise ltering, robust perfect adaptation, signal amplication, etc."

But macro conditions can then change parameters in the Lagrangian.

The key issue is whether downward causation is real, having real causal powers. I argue that it is; and that this kind of causation does not required any compromising of the underlying physics. It works by changing constraints

Microphysics enables this but does not determine the outcomes. The basic physics interactions of course enable all this to happen: they allow incredible complexity to emerge. Higher level organising principles such as Darwin's theory of evolution then come into play. That then changes the macro level context and hence the micro level context. This downward process [Campbell 1974] relies on concepts such as 'living' that simply cannot be represented at the microlevel, but determine outcomes.
5 The reductionist physics view is based in a linear view of causation. Central to the way biology works are the closely related ideas of self cause and circular causation.

Synchronic emergence is when the timescale t := tb − ta of the considered microdynamic outcomes is very short relative to the timescale T of change of structures at the micro scale: T ≫ t. It is the issue of emergent dynamics when parameters are constant
7

and constraints unchanging. In the case of the brain this would for example be the flow of electrons in axons leading to mental outcomes at that time, with this micro structure taken as unchanging. Electrons and ions flow in a given set of neural connections. Diachronic emergence is when the timescale of micro dynamic outcomes considered t is of the same order or larger than the timescale T of change of structure at the micro scale: $T \leq t$, so microdynamics contexts alters significantly during this time. It is the case when parameters or constraints change because of interactions that are taking place. In the case of the brain this would for example be when something new is learned so that strengths of neural connections are altered.

Life is an ongoing adaptive process involving metabolism, homeostasis, defence, and learning in the short term, reproduction, growth, and development in the medium term, and evolution in the long term. It uses energy, disposes of waste heat and products, and uses contextual information to attain its purposes.

Basic biological needs and functions In the case of animal life,5 the basic biological functions are,
B1: Metabolism (acquiring energy and matter, getting rid of waste),
B2: Homeostasis and defence,
B3: Reproduction and subsequent development,
B4: Mobility and the ability to act,
B5: Information acquisition and processing.
They serve as attractors when variation takes places ([Ginsburg and Jablonka 2019]:245). They are the higher level organising principles that evolution discovers and then embodies in hierarchically structured physiological systems, where the macro functions are supported at the micro level by metabolic networks, gene regulatory networks, and cell signalling networks, selected from an abstract space of possibilities and realised through specific proteins [Wagner 2014]. Information is central to what happens [Nurse 2008] [Davies 2019]. These principles cannot be described or identified at the underlying microphysical levels not just because the relevant variables are not available at that level, but because their multiple realisability at lower levels means they do not correspond to specific patterns of interactions at the ion and electron level. They correspond to a whole equivalence class of such patterns of interactions (Section 2.6).

[Noble 2012] [Ellis 2016]. The second one is that as well as efficient causation, Aristotle's other forms of causation play a key role in real world outcomes.

Higher level physiological needs reach down to the underlying physical level L2 via time dependent constraints [Ellis and Kopel 2019]. The set of interactions between elements at that level is uniquely characterised by the laws of physics L, but their specific outcomes are determined by the biological context in which they operate.

"The network in question contains about 28 neurons and serves to drive the muscles controlling the teeth of the gastric mill so that food can be ground up for digestion. The output of the network is rhythmic, and hence the muscular action and the grinders movements are correspondingly rhythmic. The basic electrophysiological and anatomical features of the neurons have been catalogued, so that the microlevel vitae for each cell in the network is impressively detailed. What is not understood is how the cells in the network interact to constitute a circuit that produces the rhythmic pattern. No one cell is a repos- itory for the cells rhythmic output; no one cell is itself the repository for the properties displayed by the network as a whole. Where then does the rhythmicity come from? Very roughly speaking, from the patterns of interactions among cells and the intrinsic properties of the component cells.
The network produces rhythmic patterns in the cells, which produce rhythmic activity in the constitutive electrons and ions. This is a classic example of higher level order controlling both macro and micro level outcomes.

The Principle of Biological Relativity: There is no privileged level of causation in biology: living organisms are multi-level open stochastic systems in which the behaviour at any level depends on higher and lower levels and cannot be fully understood in isolation

The first key claim I make is that as well as upward causation, downward causation takes place
15

This is an emergent phenomenon that cannot be deduced from the underlying physics per se because they involve constants that are not fundamental physical constants.

Thus this is an example of multiple realisability (Section 2.6), which characterizes topdown causation [Ellis 2016].

The theories may have to be described in terms of propensities rather than mathematical laws.

place by manufacturing processes such as making a transistor. In biology this occurs when cell division takes place at the cellular level, and when an organism gives birth to progeny at the organism level. The context of lower level dynamics changes completely in both cases. In the latter case, as Darwin already recognised, sexual selection takes place and determines outcomes, involving very complex social and psychological interactions that alter outcomes at the genetic and physical levels.
TD3B Deleting or Altering Lower Level elements is the complementary process that is crucial in biology. In developmental biology, apoptosis (programmed cell death) plays a key role for example in digit formation (separating fingers and thumbs), while in neural development, synaptic connections are pruned as development takes place [Wolpert et al 2002]. Cell are specialised to perform specific functions as growth takes place, altering their nature and behaviour. A fundamental biological process is Adaptive selection due to selection criteria which alters either the set of lower level elements by deletion as in Darwinian selection [Campbell 1974] and the functioning of the immune system, or selecting optimal configurations, as in neural network plasticity involved in learning.
The higher level types of downward causation: TD4 (Adaptive selection of goals) and TD5 (Adaptive selection of selection criteria) build on these ones [Ellis 2012] [Ellis 2016]. The key issue is whether any of these types of downward causation are really causally effective, or just redescriptions in convenient form of microphysical causation.
Aristotle's kinds of causation There is an important further point as regards causation. As Aristotle pointed out [Bodnar 2018], there are four kinds of causation that occur in the real world. This is discussed by ([Juarrero 2002]:2,125-128,143) ([Noble 2016]:176-179) and ([Scott 2002]:298-300) They are
• Material Cause: the physical stuff that is needed for an outcome; the stuff out of which it is made, e.g., the bronze of a statue. In biology this is the physical stuff, the chemical elements as characterised by the periodic table, that make biology possible.
• Formal Cause: which makes anything what it is and no other; the material cause necessary for some outcome must be given the appropriate form through the way in which the material is arranged e.g., the shape of a statue. In biology, this is the structure at each level that underlies function at that level: physiological systems [Hall 2016] and the underlying biomolecules such as proteins [Petsko and Ringe 2009].
• Efficient Cause: The primary source of the change or rest, the force that brings an action into being; nowadays in the Newtonian case taken to be the effect of forces on inert matter, in the quantum chemistry case, Schr¨odinger's equation.
• Final Cause: the goal or purpose towards which something aims: "that for the sake of which a thing is done".

the autonomy is relative
to certain conditions that hold for a given equivalence class

Essentially higher level variables and dynamics The higher level concepts are indispensible when multiple realisability occurs, firstly because they define the space of data dL relevant at Level L, and secondly because of (4), variables in this space cannot be represented as natural kinds at the lower level. Effective Laws EFL at level L can only be expressed at level L-1 in terms of an entire equivalence class at that level. One can only define that equivalence class by using concepts defined at level L.

A key point is that multiple realisability plays a fundamental role in strong emergence [Menzies 2003]. Any particular higher level state can be realised in a multiplicity of ways in terms of lower level states. In engineering or biological cases, a high level need determines the high level function and thus a high level structure that fulfills it. This higher structure is realised by suitable lower level structures, but there are billions of ways this can happen. It does not matter which of the equivalence class of lower level realisations is used to fulfill the higher level need, as long as it is indeed fulfilled. Consequently you cannot even express the dynamics driving what is happening in a sensible way at a lower level.

Types of downward causation The basic type of downward causation are as follows (developed from [Ellis 2012] [Noble 2012] [Noble 2016] [Ellis 2016]):
TD1A Boundary conditions are constraints on particles in a system arising from the environment7 as in the case of a cylinder determining pressure and temperature of the enclosed gas, or the shape of tongue and lips determining air vibrations and so spoken words. Structural Constraints are fairly rigid structures that determine possible micro states of particles that make up the structure, as in the case of a cylinder constraining the motion of a piston, or a skeleton that supports a body.
TD1B Channeling and Containing constraints are key forms of contextual causation shaping microbiological and neural outcomes. Channeling constraints determine where reactants or electrical currents can flow, as in blood capillaries in a body, wires in a computer, or neural axons and dendrites in a brain. Containing constraints confine reactants to a limited region, so preventing them from diffusing away and providing the context for reaction networks to function. A key case is a cell wall.
TD2A Gating and signalling constraints Gating constraints control ingress and egress to a container, as in the case of voltage gated ion channels in axons, or ligand gated ion channels in synapses. They function via conformational changes controlled by voltage differential in the former case, and molecular recognition of ligands in the latter case, thus underlying cell signalling processes [Berridge 2014].
TD2B Feedback control to attain goals is a cybernetic process where the difference between a goal and the actual state of a system generates an error signal that is fed back to a controller and causes corrective action, as in thermostats and engine governors [Wiener 1948]. In biology this is homeostasis, a crucial feature of physiology at all levels [Hall 2016]. Because of this closed causal loop, goals determine outcomes. Changing the goals changes both macro and micro outcomes, as in altering the setting on a thermostat. In biology, multilevel homeostatic systems are continually responding to internal changes and external perturbations [Billman 2020].
TD3A Creation of New Elements takes place in two ways. Creation of new lower level elements occurs in physics when crystal level conditions create quasiparticles such as phonons that play a key role in dynamics at the electron level [Ellis 2020a]. This is what [Gillett 2019] calls a Downward Constitutive relation. It occurs in biology when genes are read to create proteins, a contextual process [Gilbert and Epel 2009] controlled by gene regulatory networks according to higher level needs [Noble 2016]. Creation of new higher level elements restructures lower level relations and so alters lower level dynamics. In engineering this takes

7Carlo's example of Jupiter causing tides on Earth fits here: Jupiter is part of the Earth's environment, causing a detectable gravitational field at Marseilles.

if we interpret
top-down effects as higher-level constraining-relations on the possible lower-level states of a given system, top-down causation becomes a matter of understanding how higher levels define the boundary conditions of lower-level dynamics (see e.g., Ellis 2008; 2012; 2016; Green 2018; Moreno and Mossio 2015; Mossio and Moreno 2010). Constraints are here understood as physical conditions that limit the degree of freedom of a dynamic process, thus enabling only selected system states (Christiansen 2000; Hooker 2013; Juarrero 1998; 1999). Constraints are typically regarded as being at a higher spatial scale than the entities and operations of the constrained processes. In biology, constraints are often further defined according to functional levels in a hierarchically organized system (Pattee 1971; 1973; Salthe 1985; Wimsatt 2007). For instance, the shape and size of blood vessels enable efficient circulation by limiting the degrees of freedom of liquid motion, or blood flow. The interpretation of top-down causation as the ability of upper-level variables to set the context for lower-level ones may be seen as a "weaker" form of top-down causation (Emmeche et al. 2000). Yet, it allows for an understanding of how constraints productively can channel system states that are not possible to reach for an unconstrained system. The productive aspect of constraints can be illustrated by how an open respiratory system would not be able to provide sufficient gas exchange for a large organism. Similarly, the constraints provided by a sail on wind flow enable a sailboat to move. When one increases the drag by trimming the sail, one does not (effectively) change the operating cause (the wind). Rather, one modifies the structural constraints that channel a pressure difference across the windward and leeward side of the sail. Constraints thus have causal power by delimiting the space of possibilities


for lower-level causes. Without appeals to top-down constraints, we would not be able to explain why specific states are realized among multiple possible lower-level states and through such selections give rise to emergent properties.

biological functions must be
understood through goals of whole organisms and species, which again depends on higher-level features such as the environmental and evolutionary background

the requirement of higher-level boundary conditions to constrain the space of causal possibilities implies that some lower-level states cannot be accounted for without the boundary conditions imposed by higher-level structures.

certain emergent properties become possible only when constraining relations are applied on lower-level states.

many biological processes would not be possible without inter-level constraining relations

equivalences classes of lower level operations that give the same higher level response for a certain goal.

functional goals are higher-level concepts

In biology, the relations between higher and lower-scale variables are often illustrated in diagrams through feedback relations going both up and down (see e.g., Noble 2012; Lesne 2013)

generic models are explanatory because they show how many causal details are explanatorily irrelevant as long as stabilizing structures defining an equivalence class are in place

In both contexts, simple averaging over lower-scale variables would not enable scientists to predict macroscale material properties.

Stiffness is understood as the resistance of a material to deformation in response to applied force

For Ellis, a crucial feature of top-down causation is how coherent higher-level actions emerge from top-down constraints on lower-level dynamics (Ellis 2012, p. 128). Such constraints are often mathematically described as boundary conditions that delimit the set of lower-level variables (e.g., the set of initial conditions as inputs to the dynamics) within which a given function can be realized.

a feedback loop is here interpreted as a structural constraint that delimits the space of possibilities for lowerlevel dynamics.