Seven Pines Symposium XI

"Emergence: From Physics to Biology"

May 2-6, 2007

Program

(Breakfast 8:30-9:30 A.M., Lunch and Free Time 12:30-3:30 P.M.
Cocktails and Dinner 7:00 P.M.)

May 2 (Wednesday)

6:30 P.M. Welcome, Cocktails and Dinner

May 3 (Thursday)

9:15-9:45 Introduction: The Goals of the Seven Pines Symposium: Lee Gohlike (Outing Lodge)

9:45-10:45 Conceptual Framework: Concepts of Emergence

I. With Illustrations from Physics: Jeremy Butterfield (Oxford)

II. With Illustrations from Biology: Kenneth Schaffner (Pittsburgh)

10:45-11:15 Break

11:15-12:30 Discussion

3:30-4:30 History of Concepts of Emergence

I. In Physics: Michael Silberstein (Elizabethtown College)

II. In Biology: Manfred Laubichler (Arizona State)

4:30-5:00 Break

5:00-6:30 Discussion

May 4 (Friday)

9:30-10:30 Can Chemistry be Fully Reduced to Physics?

I. One Perspective: Jeff Ramsey (Smith)

II. Another Perspective: Eric Scerri (UCLA)

10:30-11:00 Break

11:00-12:30 Discussion

3:30-4:30 Can Biology be Reduced to Physics and Chemistry?

I. One Perspective: Eörs Szathmáry (Collegium Budapest)

II. Another Perspective: Kenneth Waters (Minnesota)

4:30-5:00 Break

5:00-6:30 Discussion

May 5 (Saturday)

9:30-10:30 Emergence and Reductionism

I. In Physics: Leo Kadanoff (Chicago)

II. In Biology: Ricardo Azevedo (Houston)

10:30-11:00 Break

11:00-12:30 Discussion

3:30-4:30 Complex Networks

I. In Physics: Stuart Kaufman (Calgary)

II. In Biololgy: Michael Travisano (Minnesota)

4:30-5:00 Break

5:00-6:30 Discussion

May 6 (Sunday)

9:30-11:30 Wrap-Up Session

Participants

Founder and Introducer

Lee Gohlike

Outing Lodge at Pine Point, 11661 Myeron Road, Stillwater, MN 55082

Other Speakers

Jeremy Butterfield (University of Oxford) Ph.D. 1984 University of Cambridge

Research Interests: Philosophy of quantum theory, relativity, and classical mechanics.

Kenneth F. Schaffner (University of Pittsburgh) Ph.D., M.D. 1967, '86 Columbia U, U Pittsburgh

Research Interests: Ethical and philosophical issues in human behavioral and psychiatric genetics.

Michael Silberstein (Elizabethtown College) Ph.D. 1994 University of Oklahoma

Research Interests: Philosophy of physics and philosophy of cognitive neuroscience and how both bear on questions of reduction, emergence, and explanation.

Manfred D. Laubichler (Arizona State University) Ph.D. 1997 Yale University

Research Interests: Theoretical biology; evolutionary developmental biology; evolutionary innovations and novelties; evolution of complex social systems; history of biology.

Jeffry L. Ramsey (Smith College) Ph.D. 1990 University of Chicago

Research Interests: Explanation, reduction, and theory structure in the modern physical and biological sciences, especially chemistry and biochemistry.

Eric R. Scerri (UCLA) Ph.D. 1992 King's College London

Research Interests: Philosophy of chemistry, reduction of chemistry to quantum mechanics, emergence, history and philosophy of the periodic system, chemical education.

Eörs Szathmáry (Collegium Budapest) Ph.D. 1987 Roland Eötvös University

Research Interests: Major transitions in evolution such as the origin of the cell, the appearance of multicellularity, or the emergence of natural language.

C. Kenneth Waters (University of Minnesota) Ph.D. 1985 Indiana University

Research Interests: The nature and development of scientific knowledge and clarification of the conceptual basis of particular biological systems.

Leo Kadanoff (University of Chicago) Ph.D. 1960 Harvard University

Research Interests: Condensed matter physics; hydrodynamics; informal education.

Ricardo B.R. Azevedo (University of Houston) Ph.D. 1997 University of Edinburgh

Research Interests: Phenotypic variation and its influence on evolution.

Stuart A. Kauffman (University of Calgary) M.D. 1968 U of California, San Francisco

Research Interests: Developmental genetics, theoretical biology, evolution, and the origin of life.

Michael Travisano (University of Minnesota) Ph.D. 1993 Michigan State University

Research Interests: The causes and consequences of adaptation.

Other Participants

John Beatty (University of British Columbia) Ph.D. 1979 Indiana University

Research Interests: Theoretical foundations, methodology, and socio-political dimensions of genetics and evolutionary biology.

Mark E. Borrello (University of Minnesota) Ph.D. 2002 Indiana University

Research Interests: The question of the levels (gene to species) at which natural selection operates; history of behavioral biology.

John Earman (University of Pittsburgh) Ph.D. 1968 Princeton University

Research Interests: History of physics, foundations of physics.

Geoffrey Hellman (University of Minnesota) Ph.D. 1972 Harvard University

Research Interests: Philosophy of mathematics, philosophy of quantum mechanics, physics and science more generally, philosophy of logic, musical aesthetics.

Don Howard (Notre Dame University) Ph.D. 1979 Boston University
Research Interests: History and philosophy of physics, especially foundations of quantum mechanics and spacetime theory, Einstein, Bohr, history of the philosophy of science.

Michel Janssen (University of Minnesota) Ph.D. 1995 University of Pittsburgh
Research Interests: History of modern physics, especially relativity and quantum revolutions, Einstein, philosophy of science.

Alan C. Love (University of Minnesota) Ph.D. 2005 University of Pittsburgh
Research Interests: The nature of conceptual change and explanation in the biological sciences, specifically within evolutionary developmental biology.

Antigone M. Nounou (University of Minnesota) Ph.D. 2002 University of London
Research Interests: Philosophy of physics with emphasis on quantum field theories.

Serge Rudaz (University of Minnesota) Ph.D. 1979 Cornell University
Research Interests: Theoretical high energy physics.

Philip Stamp (University of British Columbia) Ph.D. 1984 University of Sussex
Research Interests: Large-scale quantum phenomena, strongly-correlated condensed matter systems, statistical mechanics, and field theory.

Roger H. Stuewer (University of Minnesota) Ph.D. 1968 University of Wisconsin
Research Interests: History of quantum and nuclear physics.

William G. Unruh (University of British Columbia) Ph.D. 1971 Princeton University
Research Interests: Quantum gravity, quantum computing, philosophical issues in the interpretation of quantum mechanics.

Robert M. Wald (University of Chicago) Ph.D. 1972 Princeton University
Research Interests: General relativity, quantum field theory in curved spacetime, quantum gravity.

Jeremy Butterfield

Statement

and

Readings

Emergence in Physics[1]

J. Butterfield: jb56@cam.ac.uk: 20 February 2007

0. Introduction I expect my talk will cover the following topics.

(1) I will contrast emergence with the more general idea of "good" variables.

(2) I will contrast emergence with the failure of (i) reduction in the sense of definitional extension; and of (ii) supervenience.

(3) I propose to think of emergence in terms of limiting relations between theories, or more generally in terms of *regimes* of a theory.

(4) I will illustrate this with the $N \to \infty$ of algebraic quantum statistical mechanics.

A few more detailed notes, and some references, follow.

1. Emergence vs. good variables Emergence as properties/behaviour that are both novel and robust relative to some comparison class: especially one given by a theory of the micro-details.

Novelty and robustness are liable to be ambiguous, even controversial or subjective: even for a fixed comparison class. (E.g.: for novelty, the philosophical debate about identity of properties; for robustness, the different definitions of stability.) But nevermind!

"Good" variables and-or approximation schemes:—
Here "good" is ambiguous between:
small in number[2] and autonomous (uncoupled equations);
easily calculated with;
suited to *given* problem;
insightful, eg by suggestiveness for other theory, or suitability to alien procedure eg quantization.

Accordingly, good variables/schemes vary:
(i) in their scope (from a single problem, eg set by boundary conditions, to a whole theory); and
(ii) in our ways to seek them, e.g. reduction by exploiting a symmetry, or neglecting the negligible—itself various, e.g. coarse-graining (averaging) or dressing (re-factorization of state-space).

Emergence is closest to the first meaning of "good": robust $\approx$ autonomous. But emergence is often taken to include:
(i) more about the idea of novelty; and-or
(ii) other general ideas such as non-linearity, heirarchy, scaling, complexity; and-or
(iii) proposed paradigms for theory-development, eg renormalization group or cellular au-

[1] Draft Summary of a talk for the Seven Pines Symposium, 2-6 May 2007.

[2] 'Small in number' does not always involve reduction. Often we understand/model a finite-dimensional system more successfully by idealizing it as infinite-dimensional: despite atomism, continuous models of sound or fluid flow are successful. I think the idealization $N \to \infty$ for large quantum systems is similar—and similarly justified.

tomata or self-organized criticality. (Frigg (2003) criticizes the claims that self-organized criticality is a universal theory.)

2. Emergence vs. reduction; and vs. supervenience All the above ideas, about both good variables/schemes and emergence, seem independent of philosophers' proposals that emergence is: either (i) failure of reduction (in logicians' sense of *definitional extension*); or (ii) failure of supervenience.

(Note: The idea of supervenience, also called 'determination', is a relation between families of properties: viz. that total matching of any two entities as regards one family of properties (called the *subvening* family) implies their total matching as regards the other family (the *supervening* family). Most ways of making this idea precise make it a weakening of definitional extension: namely, a weakening that allows one or more of the definitions (of a property in the supervening family in terms of the subvening family) to be infinitely long. But I will not pause on the technicalities here.)

As to (i):— I think a theory could describe novel and robust properties/behaviour, while being a definitional extension of another. In other words: the power of reduction is stronger than commonly thought. The $N \to \infty$ limit in quantum statistical mechanics (Paragraph 4) is an example: there is reduction (using a suitably strong mathematical language), with novelty—viz. superselection sectors. (Other (nearly synonymous!) buzz-words are: classical observables, inequivalent representations, symmetry breaking.)

That is compatible with saying that, since supervenience is weaker than definitional extension, all cases of "supervenience-but-not-definitional-extension" are cases of emergence. But I also deny this (as do other philosophers); (Butterfield and Isham (1999), Section 2, pp. 114-126; Humphreys 1997, Section 2).

As to (ii):— It of course follows from what I just said—that there are cases of emergence which are also definitional extensions—that (ii) is false. That is: emergence is not failure of supervenience.

But there is a more interesting (i.e. controversial!) point here. Several philosophers (some of them at 7 Pines!) have argued that

(a): quantum entanglement is an important clear-cut case of a failure of supervenience: more precisely, failure of supervenience of the state of a whole on the states of its parts—also known as *mereological* supervenience.

(b): quantum entanglement underpins striking cases of emergence, including the cases under (i), i.e. superselection sectors. (And they emphasise that these cases are not just striking, but also well-understood, and so a surer guide to philosophical understanding of emergence than eg the mind-body relation.)

(References for (a) and (b) include: Howard (2003, pp. 6-17), Humphreys (1997a Section 6), Silberstein (2001, pp. 73-78; 2002, pp. 96-98)), Silberstein & McGeever (1999, p. 187-189).)

I have a bone to pick here! In effect, I agree with the letter of (a) and (b), but not the spirit. (Huttemann's position (2005) is broadly similar to mine.) More precisely:

Though (a) is true:— Quantum theory, with its entangled states, conforms to close cousins of mereological supervenience. For think in terms of the quantum state as a complex-valued

function ψ on the configuration space, e.g. $\psi : \mathbb{R}^6 \rightarrow$ |

C for two spinless particles. Quantum entanglement means that ψ cannot be represented as an assignment of two complex numbers to each point of physical space $\mathbb{R}^3$ ... but only as an assignment of a single complex number to a point of $\mathbb{R}^3 \times \mathbb{R}^3$. But these are, I submit, close cousins! So I would not want broad, cherished, metaphysical theses of reduction, or of anti-holistic supervenience, to be violated by the latter sort of state-space.

More generally, we should recall how quantum physics, no less than classical physics, illustrates two triumphs of reductionism that are so endemic in the development of (and successes of) both quantum and classical physics that we tend to forget them. Namely:

[1]; The uniform rules for defining a composite system's state-space and its quantities; (viz. Cartesian products in classical physics; tensor products in quantum theory); and

[2]: *Pace* the 'British emergentism' of Broad et al. in the inter-war period: the non-existence of "configurational forces", i.e. forces that only come into play when the number of bodies/particles/degrees of freedom exceeds some number. Or to put it more positively: the fact that both quantum and classical physics manage with only 2-body forces (potentials).

Though (b) is true:— In the cited cases of emergence (i.e. superselection sectors, superconductivity), quantum entanglement is not, I submit, the "main fuel". Other features are at least equally important: especially, the $N \rightarrow \infty$ limit—cf. Paragraph 4.

(Humphreys (1997, Sections 3 and 4) makes essentially this point: but ties it, in my view unnecessarily, to his advocacy (1997a, Section 5) of a physical operation of "fusion".)

(I should admit, of course, that perhaps the most important "fuel" is not the fancy mathematical physics of this limit!... but the creative, heuristic physics of writing down the "right" interaction, e.g. the BCS Hamiltonian. This returns us to Paragraph 1 on good variables, and to [2] just above.)

3. Emergence and limiting relations I now propose to steer a middle course between generalities about emergence (Paragraph 1 and 2 above), and the proposed paradigms listed at the end of Paragraph 1. I propose to consider limiting relations (in general: for some states, some quantities, some parameter-values) between theories: or more generally, *regimes* of a theory.

In this framework, emergence will especially concern regimes for composite, especially "large", systems. Accordingly, Paragraph 4 will look at the case of quantum systems with an infinite number of particles/degrees of freedom, i.e. the $N \rightarrow \infty$ limit of quantum mechanics. (This will be a special case of the general idea of a classical limit of quantum theory, given by $\hbar \rightarrow 0$.)

Theories T_0 and T_κ postulate state-spaces Γ_0 and Γ_κ, and sets (algebras) of quantities $\mathcal{A}_0$ and $\mathcal{A}_\kappa$. Think of κ as a real parameter labelling a "version" of a generic theory: in our case, $\kappa \equiv \hbar$, the generic theory is quantum mechanics, and T_0 is classical mechanics.

So there are two main kinds of limiting relation: about states and quantities.

For all, or maybe just some, states $s_0 \in \Gamma_0$, there is a sequence of states $s_\kappa \in \Gamma_\kappa$ such that $\mathrm{s}_\kappa \rightarrow s_0$.

For all, maybe some, quantities $A_0 \in \mathcal{A}_0$, there is a sequence of quantities $A_\kappa \in \mathcal{A}_\kappa$ such

that $\mathrm{A}_\kappa \to A_0$.

Since the state-spaces/algebras can have different mathematical structures (evidently do so, for the quantum-classical case!), both $\to$ s need to be clarified.

In Paragraph 4, we will follow Landsman (2006) in using deformation quantization: a recently-developed framework which, Landsman argues, makes the limiting process "as clear as it can be".

For the moment, note just that in the philosophical literature, Batterman has stressed the importance of such "singular limits" for understanding inter-theoretic relations. Indeed, he considers the quantum-classical case, approaching it on analogy with the case of wave optics/geometric optics, ie geometric optics as the short-wavelength limit of wave optics. In the quantum-classical case, this amounts to the WKB or "semi-classical" approach to understanding the $\hbar \to 0$ limit. Landsman (2006, Section 5.5) argues that this is a very limited approach, and I will not go into it, except to report that:—

Batterman argues that in both these cases (mechanics, quantum or classical; and optics, wave or geometric), there is emergence in the strong sense that the "deeper/later" theory, i.e. quantum mechanics/wave optics, *cannot* explain all the phenomena that occur in the short-wavelength limit: that explanations need to appeal to the concepts of the "shallower/earlier" ("supervening") theory. For references, and a critique of Batterman's position (to my mind: persuasive), cf. Belot (2003).

Anyway: in general: we expect the $\to$ s to mesh in that:
(i) at (appropriate) s_0, A_0, the values obey the corresponding relation:

$$\mathrm{A}_\kappa(s_\kappa) \to A_0(s_0);$$

and maybe (ii) commutation with time-evolution.

I propose that we should not mind which regimes—combinations of states, quantities, and parameter-values—to call 'emergent'. This depends on which ideas from Paragraphs 1 and 2 above are emphasised, and so is in part ambiguous/subjective.

4. The $N \to \infty$ limit of quantum mechanics I propose (if I have time after the fights above!) to expound aspects of the $N \to \infty$ limit of quantum mechanics: (a special case of the general idea of $\hbar \to 0$). I will follow Landsman (2006, Section 6), and talk in terms of deformation quantization.

Though this material is very restricted, it is enough to capture phenomena often called 'emergent': eg chirality and knot-type of molecules, temperature and other macroscopic observables of substances.

It is also a preamble to other examples of emergence: eg KMS states, with their various kinds of robustness (cf. Emch 2006).

This material will be in effect an ode to reductionism: even to philosophers' definitional extension—using amazingly short definitions! (... provided you help yourself to a sufficiently powerful mathematical language...)

I therefore recommend reading Landsman (2006, Section 6). Since Landsman's paper is technically daunting, let me add some guidelines, based on limitations of my discussion.

Within Section 6, I will only hope to treat Sections 6.1-6.4; i.e. I will ignore Section 6.5-6.6.

Within Landsman's earlier Sections, the only essential preliminaries are some parts of (i) Section 4.3 on deformation quantization, and (ii) Section 5.1. Namely, we need: from Section 4.3, the ideas of a continuous field of algebras, and (thereby) a deformation quantization; and from Section 5.1, the idea of a continuous field of states.

(By the way: for the limitations of geometric quantization, and the WKB approach to the classical limit, cf. also Landsman's Sections 4.4 and 5.5, respectively. And for decoherence, cf. Landsman's Section 7.1.)

References :

Belot, G. (2003), 'Whose Devil? Which Details?', available at:
http://philsci-archive.pitt.edu/archive/00001515/

Butterfield, J. and Isham, C. (1999), 'On the Emergence of Time in Quantum Gravity', in *The Arguments of Time*, British Academy and O.U.P., 1999, pp. 111-168; and at: gr-qc/9901024.

Emch, G. (2006), 'Quantum statistical physics', in J. Butterfield and J. Earman (eds.) *The Handbook of the Philosophy of Physics*, North-Holland: Elsevier.

Frigg, R. (2003), 'Self-organised criticality-what it is and what it isn't' *Studies in the History and Philosophy of Science* **34**, pp. 613-632.

Howard, D. (2003), 'Reduction and emergence in the physical sciences: some lessons from the particle physics—condensed matter physics debate', available at:
http://www.nd.edu/ dhoward1/ReductionandEmergence.pdf

Humphreys, P. (1997), 'Emergence, not supervenience', *Philosophy of Science* **64**, pp. S337-S345.

Humphreys, P. (1997a), 'How properties emerge', *Philosophy of Science* **64**, pp. 1-17.

Huttemann, A. (2005), 'Explanation, Emergence and Quantum Entanglement', *Philosophy of Science* **72**, pp. 114-127).

Landsman, N. (2006), 'Between classical and quantum', in J. Butterfield and J. Earman (eds.) *The Handbook of the Philosophy of Physics*, Elsevier. Available at: http://philsci-archive.pitt.edu/archive/00002328/
or at: http://arxiv.org/abs/quant-ph/0506082

Silberstein, M. and McGeever, J. (1999). 'The search for ontological emergence' *Philosophical Quarterly* **49**, pp. 182-200.

Silberstein, M. (2001), 'Converging on emergence: Consciousness, causation, and explanation', *Journal of Consciousness Studies* **8**, pp. 61-98.

Silberstein, M. (2002), 'Reduction, emergence, and explanation' In Peter Machamer and Michael Silberstein (eds.) *The Blackwell Guide to the Philosophy of Science*, pp 80-107. Malden: Blackwell Publishers.

Recommended Reading: Landsman, N. (2006): Sections 6.1 to 6.4 only.

Between classical and quantum*

N.P. Landsman

Radboud Universiteit Nijmegen
Institute for Mathematics, Astrophysics, and Particle Physics
Toernooiveld 1, 6525 ED NIJMEGEN
THE NETHERLANDS

email landsman@math.ru.nl

May 23, 2006

Abstract

The relationship between classical and quantum theory is of central importance to the philosophy of physics, and any interpretation of quantum mechanics has to clarify it. Our discussion of this relationship is partly historical and conceptual, but mostly technical and mathematically rigorous, including over 500 references. For example, we sketch how certain intuitive ideas of the founders of quantum theory have fared in the light of current mathematical knowledge. One such idea that has certainly stood the test of time is Heisenberg's 'quantum-theoretical *Umdeutung* (reinterpretation) of classical observables', which lies at the basis of quantization theory. Similarly, Bohr's correspondence principle (in somewhat revised form) and Schrödinger's wave packets (or coherent states) continue to be of great importance in understanding classical behaviour from quantum mechanics. On the other hand, no consensus has been reached on the Copenhagen Interpretation, but in view of the parodies of it one typically finds in the literature we describe it in detail.

On the assumption that quantum mechanics is universal and complete, we discuss three ways in which classical physics has so far been believed to emerge from quantum physics, namely in the limit $\hbar \to 0$ of small Planck's constant (in a finite system), in the limit $N \to \infty$ of a large system with N degrees of freedom (at fixed $\hbar$), and through decoherence and consistent histories. The first limit is closely related to modern quantization theory and microlocal analysis, whereas the second involves methods of C^*-algebras and the concepts of superselection sectors and macroscopic observables. In these limits, the classical world does not emerge as a sharply defined objective reality, but rather as an approximate *appearance* relative to certain "classical" states and observables. Decoherence subsequently clarifies the role of such states, in that they are "einselected", i.e. robust against coupling to the environment. Furthermore, the nature of classical observables is elucidated by the fact that they typically define (approximately) consistent sets of histories.

This combination of ideas and techniques does not quite resolve the measurement problem, but it does make the point that classicality results from the *elimination* of certain states and observables from quantum theory. Thus the classical world is not created by observation (as Heisenberg once claimed), but rather by the lack of it.

*To appear in Elsevier's forthcoming *Handbook of the Philosophy of Science*, Vol. 2: *Philosophy of Physics* (eds. John Earman & Jeremy Butterfield). The author is indebted to Stephan de Bièvre, Jeremy Butterfield, Dennis Dieks, Jim Hartle, Gijs Tuynman, Steven Zelditch, and Wojciech Zurek for detailed comments on various drafts of this paper. The final version has greatly benefited from the 7 Pines Meeting on 'The Classical-Quantum Borderland' (May, 2005); the author wishes to express his gratitude to Lee Gohlike and the Board of the 7 Pines Meetings for the invitation, and to the other speakers (M. Devoret, J. Hartle, E. Heller, G. 't Hooft, D. Howard, M. Gutzwiller, M. Janssen, A. Leggett, R. Penrose, P. Stamp, and W. Zurek) for sharing their insights with him.

6 The limit $N \to \infty$

In this section we show to what extent classical physics may approximately emerge from quantum theory when the size of a system becomes large. Strictly classical behaviour would be an idealization reserved for the limit where this size is infinite, which we symbolically denote by "$\lim N \to \infty$". As we shall see, mathematically speaking this limit is a special case of the limit $\hbar \to 0$ discussed in the previous chapter. What is more, we shall show that formally the limit $N \to \infty$ even falls under the heading of continuous fields of C^*-algebras and deformation quantization (see Subsection 4.3.) Thus the 'philosophical' nature of the idealization involved in assuming that a system is infinite is much the same as that of assuming $\hbar \to 0$ in a quantum system of given (finite) size; in particular, the introductory comments in Section 1 apply here as well.

An analogous discussion pertains to the derivation of thermodynamics from statistical mechanics (Emch & Liu, 2002; Batterman, 2005). For example, *in theory* phase transitions only occur in infinite systems, but *in practice* one sees them every day. Thus it appears to be valid to approximate a pot of 10^{23} boiling water molecules by an infinite number of such molecules. The basic point is that the distinction between microscopic and macroscopic regimes is unsharp unless one admits infinite systems as an idealization, so that one can simply say that microscopic systems are finite, whereas macroscopic systems are infinite. This procedure is eventually justified by the results it produces.

Similarly, in the context of quantum theory classical behaviour is simply not found in finite systems (when $\hbar > 0$ is fixed), whereas, as we shall see, it *is* found in infinite ones. Given the observed classical nature of the macroscopic world,[255] at the end of the day one concludes that the idealization in question is apparently a valid one. One should not be confused by the fact that the error in the number of particles this approximation involves (viz. $\infty - 10^{23} = \infty$) is considerably larger than the number of particles in the actual system. If all of the 10^{23} particles in question were *individually* tracked down, the approximation is indeed a worthless ones, but the point is rather that the limit $N \to \infty$ is valid whenever *averaging* over $N = 10^{23}$ particles is well approximated by averaging over an arbitrarily larger number N (which, then, one might as well let go to infinity). Below we shall give a precise version of this argument.

Despite our opening comments above, the quantum theory of infinite systems has features of its own that deserve a separate section. Our treatment is complementary to texts such as Thirring (1983), Strocchi (1985), Bratteli & Robinson (1987), Haag (1992), Araki (1999), and Sewell (1986, 2002), which should be consulted for further information on infinite quantum systems. The theory in Subsections 6.1 and 6.5 is a reformulation in terms of continuous field of C^*-algebras and deformation quantization of the more elementary parts of a remarkable series of papers on so-called quantum mean-field systems by Raggio & Werner (1989, 1991), Duffield & Werner (1992a,b,c), and Duffield, Roos, & Werner (1992). These models have their origin in the treatment of the BCS theory of superconductivity due to Bogoliubov (1958) and Haag (1962), with important further contributions by Thirring & Wehrl (1967), Thirring (1968), Hepp (1972), Hepp & Lieb (1973), Rieckers (1984), Morchio & Strocchi (1987), Duffner & Rieckers (1988), Bona (1988, 1989, 2000), Unnerstall (1990a, 1990b), Bagarello & Morchio (1992), Sewell (2002), and others.

6.1 Macroscopic observables

The large quantum systems we are going to study consist of N copies of a single quantum system with unital algebra of observables $\mathcal{A}_1$. Almost all features already emerge in the simplest example $\mathcal{A}_1 = M_2(\mathbb{C})$ (i.e. the complex 2×2 matrices), so there is nothing wrong with having this case in mind as abstraction increases.[256] The aim of what follows is to describe in what precise sense macroscopic

relating (in)complete classical motion in a potential to (lack of) essential selfadjointness of the corresponding Schrödinger operator, it is usually the case that completeness implies essential selfadjointness, and vice versa. See Reed & Simon (1975), Appendix to §X.1, where the reader may also find examples of classically incomplete but quantum-mechanically complete motion, and vice versa. Now, here is the central point for the present discussion: as probably first noted by Hepp (1974), *different self-adjoint extensions have the same classical limit* (in the sense of (5.20) or similar criteria), namely the given *incomplete* classical dynamics. This proves that complete quantum dynamics can have incomplete motion as its classical limit. However, much remains to be understood in this area. See also Earman (2005, 2006).

[255] With the well-known mesoscopic exceptions (Leggett, 2002; Brezger et al., 2002; Chiorescu et al., 2003; Marshall et al., 2003; Devoret et al., 2004).

[256] In the opposite direction of greater generality, it is worth noting that the setting below actually incorporates quantum systems defined on general lattices in $\mathbb{R}^n$ (such as $\mathbb{Z}^n$). For one could relabel things so as to make $\mathcal{A}_{1/N}$ below the algebra

observables (i.e. those obtained by averaging over an infinite number of sites) are "classical".

From the single C^*-algebra $\mathcal{A}_1$, we construct a continuous field of C^*-algebras $\mathcal{A}^{(c)}$ over

$$I = 0 \cup 1/\mathbb{N} = \{0, \ldots, 1/N, \ldots, \tfrac{1}{3}, \tfrac{1}{2}, 1\} \subset [0,1], \tag{6.1}$$

as follows. We put

$$\begin{aligned} \mathcal{A}_0^{(c)} &= C(\mathcal{S}(\mathcal{A}_1)); \\ \mathcal{A}_{1/N}^{(c)} &= \mathcal{A}_1^N, \end{aligned} \tag{6.2}$$

where $\mathcal{S}(\mathcal{A}_1)$ is the state space of $\mathcal{A}_1$ (equipped with the weak*-topology)[257] and $\mathcal{A}_1^N = \hat{\otimes}^N \mathcal{A}_1$ is the (spatial) tensor product of N copies of $\mathcal{A}_1$.[258] This explains the suffix c in $\mathcal{A}^{(c)}$: it refers to the fact that the limit algebra $\mathcal{A}_0^{(c)}$ is *c*lassical or *c*ommutative.

For example, take $\mathcal{A}_1 = M_2(\mathbb{C})$. Each state is given by a density matrix, which is of the form

$$\rho(x,y,z) = \tfrac{1}{2} \begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix}, \tag{6.3}$$

for some $(x,y,z) \in \mathbb{R}^3$ satisfying $x^2 + y^2 + z^2 \leq 1$. Hence $\mathcal{S}(M_2(\mathbb{C}))$ is isomorphic (as a compact convex set) to the three-ball B^3 in $\mathbb{R}^3$. The pure states are precisely the points on the boundary,[259] i.e. the density matrices for which $x^2 + y^2 + z^2 = 1$ (for these and these alone define one-dimensional projections).[260]

In order to define the continuous sections of the field, we introduce the *symmetrization maps* $j_{NM} : \mathcal{A}_1^M \to \mathcal{A}_1^N$, defined by

$$j_{NM}(A_M) = S_N(A_M \otimes 1 \otimes \cdots \otimes 1), \tag{6.4}$$

where one has $N - M$ copies of the unit $1 \in \mathcal{A}_1$ so as to obtain an element of $\mathcal{A}_1^N$. The symmetrization operator $S_N : \mathcal{A}_1^N \to \mathcal{A}_1^N$ is given by (linear and continuous) extension of

$$S_N(B_1 \otimes \cdots \otimes B_N) = \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} B_{\sigma(1)} \otimes \cdots \otimes B_{\sigma(N)}, \tag{6.5}$$

where $\mathfrak{S}_N$ is the permutation group (i.e. symmetric group) on N elements and $B_i \in \mathcal{A}_1$ for all $i = 1, \ldots, N$. For example, $j_{N1} : \mathcal{A}_1 \to \mathcal{A}_1^N$ is given by

$$j_{N1}(B) = \overline{B}^{(N)} = \frac{1}{N} \sum_{k=1}^{N} 1 \otimes \cdots \otimes B_{(k)} \otimes 1 \cdots \otimes 1, \tag{6.6}$$

where $B_{(k)}$ is B seen as an element of the k'th copy of $\mathcal{A}_1$ in $\mathcal{A}_1^N$. As our notation $\overline{B}^{(N)}$ indicates, this is just the 'average' of B over all copies of $\mathcal{A}_1$. More generally, in forming $j_{NM}(A_M)$ an operator $A_M \in \mathcal{A}_1^M$ that involves M sites is averaged over $N \geq M$ sites. When $N \to \infty$ this means that one forms a *macroscopic* average of an M-particle operator.

---

of observables of all lattice points Λ contained in, say, a sphere of radius N. The limit $N \to \infty$ then corresponds to the limit $\Lambda \to \mathbb{Z}^n$.

[257] In this topology one has $\omega_\lambda \to \omega$ when $\omega_\lambda(A) \to \omega(A)$ for each $A \in \mathcal{A}_1$.

[258] When $\mathcal{A}_1$ is finite-dimensional the tensor product is unique. In general, one needs the *projective* tensor product at this point. See footnote 90. The point is the same here: any tensor product state $\omega_1 \otimes \cdots \otimes \omega_N$ on $\otimes^N \mathcal{A}_1$ - defined on elementary tensors by $\omega_1 \otimes \cdots \otimes \omega_N(A_1 \otimes \cdots \otimes A_N) = \omega_1(A_1) \cdots \omega_N(A_N)$ - extends to a state on $\hat{\otimes}^N \mathcal{A}_1$ by continuity.

[259] The *extreme boundary* $\partial_e K$ of a convex set K consists of all $\omega \in K$ for which $\omega = p\rho + (1-p)\sigma$ for some $p \in (0,1)$ and $\rho, \sigma \in K$ implies $\rho = \sigma = \omega$. If $K = \mathcal{S}(\mathcal{A})$ is the state space of a C^*-algebra $\mathcal{A}$, the extreme boundary consists of the pure states on $\mathcal{A}$ (the remainder of $\mathcal{S}(\mathcal{A})$ consisting of mixed states). If K is embedded in a vector space, the extreme boundary $\partial_e K$ may or may not coincide with the geometric boundary ∂K of K. In the case $K = B^3 \subset \mathbb{R}^3$ it does, but for an equilateral triangle in $\mathbb{R}^2$ it does not, since $\partial_e K$ merely consists of the corners of the triangle whereas the geometric boundary includes the sides as well.

[260] Eq. (6.3) has the form $\rho(x,y,z) = \frac{1}{2}(x\sigma_x + y\sigma_y + z\sigma_z)$, where the σ_i are the Pauli matrices. This yields an isomorphism between $\mathbb{R}^3$ and the Lie algebra of $SO(3)$ in its spin-$\frac{1}{2}$ representation $\mathcal{D}_{1/2}$ on $\mathbb{C}^2$. This isomorphism intertwines the defining action of $SO(3)$ on $\mathbb{R}^3$ with its adjoint action on $M_2(\mathbb{C})$. I.e., for any rotation R one has $\rho(R\mathbf{x}) = \mathcal{D}_{1/2}(R)\rho(\mathbf{x})\mathcal{D}_{1/2}(R)^{-1}$. This will be used later on (see Subsection 6.5).

We say that a sequence $A = (A_1, A_2, \cdots)$ with $A_N \in \mathcal{A}_1^N$ is *symmetric* when

$$A_N = j_{NM}(A_M) \tag{6.7}$$

for some fixed M and all $N \geq M$. In other words, the tail of a symmetric sequence entirely consists of 'averaged' or 'intensive' observables, which become macroscopic in the limit $N \to \infty$. Such sequences have the important property that they commute in this limit; more precisely, if A and A' are symmetric sequences, then

$$\lim_{N\to\infty} \|A_N A'_N - A'_N A_N\| = 0. \tag{6.8}$$

As an enlightening special case we take $A_N = j_{N1}(B)$ and $A'_N = j_{N1}(C)$ with $B, C \in \mathcal{A}_1$. One immediately obtains from the relation $[B_{(k)}, C_{(l)}] = 0$ for $k \neq l$ that

$$\left[\overline{B}^{(N)}, \overline{C}^{(N)}\right] = \frac{1}{N}\overline{[B,C]}^{(N)}. \tag{6.9}$$

For example, if $\mathcal{A}_1 = M_2(\mathbb{C})$ and if for B and C one takes the spin-$\frac{1}{2}$ operators $S_j = \frac{\hbar}{2}\sigma_j$ for $j = 1, 2, 3$ (where σ_j are the Pauli matrices), then

$$\left[\overline{S}_j^{(N)}, \overline{S}_k^{(N)}\right] = i\frac{\hbar}{N}\epsilon_{jkl}\overline{S}_l^{(N)}. \tag{6.10}$$

This shows that averaging one-particle operators leads to commutation relations formally like those of the one-particle operators in question, but with Planck's constant $\hbar$ replaced by a variable $\hbar/N$. For constant $\hbar = 1$ this leads to the interval (6.1) over which our continuous field of C^*-algebras is defined; for any other constant value of $\hbar$ the field would be defined over $I = 0 \cup \hbar/\mathbb{N}$, which of course merely changes the labeling of the C^*-algebras in question.

We return to the general case, and denote a section of the field with fibers (6.2) by a sequence $A = (A_0, A_1, A_2, \cdots)$, with $A_0 \in \mathcal{A}_0^{(c)}$ and $A_N \in \mathcal{A}_1^N$ as before (i.e. the corresponding section is $0 \mapsto A_0$ and $1/N \mapsto A_N$). We then complete the definition of our continuous field by declaring that a sequence A defines a *continuous* section iff:

- $(A_1, A_2, \cdots)$ is *approximately symmetric*, in the sense that for any $\varepsilon > 0$ there is an N_ε and a symmetric sequence A' such that $\|A_N - A'_N\| < \varepsilon$ for all $N \geq N_\varepsilon$;[261]
- $A_0(\omega) = \lim_{N\to\infty} \omega^N(A_N)$, where $\omega \in \mathcal{S}(\mathcal{A}_1)$ and $\omega^N \in \mathcal{S}(\mathcal{A}_1^N)$ is the tensor product of N copies of ω, defined by (linear and continuous) extension of

$$\omega^N(B_1 \otimes \cdots \otimes B_N) = \omega(B_1) \cdots \omega(B_N). \tag{6.11}$$

  This limit exists by definition of an approximately symmetric sequence.[262]

It is not difficult to prove that this choice of continuous sections indeed defines a continuous field of C^*-algebras over $I = 0 \cup 1/\mathbb{N}$ with fibers (6.2). The main point is that

$$\lim_{N\to\infty} \|A_N\| = \|A_0\| \tag{6.12}$$

whenever $(A_0, A_1, A_2, \cdots)$ satisfies the two conditions above.[263] This is easy to show for symmetric sequences,[264] and follows from this for approximately symmetric ones.

Consistent with (6.8), we conclude that in the limit $N \to \infty$ the macroscopic observables organize themselves in a commutative C^*-algebra isomorphic to $C(\mathcal{S}(\mathcal{A}_1))$.

[261] A symmetric sequence is evidently approximately symmetric.

[262] If $(A_1, A_2, \cdots)$ is symmetric with (6.7), one has $\omega^N(A_N) = \omega^M(A_M)$ for $N > M$, so that the tail of the sequence $(\omega^N(A_N))$ is even independent of N. In the approximately symmetric case one easily proves that $(\omega^N(A_N))$ is a Cauchy sequence.

[263] Given (6.12), the claim follows from Prop. II.1.2.3 in Landsman (1998) and the fact that the set of functions A_0 on $\mathcal{S}(\mathcal{A}_1)$ arising in the said way are dense in $C(\mathcal{S}(\mathcal{A}_1))$ (equipped with the supremum-norm). This follows from the Stone–Weierstrass theorem, from which one infers that the functions in question even exhaust $\mathcal{S}(\mathcal{A}_1)$.

[264] Assume (6.7), so that $\|A_N\| = \|j_{NN}(A_N)\|$ for $N \geq M$. By the C^*-axiom $\|A^*A\| = \|A^2\|$ it suffices to prove (6.12) for $A_0^* = A_0$, which implies $A_M^* = A_M$ and hence $A_N^* = A_N$ for all $N \geq M$. One then has $\|A_N\| = \sup\{|\rho(A_N)|, \rho \in \mathcal{S}(\mathcal{A}_1^N)\}$. Because of the special form of A_N one may replace the supremum over the set $\mathcal{S}(\mathcal{A}_1^N)$ of all states on $\mathcal{A}_1^N$ by the supremum over the set $\mathcal{S}^p(\mathcal{A}_1^N)$ of all permutation invariant states, which in turn may be replaced by the supremum over the extreme boundary $\partial\mathcal{S}^p(\mathcal{A}_1^N)$ of $\mathcal{S}^p(\mathcal{A}_1^N)$. It is well known (Størmer, 1969; see also Subsection 6.2) that the latter consists of all states of the form $\rho = \omega^N$, so that $\|A_N\| = \sup\{|\omega^N(A_N)|, \omega \in \mathcal{S}(\mathcal{A}_1)\}$. This is actually equal to $\|A_M\| = \sup\{|\omega^M(A_M)|\}$. Now the norm in $\mathcal{A}_0^{(c)}$ is $\|A_0\| = \sup\{|A_0(\omega)|, \omega \in \mathcal{S}(\mathcal{A}_1)\}$, and by definition of A_0 one has $A_0(\omega) = \omega^M(A_M)$. Hence (6.12) follows.

6.2 Quasilocal observables

In the C^*-algebraic approach to quantum theory, infinite systems are usually described by means of inductive limit C^*-algebras and the associated quasilocal observables (Thirring, 1983; Strocchi, 1985; Bratteli & Robinson, 1981, 1987; Haag, 1992; Araki, 1999; Sewell, 1986, 2002). To arrive at these notions in the case at hand, we proceed as follows (Duffield & Werner, 1992c).

A sequence $A = (A_1, A_2, \cdots)$ (where $A_N \in \mathcal{A}_1^N$, as before) is called *local* when for some fixed M and all $N \geq M$ one has $A_N = A_M \otimes 1 \otimes \cdots \otimes 1$ (where one has $N - M$ copies of the unit $1 \in \mathcal{A}_1$); cf. (6.4). A sequence is said to be *quasilocal* when for any $\varepsilon > 0$ there is an N_ε and a local sequence A' such that $\|A_N - A'_N\| < \varepsilon$ for all $N \geq N_\varepsilon$. On this basis, we define the *inductive limit* C^*-algebra

$$\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N} \tag{6.13}$$

of the family of C^*-algebras $(\mathcal{A}_1^N)$ with respect to the inclusion maps $\mathcal{A}_1^N \hookrightarrow \mathcal{A}_1^{N+1}$ given by $A_N \mapsto A_N \otimes 1$. As a set, (6.13) consists of all equivalence classes $[A] \equiv A_0$ of quasilocal sequences A under the equivalence relation $A \sim B$ when $\lim_{N\to\infty} \|A_N - B_N\| = 0$. The norm on $\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$ is

$$\|A_0\| = \lim_{N\to\infty} \|A_N\|, \tag{6.14}$$

and the rest of the C^*-algebraic structure is inherited from the quasilocal sequences in the obvious way (e.g., $A_0^* = [A^*]$ with $A^* = (A_1^*, A_2^*, \cdots)$, etc.). As the notation suggests, each $\mathcal{A}_1^N$ is contained in $\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$ as a C^*-subalgebra by identifying $A_N \in \mathcal{A}_1^N$ with the local (and hence quasilocal) sequence $A = (0, \cdots, 0, A_N \otimes 1, A_N \otimes 1 \otimes 1, \cdots)$, and forming its equivalence class A_0 in $\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$ as just explained.[265] The assumption underlying the common idea that (6.13) is "the" algebra of observables of the infinite system under study is that by locality or some other human limitation the infinite tail of the system is not accessible, so that the observables must be arbitrarily close (i.e. in norm) to operators of the form $A_N \otimes 1 \otimes 1, \cdots$ for some *finite* N.

This leads us to a second continuous field of C^*-algebras $\mathcal{A}^{(q)}$ over $0 \cup 1/\mathbb{N}$, with fibers

$$\begin{aligned} \mathcal{A}_0^{(q)} &= \overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}; \\ \mathcal{A}_{1/N}^{(q)} &= \mathcal{A}_1^N. \end{aligned} \tag{6.15}$$

Thus the suffix q reminds one of that fact that the limit algebra $\mathcal{A}_0^{(q)}$ consists of *q*uasilocal or *q*uantum-mechanical observables. We equip the collection of C^*-algebras (6.15) with the structure of a continuous field of C^*-algebras $\mathcal{A}^{(q)}$ over $0 \cup 1/\mathbb{N}$ by declaring that the continuous sections are of the form $(A_0, A_1, A_2, \cdots)$ where $(A_1, A_2, \cdots)$ is quasilocal and A_0 is defined by this quasilocal sequence as just explained.[266] For $N < \infty$ this field has the same fibers

$$\mathcal{A}_{1/N}^{(q)} = \mathcal{A}_{1/N}^{(c)} = \mathcal{A}_1^N \tag{6.16}$$

as the continuous field $\mathcal{A}$ of the previous subsection, but the fiber $\mathcal{A}_0^{(q)}$ is completely different from $\mathcal{A}_0^{(c)}$. In particular, if $\mathcal{A}_1$ is noncommutative then so is $\mathcal{A}_0^{(q)}$, for it contains all $\mathcal{A}_1^N$.

The relationship between the continuous fields of C^*-algebras $\mathcal{A}^{(q)}$ and $\mathcal{A}^{(c)}$ may be studied in two different (but related) ways. First, we may construct concrete representations of all C^*-algebras $\mathcal{A}_1^N$, $N < \infty$, as well as of $\mathcal{A}_0^{(c)}$ and $\mathcal{A}_0^{(q)}$ on a single Hilbert space; this approach leads to superselections rules in the traditional sense. This method will be taken up in the next subsection. Second, we may look at those families of states $(\omega_1, \omega_{1/2}, \cdots, \omega_{1/N}, \cdots)$ (where $\omega_{1/N}$ is a state on $\mathcal{A}_1^N$) that admit limit states $\omega_0^{(c)}$ *and* $\omega_0^{(q)}$ on $\mathcal{A}_0^{(c)}$ and $\mathcal{A}_0^{(q)}$, respectively, such that the ensuing families of states $(\omega_0^{(c)}, \omega_1, \omega_{1/2}, \cdots)$ and $(\omega_0^{(q)}, \omega_1, \omega_{1/2}, \cdots)$ are *continuous* fields of states on $\mathcal{A}^{(c)}$ and on $\mathcal{A}^{(q)}$, respectively (cf. the end of Subsection 5.1).

Now, any state $\omega_0^{(q)}$ on $\mathcal{A}_0^{(q)}$ defines a state $\omega_{0|1/N}^{(q)}$ on $\mathcal{A}_1^N$ by restriction, and the ensuing field of states on $\mathcal{A}^{(q)}$ is clearly continuous. Conversely, any continuous field $(\omega_0^{(q)}, \omega_1, \omega_{1/2}, \ldots, \omega_{1/N}, \ldots)$ of states on

[265] Of course, the entries $A_1, \cdots A_{N-1}$, which have been put to zero, are arbitrary.

[266] The fact that this defines a continuous field follows from (6.14) and Prop. II.1.2.3 in Landsman (1998); cf. footnote 263.

$\mathcal{A}^{(q)}$ becomes arbitrarily close to a field of the above type for N large.[267] However, the restrictions $\omega^{(q)}_{0|1/N}$ of a given state $\omega^{(q)}_0$ on $\mathcal{A}^{(q)}_0$ to $\mathcal{A}^N_1$ may not converge to a state $\omega^{(c)}_0$ on $\mathcal{A}^{(c)}_0$ for $N \to \infty$.[268]. States $\omega^{(q)}_0$ on $\overline{\cup_{N\in\mathbb{N}}\mathcal{A}^N_1}$ that do have this property will here be called *classical*. In other words, $\omega^{(q)}_{0|1/N}$ is classical when there exists a probability measure μ_0 on $\mathcal{S}(\mathcal{A}_1)$ such that

$$\lim_{N\to\infty} \int_{\mathcal{S}(\mathcal{A}_1)} d\mu_0(\rho)\, (\rho^N(A_N) - \omega^{(q)}_{0|1/N}(A_N)) = 0 \tag{6.17}$$

for each (approximately) symmetric sequence $(A_1, A_2, \ldots)$. To analyze this notion we need a brief intermezzo on general C^*-algebras and their representations.

- A *folium* in the state space $\mathcal{S}(\mathcal{B})$ of a C^*-algebra $\mathcal{B}$ is a convex, norm-closed subspace $\mathcal{F}$ of $\mathcal{S}(\mathcal{B})$ with the property that if $\omega \in \mathcal{F}$ and $B \in \mathcal{B}$ such that $\omega(B^*B) > 0$, then the "reduced" state $\omega_B : A \mapsto \omega(B^*AB)/\omega(B^*B)$ must be in $\mathcal{F}$ (Haag, Kadison, & Kastler, 1970).[269] For example, if π is a representation of $\mathcal{B}$ on a Hilbert space $\mathcal{H}$, then the set of all density matrices on $\mathcal{H}$ (i.e. the π-normal states on $\mathcal{B}$)[270] comprises a folium $\mathcal{F}_\pi$. In particular, each state ω on $\mathcal{B}$ defines a folium $\mathcal{F}_\omega \equiv \mathcal{F}_{\pi_\omega}$ through its GNS-representation π_ω.

- Two representations π and π' are called *disjoint*, written $\pi \perp \pi'$, if no subrepresentation of π is (unitarily) equivalent to a subrepresentation of π' and vice versa. They are said to be *quasi-equivalent*, written $\pi \sim \pi'$, when π has no subrepresentation disjoint from π', and vice versa.[271] Quasi-equivalence is an equivalence relation $\sim$ on the set of representations. See Kadison & Ringrose (1986), Ch. 10.

- Similarly, two states ρ, σ are called either quasi-equivalent ($\rho \sim \sigma$) or disjoint ($\rho \perp \sigma$) when the corresponding GNS-representations have these properties.

- A state ω is called *primary* when the corresponding von Neumann algebra $\pi_\omega(\mathcal{B})''$ is a factor.[272] Equivalently, ω is primary iff each subrepresentation of $\pi_\omega(\mathcal{B})$ is quasi-equivalent to $\pi_\omega(\mathcal{B})$, which is the case iff $\pi_\omega(\mathcal{B})$ admits no (nontrivial) decomposition as the direct sum of two disjoint subrepresentations.

Now, there is a bijective correspondence between folia in $\mathcal{S}(\mathcal{B})$ and quasi-equivalence classes of representations of $\mathcal{B}$, in that $\mathcal{F}_\pi = \mathcal{F}_{\pi'}$ iff $\pi \sim \pi'$. Furthermore (as one sees from the GNS-construction), any folium $\mathcal{F} \subset \mathcal{S}(\mathcal{B})$ is of the form $\mathcal{F} = \mathcal{F}_\pi$ for some representation $\pi(\mathcal{B})$. Note that if π is injective (i.e. faithful), then the corresponding folium is dense in $\mathcal{S}(\mathcal{B})$ in the weak*-topology by Fell's Theorem. So in case that $\mathcal{B}$ is simple,[273] any folium is weak*-dense in the state space.

Two states need not be either disjoint or quasi-equivalent. This dichotomy does apply, however, within the class of primary states. Hence *two primary states are either disjoint or quasi-equivalent.* If ω is primary, then each state in the folium of π_ω is primary as well, and is quasi-equivalent to ω. If, on the other hand, ρ and σ are primary and disjoint, then $\mathcal{F}_\rho \cap \mathcal{F}_\sigma = \emptyset$. Pure states are, of course, primary.[274] Furthermore, in thermodynamics pure phases are described by primary KMS states (Emch & Knops, 1970; Bratteli & Robinson, 1981; Haag, 1992; Sewell, 2002). This apparent relationship between primary states and "purity" of some sort is confirmed by our description of macroscopic observables:[275]

[267] For any fixed quasilocal sequence $(A_1, A_2, \cdots)$ and $\varepsilon > 0$, there is an N_ε such that $|\omega_{1/N}(A_N) - \omega^{(q)}_{0|1/N}(A_N)| < \varepsilon$ for all $N > N_\varepsilon$.

[268] See footnote 288 below for an example

[269] See also Haag (1992). The name 'folium' is very badly chosen, since $\mathcal{S}(\mathcal{B})$ is by no means foliated by its folia; for example, a folium may contain subfolia.

[270] A state ω on $\mathcal{B}$ is called π-normal when it is of the form $\omega(B) = \mathrm{Tr}\,\rho\pi(B)$ for some density matrix ρ. Hence the π-normal states are the normal states on the von Neumann algebra $\pi(\mathcal{B})''$.

[271] Equivalently, two representations π and π' are disjoint iff no π-normal state is π'-normal and vice versa, and quasi-equivalent iff each π-normal state is π'-normal and vice versa.

[272] A von Neumann algebra $\mathcal{M}$ acting on a Hilbert space is called a *factor* when its center $\mathcal{M} \cap \mathcal{M}'$ is trivial, i.e. consists of multiples of the identity.

[273] In the sense that it has no *closed* two-sided ideals. For example, the matrix algebra $M_n(\mathbb{C})$ is simple for any n, as is its infinite-dimensional analogue, the C^*-algebra of all compact operators on a Hilbert space. The C^*-algebra of quasilocal observables of an infinite quantum systems is typically simple as well.

[274] Since the corresponding GNS-representation π_ω is irreducible, $\pi_\omega(\mathcal{B})'' = \mathcal{B}(\mathcal{H}_\omega)$ is a factor.

[275] These claims easily follow from Sewell (2002), §2.6.5, which in turn relies on Hepp (1972).

- *If $\omega_0^{(q)}$ is a classical primary state on $\mathcal{A}_0^{(q)} = \overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$, then the corresponding limit state $\omega_0^{(c)}$ on $\mathcal{A}_0^{(c)} = C(\mathcal{S}(\mathcal{A}_1))$ is pure (and hence given by a point in $\mathcal{S}(\mathcal{A}_1)$).*
- *If $\rho_0^{(q)}$ and $\sigma_0^{(q)}$ are classical primary states on $\mathcal{A}_0^{(q)}$, then*

$$\rho_0^{(c)} = \sigma_0^{(c)} \quad \Leftrightarrow \quad \rho_0^{(q)} \sim \sigma_0^{(q)}; \tag{6.18}$$

$$\rho_0^{(c)} \neq \sigma_0^{(c)} \quad \Leftrightarrow \quad \rho_0^{(q)} \perp \sigma_0^{(q)}. \tag{6.19}$$

As in (6.17), a general classical state $\omega_0^{(q)}$ with limit state $\omega_0^{(c)}$ on $C(\mathcal{S}(\mathcal{A}_1))$ defines a probability measure μ_0 on $\mathcal{S}(\mathcal{A}_1)$ by

$$\omega_0^{(c)}(f) = \int_{\mathcal{S}(\mathcal{A}_1)} d\mu_0\, f, \tag{6.20}$$

which describes the probability distribution of the macroscopic observables in that state. As we have seen, this distribution is a delta function for primary states. In any case, it is insensitive to the microscopic details of $\omega_0^{(q)}$ in the sense that local modifications of $\omega_0^{(q)}$ do not affect the limit state $\omega_0^{(c)}$ (Sewell, 2002). Namely, it easily follows from (6.8) and the fact that the GNS-representation is cyclic that one can strengthen the second claim above:

> *Each state in the folium $\mathcal{F}_{\omega_0^{(q)}}$ of a classical state $\omega_0^{(q)}$ is automatically classical and has the same limit state on $\mathcal{A}_0^{(c)}$ as $\omega_0^{(q)}$.*

To make this discussion a bit more concrete, we now identify an important class of classical states on $\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$. We say that a state ω on this C^*-algebra is *permutation-invariant* when each of its restrictions to $\mathcal{A}_1^N$ is invariant under the natural action of the symmetric group $\mathfrak{S}_N$ on $\mathcal{A}_1^N$ (i.e. $\sigma \in \mathfrak{S}_N$ maps an elementary tensor $A_N = B_1 \otimes \cdots \otimes B_N \in \mathcal{A}_1^N$ to $B_{\sigma(1)} \otimes \cdots \otimes B_{\sigma(N)}$, cf. (6.5)). The structure of the set $\mathcal{S}^{\mathfrak{S}}$ of all permutation-invariant states in $\mathcal{S}(\mathcal{A}_0^{(q)})$ has been analyzed by Størmer (1969). Like any compact convex set, it is the (weak*-closed) convex hull of its extreme boundary $\partial_e\mathcal{S}^{\mathfrak{S}}$. The latter consists of all infinite product states $\omega = \rho^\infty$, where $\rho \in \mathcal{S}(\mathcal{A}_1)$. I.e. if $A_0 \in \mathcal{A}_0^{(q)}$ is an equivalence class $[A_1, A_2, \cdots]$, then

$$\rho^\infty(A_0) = \lim_{N\to\infty} \rho^N(A_N); \tag{6.21}$$

cf. (6.11). Equivalently, the restriction of ω to any $\mathcal{A}_1^N \subset \mathcal{A}_0^{(q)}$ is given by $\otimes^N\rho$. Hence $\partial_e\mathcal{S}^{\mathfrak{S}}$ is isomorphic (as a compact convex set) to $\mathcal{S}(\mathcal{A}_1)$ in the obvious way, and the primary states in $\mathcal{S}^{\mathfrak{S}}$ are precisely the elements of $\partial_e\mathcal{S}^{\mathfrak{S}}$.

A general state $\omega_0^{(q)}$ in $\mathcal{S}^{\mathfrak{S}}$ has a unique decomposition[276]

$$\omega_0^{(q)}(A_0) = \int_{\mathcal{S}(\mathcal{A}_1)} d\mu(\rho)\, \rho^\infty(A_0), \tag{6.22}$$

where μ is a probability measure on $\mathcal{S}(\mathcal{A}_1)$ and $A_0 \in \mathcal{A}_0^{(q)}$.[277] The following beautiful illustration of the abstract theory (Unnerstall, 1990a,b) is then clear from (6.17) and (6.22):

> *If $\omega_0^{(q)}$ is permutation-invariant, then it is classical. The associated limit state $\omega_0^{(c)}$ on $\mathcal{A}_0^{(c)}$ is characterized by the fact that the measure μ_0 in* (6.20) *coincides with the measure μ in* (6.22).[278]

[276] This follows because $\mathcal{S}^{\mathfrak{S}}$ is a so-called Bauer simplex (Alfsen, 1970). This is a compact convex set K whose extreme boundary $\partial_e K$ is closed and for which every $\omega \in K$ has a *unique* decomposition as a probability measure supported by $\partial_e K$, in the sense that $a(\omega) = \int_{\partial_e K} d\mu(\rho)\, a(\rho)$ for any continuous affine function a on K. For a unital C^*-algebra $\mathcal{A}$ the continuous affine functions on the state space $K = \mathcal{S}(\mathcal{A})$ are precisely the elements A of $\mathcal{A}$, reinterpreted as functions $\hat{A}$ on $\mathcal{S}(\mathcal{A})$ by $\hat{A}(\omega) = \omega(A)$. For example, the state space $\mathcal{S}(\mathcal{A})$ of a commutative unital C^*-algebra $\mathcal{A}$ is a Bauer simplex, which consists of all (regular Borel) probability measures on the pre state space $\mathcal{P}(\mathcal{A})$.

[277] This is a quantum analogue of De Finetti's representation theorem in classical probability theory (Heath & Sudderth, 1976; van Fraassen, 1991); see also Hudson & Moody (1975/76) and Caves et al. (2002).

[278] In fact, each state in the folium $\mathcal{F}^{\mathfrak{S}}$ in $\mathcal{S}(\mathcal{A}_0^{(q)})$ corresponding to the (quasi-equivalence class of) the representation $\oplus_{[\omega\in\mathcal{S}^{\mathfrak{S}}]}\pi_\omega$ is classical.

6.3 Superselection rules

Infinite quantum systems are often associated with the notion of a superselection rule (or sector), which was originally introduced by Wick, Wightman, & Wigner (1952) in the setting of standard quantum mechanics on a Hilbert space $\mathcal{H}$. The basic idea may be illustrated in the example of the boson/fermion (or "univalence") superselection rule.[279] Here one has a *projective* unitary representation $\mathcal{D}$ of the rotation group $SO(3)$ on $\mathcal{H}$, for which $\mathcal{D}(R_{2\pi}) = \pm 1$ for any rotation $R_{2\pi}$ of 2π around some axis. Specifically, on bosonic states Ψ_B one has $\mathcal{D}(R_{2\pi})\Psi_B = \Psi_B$, whereas on fermionic states Ψ_F the rule is $\mathcal{D}(R_{2\pi})\Psi_F = -\Psi_F$. Now the argument is that a rotation of 2π accomplishes nothing, so that it cannot change the physical state of the system. This requirement evidently holds on the subspace $\mathcal{H}_B \subset \mathcal{H}$ of bosonic states in $\mathcal{H}$, but it is equally well satisfied on the subspace $\mathcal{H}_F \subset \mathcal{H}$ of fermionic states, since Ψ and $z\Psi$ with $|z| = 1$ describe the same physical state. However, if $\Psi = c_B\Psi_B + c_F\Psi_F$ (with $|c_B|^2 + |c_F|^2 = 1$), then $\mathcal{D}(R_{2\pi})\Psi = c_B\Psi_B - c_F\Psi_F$, which is not proportional to Ψ and apparently describes a genuinely different physical state from Ψ.

The way out is to deny this conclusion by declaring that $\mathcal{D}(R_{2\pi})\Psi$ and Ψ *do* describe the same physical state, and this is achieved by postulating that no physical *observables* A (in their usual mathematical guise as operators on $\mathcal{H}$) exist for which $(\Psi_B, A\Psi_F) \neq 0$. For in that case one has

$$(c_B\Psi_B \pm c_F\Psi_F, A(c_B\Psi_B \pm c_F\Psi_F)) = |c_B|^2(\Psi_B, A\Psi_B) + |c_F|^2(\Psi_F, A\Psi_F) \tag{6.23}$$

for any *observable* A, so that $(\mathcal{D}(R_{2\pi})\Psi, A\mathcal{D}(R_{2\pi})\Psi) = (\Psi, A\Psi)$ for any $\Psi \in \mathcal{H}$. Since any quantum-mechanical prediction ultimately rests on expectation values $(\Psi, A\Psi)$ for physical observables A, the conclusion is that a rotation of 2π indeed does nothing to the system. This is codified by saying that superpositions of the type $c_B\Psi_B + c_F\Psi_F$ are *incoherent* (whereas superpositions $c_1\Psi_1 + c_2\Psi_2$ with Ψ_1, Ψ_2 both in either $\mathcal{H}_B$ or in $\mathcal{H}_F$ are *coherent*). Each of the subspaces $\mathcal{H}_B$ and $\mathcal{H}_F$ of $\mathcal{H}$ is said to be a *superselection sector*, and the statement that $(\Psi_B, A\Psi_F) = 0$ for any observbale A and $\Psi_B \in \mathcal{H}_B$ and $\Psi_F \in \mathcal{H}_F$ is called a *superselection rule*.[280]

The price one pays for this solution is that states of the form $c_B\Psi_B + c_F\Psi_F$ with $c_B \neq 0$ and $c_F \neq 0$ are mixed, as one sees from (6.23). More generally, if $\mathcal{H} = \oplus_{\lambda\in\Lambda}\mathcal{H}_\lambda$ with $(\Psi, A\Phi) = 0$ whenever A is an observable, $\Psi \in \mathcal{H}_\lambda$, $\Phi \in \mathcal{H}_{\lambda'}$, and $\lambda \neq \lambda'$, and if in addition for each λ and each pair $\Psi, \Phi \in \mathcal{H}_\lambda$ there exists an observable A for which $(\Psi, A\Phi) \neq 0$, then the subspaces $\mathcal{H}_\lambda$ are called superselection sectors in $\mathcal{H}$. Again a key consequence of the occurrence of superselection sectors is that unit vectors of the type $\Psi = \sum_\lambda c_\lambda\Psi_\lambda$ with $\Psi \in \mathcal{H}_\lambda$ (and $c_\lambda \neq 0$ for at least two λ's) define mixed states

$$\psi(A) = (\Psi, A\Psi) = \sum_\lambda |c_\lambda|^2(\Psi_\lambda, A\Psi_\lambda) = \sum_\lambda |c_\lambda|^2\psi_\lambda(A).$$

This procedure is rather ad hoc. A much deeper approach to superselection theory was developed by Haag and collaborators; see Roberts & Roepstorff (1969) for an introduction. Here the starting point is the abstract C^*-algebra of observables $\mathcal{A}$ of a given quantum system, and superselection sectors are reinterpreted as equivalence classes (under unitary isomorphism) of irreducible representations of $\mathcal{A}$ (satisfying a certain selection criterion - see below). The connection between the concrete Hilbert space approach to superselection sectors discussed above and the abstract C^*-algebraic approach is given by the following lemma (Hepp, 1972):[281]

> *Two pure states ρ, σ on a C^*-algebra $\mathcal{A}$ define different sectors iff for each representation $\pi(\mathcal{A})$ on a Hilbert space $\mathcal{H}$ containing unit vectors Ψ_ρ, Ψ_σ such that $\rho(A) = (\Psi_\rho, \pi(A)\Psi_\rho)$ and $\sigma(A) = (\Psi_\sigma, \pi(A)\Psi_\sigma)$ for all $A \in \mathcal{A}$, one has $(\Psi_\rho, \pi(A)\Psi_\sigma) = 0$ for all $A \in \mathcal{A}$.*

In practice, however, most irreducible representations of a typical C^*-algebra $\mathcal{A}$ used in physics are physically irrelevant mathematical artefacts. Such representations may be excluded from consideration by some *selection criterion*. What this means depends on the context. For example, in quantum field theory this notion is made precise in the so-called DHR theory (reviewed by Roberts (1990), Haag (1992), Araki (1999), and Halvorson (2005)). In the class of theories discussed in the preceding

[279] See also Giulini (2003) for a modern mathematical treatment.

[280] In an ordinary selection rule between Ψ and Φ one merely has $(\Psi, H\Phi) = 0$ for the Hamiltonian H.

[281] Hepp proved a more general version of this lemma, in which 'Two pure states ρ, σ on a C^*-algebra $\mathcal{B}$ define different sectors iff...' is replaced by 'Two states ρ, σ on a C^*-algebra $\mathcal{B}$ are disjoint iff...'

two subsections, we take the algebra of observables $\mathcal{A}$ to be $\mathcal{A}_0^{(q)}$ - essentially for reasons of human limitation - and for pedagogical reasons define (equivalence classes of) irreducible representations of $\mathcal{A}_0^{(q)}$ as superselection sectors, henceforth often just called *sectors*, only when they are equivalent to the GNS-representation given by a permutation-invariant pure state on $\mathcal{A}_0^{(q)}$. In particular, such a state is classical. On this selection criterion, the results in the preceding subsection trivially imply that there is a bijective correspondence between pure states on $\mathcal{A}_1$ and sectors of $\mathcal{A}_0^{(q)}$. The sectors of the commutative C^*-algebra $\mathcal{A}_0^{(c)}$ are just the points of $\mathcal{S}(\mathcal{A}_1)$; note that a *mixed* state on $\mathcal{A}_1$ defines a *pure* state on $\mathcal{A}_0^{(c)}$! The role of the sectors of $\mathcal{A}_1$ in connection with those of $\mathcal{A}_0^{(c)}$ will be clarified in Subsection 6.5.

Whatever the model or the selection criterion, it is enlightening (and to some extent even in accordance with experimental practice) to consider superselection sectors entirely from the perspective of the pure states on the algebra of observables $\mathcal{A}$, removing $\mathcal{A}$ itself and its representations from the scene. To do so, we equip the space $\mathcal{P}(\mathcal{A})$ of pure states on $\mathcal{A}$ with the structure of a transition probability space (von Neumann, 1981; Mielnik, 1968).[282] A *transition probability* on a set $\mathcal{P}$ is a function

$$p : \mathcal{P} \times \mathcal{P} \to [0, 1] \tag{6.24}$$

that satisfies

$$p(\rho, \sigma) = 1 \iff \rho = \sigma \tag{6.25}$$

and

$$p(\rho, \sigma) = 0 \iff p(\sigma, \rho) = 0. \tag{6.26}$$

A set with such a transition probability is called a *transition probability space*. Now, the pure state space $\mathcal{P}(\mathcal{A})$ of a C^*-algebra $\mathcal{A}$ carries precisely this structure if we define[283]

$$p(\rho, \sigma) := \inf\{\rho(A) \mid A \in \mathcal{A}, 0 \leq A \leq 1, \sigma(A) = 1\}. \tag{6.27}$$

To give a more palatable formula, note that since pure states are primary, two pure states ρ, σ are either disjoint ($\rho \perp \sigma$) or else (quasi, hence unitarily) equivalent ($\rho \sim \sigma$). In the first case, (6.27) yields

$$p(\rho, \sigma) = 0 \quad (\rho \perp \sigma). \tag{6.28}$$

Ine the second case it follows from Kadison's transitivity theorem (cf. Thm. 10.2.6 in Kadison & Ringrose (1986)) that the Hilbert space $\mathcal{H}_\rho$ from the GNS-representation $\pi_\rho(\mathcal{A})$ defined by ρ contains a unit vector Ω_σ (unique up to a phase) such that

$$\sigma(A) = (\Omega_\sigma, \pi_\rho(A)\Omega_\sigma). \tag{6.29}$$

Eq. (6.27) then leads to the well-known expression

$$p(\rho, \sigma) = |(\Omega_\rho, \Omega_\sigma)|^2 \quad (\rho \sim \sigma). \tag{6.30}$$

In particular, if $\mathcal{A}$ is commutative, then

$$p(\rho, \sigma) = \delta_{\rho\sigma}. \tag{6.31}$$

For $\mathcal{A} = M_2(\mathbb{C})$ one obtains

$$p(\rho, \sigma) = \tfrac{1}{2}(1 + \cos\theta_{\rho\sigma}), \tag{6.32}$$

where $\theta_{\rho\sigma}$ is the angular distance between ρ and σ (seen as points on the two-sphere $S^2 = \partial_e B^3$, cf. (6.3) etc.), measured along a great circle.

Superselection sectors may now be defined for any transition probability spaces $\mathcal{P}$. A family of subsets of $\mathcal{P}$ is called *orthogonal* if $p(\rho, \sigma) = 0$ whenever ρ and σ do not lie in the same subset. The space $\mathcal{P}$ is called *reducible* if it is the union of two (nonempty) orthogonal subsets; if not, it is said to be *irreducible*. A *component* of $\mathcal{P}$ is a subset $\mathcal{C} \subset \mathcal{P}$ such that $\mathcal{C}$ and $\mathcal{P}\backslash\mathcal{C}$ are orthogonal. An irreducible component of $\mathcal{P}$ is called a *(superselection) sector*. Thus $\mathcal{P}$ is the disjoint union of its sectors. For $\mathcal{P} = \mathcal{P}(\mathcal{A})$ this reproduces the algebraic definition of a superselection sector (modulo the selection criterion) via the correspondence between states and representations given by the GNS-constructions. For example, in the commutative case $\mathcal{A} \cong C(X)$ each point in $X \cong \mathcal{P}(\mathcal{A})$ is its own little sector.

[282] See also Beltrametti & Cassinelli (1984) or Landsman (1998) for concise reviews.

[283] This definition applies to the case that $\mathcal{A}$ is unital; see Landsman (1998) for the general case. An analogous formula defines a transition probability on the extreme boundary of any compact convex set.

6.4 A simple example: the infinite spin chain

Let us illustrate the occurrence of superselection sectors in a simple example, where the algebra of observables is $\mathcal{A}_0^{(q)}$ with $\mathcal{A}_1 = M_2(\mathbb{C})$. Let $\mathcal{H}_1 = \mathbb{C}^2$, so that $\mathcal{H}_1^N = \otimes^N \mathbb{C}^2$ is the tensor product of N copies of $\mathbb{C}^2$. It is clear that $\mathcal{A}_1^N$ acts on $\mathcal{H}_1^N$ in a natural way (i.e. componentwise). This defines an irreducible representation π_N of $\mathcal{A}_1^N$, which is indeed its unique irreducible representation (up to unitary equivalence). In particular, for $N < \infty$ the quantum system whose algebra of observables is $\mathcal{A}_1^N$ (such as a chain with N two-level systems) has no superselection rules. We define the $N \to \infty$ limit "$(M_2(\mathbb{C}))^\infty$" of the C^*-algebras $(M_2(\mathbb{C}))^N$ as the inductive limit $\mathcal{A}_0^{(q)}$ for $\mathcal{A}_1 = M_2(\mathbb{C})$, as introduced in Subsection 6.2; see (6.13). The definition of "$\otimes^\infty \mathbb{C}^2$" is slightly more involved, as follows (von Neumann, 1938).

For any Hilbert space $\mathcal{H}_1$, let Ψ be a sequence $(\Psi_1, \Psi_2, \ldots)$ with $\Psi_n \in \mathcal{H}_1$. The space H_1 of such sequences is a vector space in the obvious way. Now let Ψ and Φ be two such sequences, and write $(\Psi_n, \Phi_n) = \exp(i\alpha_n)|(\Psi_n, \Phi_n)|$. If $\sum_n |\alpha_n| = \infty$, we define the (pre-) inner product (Ψ, Φ) to be zero. If $\sum_n |\alpha_n| < \infty$, we put $(\Psi, \Phi) = \prod_n (\Psi_n, \Phi_n)$ (which, of course, may still be zero!). The (vector space) quotient of H_1 by the space of sequences Ψ for which $(\Psi, \Psi) = 0$ can be completed to a Hilbert space $\mathcal{H}_1^\infty$ in the induced inner product, called the *complete* infinite tensor product of the Hilbert space $\mathcal{H}_1$ (over the index set $\mathbb{N}$).[284] We apply this construction with $\mathcal{H}_1 = \mathbb{C}^2$. If (e_i) is some basis of $\mathbb{C}^2$, an orthonormal basis of $\mathcal{H}_1^\infty$ then consists of all different infinite strings $e_{i_1} \otimes \cdots e_{i_n} \otimes \cdots$, where e_{i_n} is e_i regarded as a vector in $\mathbb{C}^2$.[285] We denote the multi-index $(i_1, \ldots, i_n, \ldots)$ simply by I, and the corresponding basis vector by e_I.

This Hilbert space $\mathcal{H}_1^\infty$ carries a natural faithful representation π of $\mathcal{A}_0^{(q)}$: if $A_0 \in \mathcal{A}_0^{(q)}$ is an equivalence class $[A_1, A_2, \cdots]$, then $\pi(A_0)e_I = \lim_{N\to\infty} A_N e_i$, where A_N acts on the first N components of e_I and leaves the remainder unchanged.[286] Now the point is that although each $\mathcal{A}_1^N$ acts irreducibly on $\mathcal{H}_1^N$, the representation $\pi(\mathcal{A}_0^{(q)})$ on $\mathcal{H}_1^\infty$ thus constructed is highly reducible. The reason for this is that by definition (quasi-) local elements of $\mathcal{A}_0^{(q)}$ leave the infinite tail of a vector in $\mathcal{H}_1^\infty$ (almost) unaffected, so that vectors with different tails lie in different superselection sectors. Without the quasi-locality condition on the elements of $\mathcal{A}_0^{(q)}$, no superselection rules would arise. For example, in terms of the usual basis

$$\left\{ \uparrow = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \downarrow = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \tag{6.33}$$

of $\mathbb{C}^2$, the vectors $\Psi_\uparrow = \uparrow \otimes \uparrow \cdots \uparrow \cdots$ (i.e. an infinite product of 'up' vectors) and $\Psi_\downarrow = \downarrow \otimes \downarrow \cdots \downarrow \cdots$ (i.e. an infinite product of 'down' vectors) lie in different sectors. The reason why the inner product $(\Psi_\uparrow, \pi(A)\Psi_\downarrow)$ vanishes for any $A \in \mathcal{A}_0^{(q)}$ is that for local observables A one has $\pi(A) = A_M \otimes 1 \otimes \cdots 1 \cdots$ for some $A_M \in \mathcal{B}(\mathcal{H}_M)$; the inner product in question therefore involves infinitely many factors $(\uparrow, 1 \downarrow) = (\uparrow, \downarrow) = 0$. For quasilocal A the operator $\pi(A)$ might have a small nontrivial tail, but the inner product vanishes nonetheless by an approximation argument.

More generally, elementary analysis shows that $(\Psi_u, \pi(A)\Psi_v) = 0$ whenever $\Psi_u = \otimes^\infty u$ and $\Psi_v = \otimes^\infty v$ for unit vectors $u, v \in \mathbb{C}^2$ with $u \neq v$. The corresponding vector states ψ_u and ψ_v on $\mathcal{A}_0^{(q)}$ (i.e. $\psi_u(A) = (\Psi_u, \pi(A)\Psi_u)$ etc.) are obviously permutation-invariant and hence classical. Identifying $\mathcal{S}(M_2(\mathbb{C}))$ with B^3, as in (6.3), the corresponding limit state $(\psi_u)_0$ on $\mathcal{A}_0^{(c)}$ defined by ψ_u is given by (evaluation at) the point $\tilde{u} = (x, y, z)$ of $\partial_e B^3 = S^2$ (i.e. the two-sphere) for which the corresponding density matrix $\rho(\tilde{u})$ is the projection operator onto u. It follows that ψ_u and ψ_v are disjoint; cf. (6.19). We conclude that each unit vector $u \in \mathbb{C}^2$ determines a superselection sector π_u, namely the GNS-representation of the corresponding state ψ_u, and that each such sector is realized as a subspace $\mathcal{H}_u$ of $\mathcal{H}_1^\infty$ (viz. $\mathcal{H}_u = \overline{\pi(\mathcal{A}_0^{(q)})\Psi_u}$). Moreover, since a permutation-invariant state on $\mathcal{A}_0^{(q)}$ is pure iff it is of the form ψ_u, we have found all superselection sectors of our system. Thus in what follows we may

[284] Each fixed $\Psi \in \mathcal{H}_1$ defines an *incomplete* tensor product $\mathcal{H}_\Psi^\infty$, defined as the closed subspace of $\mathcal{H}_1^\infty$ consisting of all Φ for which $\sum_n |(\Psi_n, \Phi_n) - 1| < \infty$. If $\mathcal{H}_1$ is separable, then so is $\mathcal{H}_\Psi^\infty$ (in contrast to $\mathcal{H}_1^\infty$, which is an uncountable direct sum of the $\mathcal{H}_\Psi^\infty$).

[285] The cardinality of the set of all such strings equals that of $\mathbb{R}$, so that $\mathcal{H}_1^\infty$ is non-separable, as claimed.

[286] Indeed, this yields an alternative way of defining $\overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N}$ as the norm closure of the union of all $\mathcal{A}_1^N$ acting on $\mathcal{H}_1^\infty$ in the stated way.

concentrate our attention on the subspace (of $\mathcal{H}_1^\infty$) and subrepresentation (of π)

$$\begin{aligned} \mathcal{H}_{\mathfrak{S}} &= \oplus_{\tilde{u}\in S^2}\mathcal{H}_u; \\ \pi_{\mathfrak{S}}(\mathcal{A}_0^{(q)}) &= \oplus_{\tilde{u}\in S^2}\pi_u(\mathcal{A}_0^{(q)}), \end{aligned} \tag{6.34}$$

where π_u is simply the restriction of π to $\mathcal{H}_u \subset \mathcal{H}_1^\infty$.

In the presence of superselection sectors one may construct operators that distinguish different sectors whilst being a multiple of the unit in each sector. In quantum field theory these are typically global charges, and in our example the macroscopic observables play this role. To see this, we return to Subsection 6.1. It is not difficult to show that for any approximately symmetric sequence $(A_1, A_2, \cdots)$ the limit

$$\overline{A} = \lim_{N\to\infty} \pi_{\mathfrak{S}}(A_N) \tag{6.35}$$

exists in the strong operator topology on $\mathcal{B}(\mathcal{H}_{\mathfrak{S}})$ (Bona, 1988). Moreover, if $A_0 \in \mathcal{A}_0^{(c)} = C(\mathcal{S}(\mathcal{A}_1))$ is the function defined by the given sequence,[287] then the map $A_0 \mapsto \overline{A}$ defines a faithful representation of $\mathcal{A}_0^{(c)}$ on $\mathcal{H}_{\mathfrak{S}}$, which we call $\pi_{\mathfrak{S}}$ as well (by abuse of notation). An easy calculation in fact shows that $\pi_{\mathfrak{S}}(A_0)\Psi = A_0(\tilde{u})\Psi$ for $\Psi \in \mathcal{H}_u$, or, in other words,

$$\pi_{\mathfrak{S}}(A_0) = \oplus_{\tilde{u}\in S^2} A_0(\tilde{u}) 1_{\mathcal{H}_u}. \tag{6.36}$$

Thus the $\pi_{\mathfrak{S}}(A_0)$ indeed serve as the operators in question.

To illustrate how delicate all this is, it may be interesting to note that even for symmetric sequences the limit $\lim_{N\to\infty}\pi(A_N)$ does not exist on $\mathcal{H}_1^\infty$, not even in the strong topology.[288] On the positive side, it can be shown that $\lim_{N\to\infty}\pi(A_N)\Psi$ exists as an element of the von Neumann algebra $\pi(\mathcal{A}_0^{(q)})''$ whenever the vector state ψ defined by Ψ lies in the folium $\mathcal{F}^{\mathfrak{S}}$ generated by all permutation-invariant states (Bona, 1988; Unnerstall, 1990a).

This observation is part of a general theory of macroscopic observables in the setting of von Neumann algebras (Primas, 1983; Rieckers, 1984; Amann, 1986, 1987; Morchio & Strocchi, 1987; Bona, 1988, 1989; Unnerstall, 1990a, 1990b; Breuer, 1994; Atmanspacher, Amann, & Müller-Herold, 1999), which complements the purely C^*-algebraic approach of Raggio & Werner (1989, 1991), Duffield & Werner (1992a,b,c), and Duffield, Roos, & Werner (1992) explained so far.[289] In our opinion, the latter has the advantage that conceptually the passage to the limit $N \to \infty$ (and thereby the idealization of a large system as an infinite one) is very satisfactory, especially in our reformulation in terms of continuous fields of C^*-algebras. Here the commutative C^*-algebra $\mathcal{A}_0^{(c)}$ of macroscopic observables of the infinite system is glued to the noncommutative algebras $\mathcal{A}_1^N$ of the corresponding finite systems in a continuous way, and the continuous sections of the ensuing continuous field of C^*-algebras $\mathcal{A}^{(c)}$ exactly describe how *macroscopic* quantum observables of the finite systems converge to classical ones. *Microscopic* quantum observables of the pertinent finite systems, on the other hand, converge to quantum observables of the infinite quantum system, and this convergence is described by the continuous sections of the continuous field of C^*-algebras $\mathcal{A}^{(q)}$. This entirely avoids the language of superselection rules, which rather displays a shocking *dis*continuity between finite and infinite systems: for superselection rules do not exist in finite systems![290]

6.5 Poisson structure and dynamics

We now pass to the discussion of time-evolution in infinite systems of the type considered so far. We start with the observation that the state space $\mathcal{S}(\mathcal{B})$ of a finite-dimensional C^*-algebra $\mathcal{B}$ (for simplicity

[287] Recall that $A_0(\omega) = \lim_{N\to\infty}\omega^N(A_N)$.

[288] For example, let us take the sequence $A_N = j_{N1}(\mathrm{diag}(1,-1))$ and the vector $\Psi = \uparrow\downarrow\downarrow\uparrow\uparrow\uparrow\uparrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\cdots$, where a sequence of 2^N factors of $\uparrow$ is followed by 2^{N+1} factors of $\downarrow$, etc. Then the sequence $\{\pi(A_N)\Psi\}_{N\in\mathbb{N}}$ in $\mathcal{H}_1^\infty$ diverges: the subsequence where N runs over all numbers 2^n with n odd converges to $\frac{1}{3}\Psi$, whereas the subsequence where N runs over all 2^n with n even converges to $-\frac{1}{3}\Psi$.

[289] Realistic models have been studied in the context of both the C^*-algebraic and the von Neumann algebraic approach by Rieckers and his associates. See, for example, Honegger & Rieckers (1994), Gerisch, Münzner, & Rieckers (1999), Gerisch, Honegger, & Rieckers (2003), and many other papers. For altogether different approaches to macroscopic observables see van Kampen (1954, 1988, 1993), Wan & Fountain (1998), Harrison & Wan (1997), Wan et al. (1998), Fröhlich, Tsai, & Yau (2002), and Poulin (2004).

[290] We here refer to superselection rules in the traditional sense of inequivalent irreducible representations of *simple* C^*-algebras. For topological reasons certain finite-dimensional systems are described by (non-simple) C^*-algebras that do admit inequivalent irreducible representations (Landsman, 1990a,b).

Kenneth F. Schaffner

Statement

and

Readings

Reduction: the Cheshire cat problem, a return to roots, and a place for emergence

Kenneth F. Schaffner
University of Pittsburgh
kfs@pitt.edu

Abstract

In this talk and accompanying paper, I propose two theses, and then examine what the consequences of those theses are for discussions of reduction and emergence. The first thesis is that what have traditionally been seen as robust, reductions of one theory or one branch of science by another more fundamental one are a largely a myth. Although there *are* such reductions in the physical sciences, they are quite rare, and depend on special requirements. In the biological sciences, these prima facie sweeping reductions fade away, like the body of the famous Cheshire cat, leaving only a smile. . . . The second thesis is that the "smiles" are fragmentary patchy explanations, and though patchy and fragmentary, they are *very important,* potentially Nobel-prize winning advances. To get the best grasp of these "smiles," I want to argue that, we need to return to the roots of discussions and analyses of scientific explanation more generally, and not focus mainly on reduction models, though three conditions based on earlier reduction models are retained in the present analysis. I briefly review the scientific explanation literature as it relates to reduction, and then sketch my account of explanation.

The account of scientific explanation I present is one I have discussed before,

but in this paper I try to simplify it, and characterize it as involving field elements (FE) and a preferred causal model system (PCMS) abbreviated as FE and PCMS. In an important sense, this FE and PCMS analysis locates an "explanation" in a typical scientific research *article*. This FE and PCMS account is very briefly illustrated in the talk, but is presented in sufficient detail in the recommended background paper, which summarizes a recent set of neurogenetic papers on two kinds of worm foraging behaviors: solitary and social feeding. One of the preferred model systems from a 2002 *Nature* article in this set is used to exemplify the FE and PCMS analysis, which is shown to have both reductive and nonreductive aspects.

The paper closes with a brief discussion of how this FE and PCMS approach differs from and is congruent with Bickle's "ruthless reductionism" and the recently revived mechanistic philosophy of science of Machamer, Darden, and Craver. The talk will also comment briefly on the "re-emergence of emergence" literature, and suggest there are important heuristic lessons in some of that literature, though the scientific examples in that literature are typically over interpreted and misanalyzed.

Synthese
DOI 10.1007/s11229-006-9031-2

ORIGINAL PAPER

Reduction: the Cheshire cat problem and a return to roots

Kenneth F. Schaffner

Abstract In this paper, I propose two theses, and then examine what the consequences of those theses are for discussions of reduction and emergence. The first thesis is that what have traditionally been seen as robust, reductions of one theory or one branch of science by another more fundamental one are a largely a myth. Although there are such reductions in the physical sciences, they are quite rare, and depend on special requirements. In the biological sciences, these prima facie sweeping reductions fade away, like the body of the famous Cheshire cat, leaving only a smile. ... The second thesis is that the "smiles" are fragmentary patchy explanations, and though patchy and fragmentary, they are very important, potentially Nobel-prize winning advances. To get the best grasp of these "smiles," I want to argue that, we need to return to the roots of discussions and analyses of scientific explanation more generally, and not focus mainly on reduction models, though three conditions based on earlier reduction models are retained in the present analysis. I briefly review the scientific explanation literature as it relates to reduction, and then offer my account of explanation. The account of scientific explanation I present is one I have discussed before, but in this paper I try to simplify it, and characterize it as involving field elements (FE) and a preferred causal model system (PCMS) abbreviated as FE and PCMS. In an important sense, this FE and PCMS analysis locates an "explanation" in a typical scientific research *article*. This FE and PCMS account is illustrated using a recent set of neurogenetic papers on two kinds of worm foraging behaviors: solitary and social feeding. One of the preferred model systems from a 2002 *Nature* article in this set is used to exemplify the FE and PCMS analysis, which is shown to have both reductive and nonreductive aspects. The paper closes with a brief discussion of how this FE and PCMS approach differs from and is congruent with Bickle's "ruthless reductionism"

K. F. Schaffner (✉)
University of Pittsburgh
Dept. HPS - 1017 CL
Pittsburgh PA 15213, USA
e-mail: kfs@pitt.edu

and the recently revived mechanistic philosophy of science of Machamer, Darden, and Craver.

Keywords Emergence explanation · Field model system · Reduction

1 Introduction: two theses about reduction

In this paper, I want to propose two theses, and then examine what the consequences of those theses might be for discussions of reduction and emergence. The first thesis is that what have traditionally been seen as robust reductions of one theory or one branch of science by another more fundamental one are largely a myth. Although there are such reductions in the physical sciences, they are quite rare, and depend on special requirements. In the biological sciences, these prima facie sweeping reductions tend to fade away, like the body of the famous Cheshire cat, leaving only a smile.... The second thesis is that the "smiles" that remain are fragmentary patchy explanations, and though patchy and fragmentary, they are very important, potentially Nobel-prize winning advances. To get the best grasp of them, I want to argue that we need to return to the roots of discussions and analyses of scientific explanation more generally, and not focus mainly on reduction models.

I did not always think that the first thesis was true. Particularly in the physical sciences, it appeared that, we had strong reductions that were constituent parts of actual science—and not mere philosophical quests for unified science. When I studied physics in 1950 s and 1960 s, thermodynamics was taught as a separate course in physics departments, but everyone knew that statistical mechanics was the science underlying thermodynamics. Similarly there were courses offered in optics, but the nature of light was known to be an electromagnetic wave (at least to a good first approximation), and Maxwell's equations could be mathematically manipulated to generate a wave equation, which in turn could be used to explain various laws of optics, such as Snell's law of refraction.

Closer inspection of the explanatory process, however, revealed difficulties.[1] Although one can get Snell's law by derivation from Maxwell's electromagnetic theory, one does not obtain the entire range of Fresnel's theory of physical optics (actually theories is more accurate, since there were several models employed by Fresnel to cover all of optics — (see Schaffner, 1972). Furthermore, to get an explanation of optical dispersion, one has to go beyond Maxwell's theory per se to Lorentz's electron theory. But even Lorentz's theory was not enough to account for all of physical optics, since to get an explanation of the photoelectric effect, one has to go beyond it to Einsteinian elementary quantum mechanics, and an explanation of the optics of moving bodies requires special relativity. The message from this prima facie strong case of intertheoretic reduction is that we get fragmentary and partial explanations of parts of a discipline, but not any type of overall sweeping reduction. The "reductions" are creeping, not sweeping.[2]

[1] These difficulties were systematically developed in the writings of Feyerabend and Kuhn about this time, in 1960s and 1970s, and will be discussed later in this paper.

[2] I first used these terms of "sweeping versus creeping" in my Schaffner (2002a). Neuroethics: reductionism, emergence, and decision-making capacities. In *Neuroethics: Mapping the Field; Conference Proceedings*, May 13–14, 2002, San Francisco: CA. Steven Marcus. New York: Dana Press. vii, 367 but

That said, it needs to be recognized that in what now seem to me to be rather special cases, *almost* sweeping reductions can be found in the physical sciences. The best example, with which I am familiar is the above mentioned reduction of physical optics by Maxwell's electromagnetic theory. The reduction does, as noted, have problems, and fails at the margins where electron theory, quantum mechanics, and special relativity need to be invoked. But the extraordinarily powerful explication of optics by electromagnetic theory needs to be acknowledged, as do the logical features and explanatory strategies of that example that come quite close to fulfilling classical Nagelian reduction conditions (more about these later). A detailed account of exactly how that reduction works, as well as where departures from classical theory are needed, can be found in two, back-to-back books by the distinguished physicist Sommerfeld. Sommerfeld published six advanced textbooks in 1940s covering all of physics, which were based on his extensive lectures on the topics delivered in 1930s. Volume III was entitled *Electrodynamics*, and volume IV, *Optics* (Sommerfeld, 1950a b). The optics in volume IV is developed reductionistically from Maxwell's theory as delineated in volume III, and the two texts represent an in-depth extended exemplar of a sweeping reduction. This is written in the Euclidean–Newtonian mode of entire fields being mathematically derived from a small number of integrated universal physical laws supplemented with simple connections between the fundamental terms in the reduced and reducing theories.

But such a comprehensive, sweeping, deductively elaboratable account seems to be dependent on some rather stringent requirements. Both reduced and reducing fields need to be representable in terms of a small number of principles or laws. Also, the connections between the two fields need to be straightforward and relatively simple, though far from obvious. (It is a simple and general statement that the electric vector *is* the light vector but it is *not obvious* that light is an electromagnetic wave.) Both of these stringent conditions, simple axiomatizablity and simple connectability, fail in significant ways in more complex sciences such as molecular genetics and neuroscience, though that they do fail, or would fail, was not necessarily obvious at the beginning of the Watson–Crick era.

That one encounters creeping rather than sweeping reductions in biology can be illustrated by Kandel's classical explanations of learning in the sea snail *Aplysia* in neuroscience. The standard accounts by Kandel provide explanations of some simple learning behaviors in *Aplysia*, but not *all* of *Aplysia*'s behaviors are explained. (e.g., *Aplysia californicum* engages in a kind of California-style sex involving multiple partners, but I have not seen any molecular cartoon describing and explaining this complex behavior). Additionally, those Kandel models are only partial neural nets and partial molecular cartoons that describe what happens to strengthen synapse connection (Kandel, James, Schwartz, & Thomas, 2000). And the Kandel cartoons (and text explanations) use interlevel language (mixing organs, cells, receptors, second messengers, and ions, among other types of entities at different levels of aggregation) — not a language involving purely chemical entities interacting with other chemical signals. So this is no robust unilevel explanation of learning—even just in A*plysia*—based solely on molecular mechanisms and chemical entities. The reasons for this have been suggested above, lack of any broad scope simple theories, plus the aggregated complexity of the parts of the mechanisms or models involved. Both of these reasons reflect the

Footnote 2 continued
the concepts are latent in my Schaffner (1993a). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.

manner, in which evolution has "designed" living organisms—by opportunistically utilizing whatever bits and pieces of mechanisms may be available and pulling them together in a Rube Goldberg assemblage—not pretty, but satisfactory if it wins in the fitness sweepstakes.

However, though we do not get sweeping reductions in the biological sciences, we do get extremely valuable potentially Nobel-prize winning progress, albeit of a creeping sort. Thus, it is important to know at a general philosophical level what is occurring when, we obtain these important results. The results are *like* reductions, but I think they are better described as *explanations*, using that term as an alternative to reduction because the e-word does not carry the conceptual freight of various reduction models and is a more appropriate general context, within which to analyze what is actually occurring in the biomedical sciences. Such explanatory reductions are in a sense *complementary* to the sweeping theoretical reductions we can find in rare instances in the physical sciences.[3] Neither impugns the character of the other, and which type of reduction one finds will depend on the structure of the disciplines and empirical results. The present paper focuses primarily on these explanatory reductions, but does so with the model of theoretical reduction as a backdrop.

2 A return to roots and a brief history of scientific explanation

In point of fact, a revisiting of the well-spring of the major reduction model—that of Nagel—suggests it was a generalization or extension (but more accurately a specification) of an ancient Aristotelian model of deductive-nomological explanation, now what is often called the Popper-Hempel model (Hempel & Oppenheim, 1948; Popper, 1959), which in the Hempel variants spawned 40–50 years of argument and criticism in the general explanation literature.[4] It is not possible to find textual evidence that Nagel was specifically generalizing Popper-Hempel, since the original publication of the Nagel model in his 1949 contains no bibliographic references. But 1961 version places reduction within the context of explanation, and explanation itself has four patterns according to Nagel, the first and oldest (actually Aristotelian) of which is the deductive model (Nagel, 1961, p. 21). And Nagel did write in 1961 that "reduction, in the sense, in which the word is here employed, is the *explanation* of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain" (p. 338); my italics.

(1) *The deductive-nomological model.* The deductive-nomological model can be illustrated by some simple examples in physics and biomedicine. The model assumes three elements: (1) A set of scientific laws (nomological statements), such as Newton's laws (e.g., $F = ma$) in mechanics, or Ohm's law $V = IR$ (or $V/R = I$), relating voltage, current, and resistance, in the physics of electricity. Additionally, we need (2) a set of initial conditions describing the particular system of interest, e.g., in the physics of electricity, we might have a circuit in which the applied voltage is 3 Volts, and the resistance is 2 Ω. A conclusion, (3), which follows deductively from the laws (here Ohm's law) and the initial

[3] Compare Mayr on the distinction between explanatory and theory reduction in his Mayr (1982). *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge, MA: Belknap Press. pp. 60–63.

[4] Hempel and Oppenheim cite Mill as well as Popper and a number of other authors as sources of their model.

conditions, is that the current in the circuit is 1.5 A. The conclusion here is the event to be explained (the *explanandum*), and the laws and initial conditions are the explainers, or *explanans*.

In the usual order of seeking an explanation, we start with a "why question," e.g., "why is the current in the circuit of interest 1.5 A?"[5] The laws and initial conditions and the derivation are the answer or the explanation. A very similar explanation can be found in biomedicine—more specifically in simple cardiology, where the law of interest is:

Q (blood flow) = pressure gradient/vascular resistance.

In this domain, problems are solved, and explanations given, as in the Ohm's law example above, but now using information about the blood pressure and the arterialvenous system's resistance.

(2) *D-N controversies.* The Hempel–Oppenheim version of the deductive model of explanation generated some 40–50 years of controversy about the adequacy of this model. Hempel himself realized it was not universally applicable, and developed the Inductive- Statistical and Deductive Statistical models to accommodate additional forms of explanation. (Some of this history is reviewed in my (Schaffner, 1993a), but for a more encyclopedic account (see Kitcher & Salmon, 1989). A number of philosophers of science found the model wanting because it seemed to require that an explanation had to involve laws and that it seemed to identify explanation and prediction. More salient for our purposes, were philosophers of science who felt a bigger picture or larger context was needed within which explanations functioned. This larger context included Kuhn's paradigms (Kuhn, 1962), Lakatos' research programmes (Lakatos, 1970), Shapere's domains (Shapere, 1977), Laudan's research traditions (Laudan, 1977), Kitcher's practices (Kitcher & Salmon, 1989), van Fraassen's pragmatic question-oriented analysis of explanation (Van Fraassen, 1980), as well as Railton's notion of an "ideal explanatory text." (Railton, 1980). As I see it, Kuhn's criticisms were in the long-run the most influential against both the Popperan falsification enterprise and the Hempelian logical empiricist tradition. It should be added that Quine, and the writings of a rediscovered Duhem, significantly assisted in this critical effort. In 1960 s, Paul Feyerabend's critiques of what we can construe as Nagel's generalization of the deductive-nomological model to inter-theory reduction was probably most important in convincing philosophers of science that some modifications were needed (Feyerabend, 1962).

As noted in connection with Nagel's views, reduction, in one important sense, is the *explanation* of a higher-level theory, or science, by a lower level more fundamental one (e.g., the reduction of biology by chemistry). In the Nagelian model of reduction, the *explanandum*—that which is to be explained—is a set of laws (theories) fully describing the higher-level or more primitive science to be reduced (e.g., biology, or Newton's mechanics). The *explanans* or explainer is the set of laws (theories) fully describing the more fundamental or more recent science (e.g., molecular chemistry or Einsteinian relativity). And also needed in this generalization are "connectability assumptions" (often called "bridge laws" and sometimes reduction functions) that define (or relate) the higher-level entities

[5] Sometimes the distinction between "why" and "how" questions is introduced into this kind of discussion, but I do not think it is a productive distinction. For some discussion of the ambiguities of "why" questions see P. Kitcher and W. C. Salmon (1989). *Scientific explanation*. Minneapolis: University of Minnesota Press, pp. 141–142.

(e.g., genes) and properties (e.g., dominance) in terms of lower-level entities and properties (e.g., DNA and enzyme action).[6]

Feyerabend and Kuhn argued in their far-reaching analyses based on historical examples that there were no such connections between either earlier and later theories or between higher-level and more fundamental theories. Their arguments were primarily from physics, e.g., citing the transition from Newton to Einstein, and the relation between thermodynamics and statistical mechanics. Both Feyerabend and Kuhn suggested the inter-theoretical relationships were ones of *replacement* of the earlier or higher-level theory by the later or lower-level theory, and that the theories were not only inconsistent—they were actually "incommensurable." Kuhn embedded his analysis in a philosophy of scientific revolutions, with major irrational aspects controlling scientific "progress," and Feyerabend drew parallels with political anarchism and argued that methodologically "anything goes."

(3) *Schaffner and Hull on Genetics.* A late 1960 s–1970 s debate between myself and David Hull mirrored the Nagel–Kuhn/Feyerabend division. In 1967, I had argued (Schaffner, 1967) that molecular biology was in the process, in the wake of Watson and Crick's work, of reducing traditional genetics (e.g., $\text{gene}_1 = \text{DNA}_1$). Hull counterargued (Hull, 1974) there was no way to systematically connect the two forms of genetics, and that possibly molecular genetics was *replacing* traditional genetics. Others joined this debate (e.g., Wimsatt, Kitcher, Rosenberg, & Waters, etc.) (see Schaffner, 2002b) for details and references)—a debate that ran into the 1980s and 1990s. Curiously, given the major strides that molecular biology was making during this period, most philosophers sided with Hull yielding the "Antireductionist Consensus," but scientists typically did not (e.g., Stent), holding to a contrarian "Reductionist Anti-consensus" (Waters, 1990). This debate continues in one form or another, and a recent set of discussions on reduction and genetics (and other sciences) can be found in the recent book edited by van Regenmortel and Hull (Van Regenmortel & Hull, 2002).

(4) *Explanation and Emergence.* One way to characterize emergence, the second Janus-related topic of the Paris conference at which this essay was one paper, is to define it as failure of any possible explanation of a whole *in terms of its parts and their relations* (and expressed only in the parts' language). In their criticisms of Nagelian reduction or analogues of it, Feyerabend and Kuhn did not ever seem to be concerned with *this kind* of an intertheoretic or inter-paradigm relation failure, nor did Hull, though there had been a long line of analogous arguments about the inability to explain or predict higher-level properties by lower-level properties, e.g., in John Stuart Mill and in Claude Bernard (see Schaffner, 1993a, pp. 415–416 for quotes and references). To situate a discussion of emergence and its relations to reduction, I want to distinguish three types of emergence:

Innocuous—The parts, without a specification of the interrelations, do not tell you what the whole will do. For example: the parts of an oscillator are a resistance, a capacitor, and a coil, plus a power source, but the system will not oscillate

[6] In actual reductions, this takes place at a considerably more specified level of detail. See for example the discussion in Watson (1987). *Molecular Biology of the Gene*. Menlo Park, CA: Benjamin/Cummings. of the *lac* operon (pp. 476–480), which specifies operator and promoter DNA sequences.

unless the connections are right, i.e., the connections must be specified for the parts to be an oscillator. This is uncontroversial.
Strong— All the information about the parts and the connections will *never* allow an explanation of the whole. This is *very* controversial; I think it is tantamount to substance pluralism. (For examples of such claims by Mayr and Weiss (see Schaffner, 1993a, pp. 415–417) though probably neither would have accepted substance pluralism as the natural implication of this position.)[7]
Pragmatic—For the immediately foreseeable future, and maybe for many years, we do not have the analytical tools that allow us to infer the behaviors of the wholes (or sometimes even the next level up) from the parts and their connections. It is this pragmatic sense that runs through my present paper. (For related views (see Wimsatt, 1976a; Simon, 1981)).

3 Further data-driven developments related to post-Nagelian reduction models: a short personal history[8]

By the early 1970 s it had become clear to me—largely from a close analysis of the development of the Jacob–Monod operon model that I had begun in 1969—that the Nagel model, and the refinements of it allowing for some aspects of the views of Popper, Feyerabend, & Kuhn (see Schaffner, 1967) had historical problems.[9] More specifically, the reduction models were neither *directive* of in-progress molecular biological research programs, nor were they fully accurate summaries of the *results* of those programs. I published this view in a "peripherality of reduction" thesis paper in (Schaffner, 1974a). Nevertheless, the fine-tuned reduction model did seem to present a reasonable template for a completed successful reduction, and the most detailed elaboration of that analysis was presented in my 1977 essay under the rubric of a generalized reduction-replacement model (GRR) (Schaffner, 1977). This variant added "replacement" to accommodate the explanation of those domains, where the previous theory had been discarded, but a cluster of experimental results remained, akin to what had been suggested earlier by Kemeny and Oppenheim (1956). (Two examples where this kind of replacement occurred where this kind of replacement occurred were in the phlogiston and aether domains; for specific details concerning optics, electricity, the aether, and special relativity (see Schaffner, 1972).

However, even as that 1977 paper was in press, it appeared to me that the nature of theory in biology had initially been misconceived, both by myself and by most philosophers of biology (compare Ruse, 1973), and that a different analysis of biological theory as a collection of overlapping causal and *interlevel* models was a much more accurate representation of what was found in *real biology*. The paper that developed

[7] By substance pluralism, I mean the existence of two independent substances, such as mind and matter were for Descartes, or matter and field seemed to be for Einstein—see his comments on Maxwell in his Autobiographical Notes, pp. 1–95 in Schilpp, and Einstein (1949). *Albert Einstein, philosopher-scientist*. Evanston, III: Library of Living Philosophers.

[8] The expression "data-driven" to describe what is reviewed here is found in Sterelny and Griffith's reduction discussion (see p. 144 and also related comments on p. 147 in Sterelny, and P.E. Griffiths (1999)). *Sex and death: An introduction to philosophy of biology. Chicago, III: University of Chicago Press.*

[9] For my historical account of the Jacob–Monod operon model see my Schaffner, (1974b). Logic of discovery and justification in regulatory genetics. *Studies in history and philosophy of science* 4: 349–385.

that thesis was submitted and accepted for publication in 1977, but its length, and the policy of the journal to which it had been sent, led to a three year delay in publication as (Schaffner 1980). In the ensuing years I re-thought the theses of this 1980 paper, and eventually the themes from the 1980 theory structure paper were partially integrated with the earlier reduction models and published in my (1993a) book.[10]

The gist of the overly long (115 pages) reduction chapter in my (1993a) was that most purported reductions, in biology are at best partial reductions, in which corrected or slightly modified fragments or parts of the reduced science are reduced (explained) by parts of the reducing science, and that in partial reductions a causal/mechanical approach (CM) is better at describing the results than is a formal reduction model (e.g., the GRR). The GRR model, however, is a good executive summary and regulative ideal for unilevel clarified—and essentially static—science; and it also pinpoints where identities operate in reductions, and emphasizes the causal generalizations inherent in and sometimes explicitly found in mechanisms. As noted earlier, some such virtually complete reductions can in point of fact be found in the history of physics. The more common partial *reductions,* though usually termed "reduction," are, paradoxically typically *multi-level* in both the reduced and reducing sciences, mixing relatively higher entities (and predicates); with relatively lower-level entities (and predicates); it is extremely rare that there are only two levels. What happens is a kind of "integration"—to use Sterelny and Griffith's term — in the sense that there is a mixing and intermingling of entities and strategies from higher level and more micro domains in a consistent way (Sterelny & Griffiths, 1999). In some ways this integration is reminiscent of what Kitcher and Culp (Culp, 1989) termed an "explanatory extension," though I have disagreed with much of the unificatory and anti-causal baggage that such a view seems to take (Schaffner, 1993a, pp. 499–500).

A table from my 1993a book is produced on the following page to illustrate these conclusions. In the 1993 reduction chapter, I also elaborated — using some of the views of Wesley Salmon, though not accepting some key features of his causal approach [11] — on the strengths of a causal mechanical approach and what value the more formal GRR model might have as well. (See table 1, following page.)

In reviewing that reduction chapter, as well as my core explanation Chap. 6 in the 1993 book, for this 2003 conference on reduction and explanation, it became even clearer to me that these views might be still further sharpened and better exemplified. The example used in the present paper has as its backdrop the fact that for the 10 years between 1993 book's publication and the Paris reduction and emergence conference, I had immersed myself in behavioral genetics. This initially began from a "simple systems" approach that was an outgrowth of a 1993 workshop convened by the NIH's National Institute of Child Health on behavioral genetics work then being done on *Caenorhabdibs elegans* and *Drosophila*. But that workshop also dealt with then just-published and emerging human discoveries (Dean Hamer presented his recent sexual orientation results at the conference, and Robert Plomin outlined his major molecular program that was to develop in the rest of 1990s). Some details of

[10] I say "partially" integrated because on rereading Chap. 3, 6, and 9 of Schaffner (1993a). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press., I think there are aspects of the theory structure account that are only partly recognized in the reduction discussion; I say more about this later in the present paper, where I hope to have accomplished a fuller integration.

[11] I found Salmon's discussion of marks, forks, and interactions not fully satisfactory. For a recent summary of Salmon's approach, and criticism, see Woodward, J. (2003). *Making things happen: A Theory of causal explanation*. New York: Oxford University Press, Chap. 8.

Table 1 CM and GRR Approaches in different states of completions of reductions

State of completion/approach	CM	GRR
Partial/patchy/fragmentary/interlevel	Box 1—CM approach usually employed; interlevel causal language is more natural than GRR connections.	Box 2—Complex GRR Model: the connections are bushy and complex when presented formally, but GRR does identify points of identity, as well as the generalizations operative in mechanisms.
Clarified science/unilevel at both levels of aggregation	Box 3—Either approach could be used here, but where theories are collection of prototypes, the bias toward axiomatization or explicit generalization built into the GRR approach will make it less simple than CM.	Box 4—Simple GRR Model: best match between Nagelian reduction and scientific practice.

this behavioral genetics work can be found in my Schaffner (1998a) and in Schaffner (2000, 1999, 2001d). And further re-analysis of reduction models with these behavioral genetics inquiries as a backdrop suggests the following conclusions presented in the remainder of this paper.

4 A return to roots: a minimalist explanation-reduction model employing a causal mechanical approach

(1) *The conditions for a partial reduction.* In attempting to return to the explanatory roots of reductions, I will begin with what distinguishes a non-reductive explanation for one that is (at least partially) reductive. One way to work toward a minimalist set of distinguishing conditions is to look at strong candidates for reductive explanations in a science of interest, for which a general reduction account is desired. The following conditions were suggested by a review of the Kandel models for *Alpysia* learning that were discussed in my 1993a book (Chap. 6), and are available in an updated and accessible form in many standard neuroscience texts, including Kandel et al. (2000). Thus, the scientific details of those examples will not be re-presented in the current article. The general conclusion of my recent review is that successful (though partial) *reductions are* causal mechanical *explanations, if,* in addition to whatever, we call adequate causal mechanical explanations (this will come later), the following three conditions hold. (I will state these in the material mode, though they can be rephrased so that they refer to sentences which describe the referents.) The first two of these are informal and the third is a formal condition that retains an important formal condition of the Nagel (and GRR model)as follows:

(1) the explainers (or *explanans*) (more on what these are later) are a part (or parts) of the organism/process, i.e., they are a (partially) decomposable microstructure(s) in the organism/process of interest.[12]

[12] Partial decomposability has been discussed by Simon (1981). *The sciences of the artificial*. Cambridge, MA: MIT Press.and by Wimsatt (1976a). Reductionism, levels of organization, and the

(2) the *explanandum* or event to be explained is a grosser (macro) typically aggregate property or end state.
(3) Connectability assumptions (CAs) need to be specified, (sometimes these CAs are called bridge laws or reduction functions), which permit the relation of macrodescriptions to microdescriptions. Sometimes these CAs are causal consequences, but in critical cases they will be identities (such as gene ① = DNA sequence ①, or aversive stimulus = bacterial odor).[13]

So far I have said little about what these three conditions are conditions of, or to what account of causal mechanical explanation they need to be added, to reflect what we find in partial and patchy reductions. Before, I sketch the explanation model, however, it is important to underscore a prima facie somewhat paradoxical aspect of partial reductions. This is their *dual* interlevel character.

2. *It's interlevel all over*. Though possibly under appreciated, I think it fair to say that it is reasonably broadly recognized that typical reducing/explaining models are interlevel (mixing together ions, molecules, cells, and cell networks, and not infrequently, even organs and organisms). Less appreciated, I think, is that the *reduced* theory/model is *also interlevel*, but not as fundamental or fine-structured as is the reducing model. Earlier I referred to the debate I had with Hull and others about Mendelian (transmission) genetics beginning in late 1960s. In that debate, and offshoot debates among a number of others in the philosophy of biology throughout 1970s and 1980s, I do not think it was fully recognized, by me or others, that Mendel's theory of heredity was *itself* vigorously interlevel. Mendel had not only summarized his discoveries in genetics in terms of laws, but in the same article he also proposed an explanation in terms of underlying factors that segregated randomly, thus mixing in his theory phenotypes and what were later called genes. To underscore the intertwined and interlevel nature of Mendelian genetics, consider the following quotation (in translation) from Mendel's 1865 paper:

> "In our experience, we find everywhere confirmation that constant progeny can be formed only when germinal cells and fertilizing pollen are alike, both endowed with the potential for creating identical individuals, as in normal fertilization of pure strains. Therefore, we must consider it inevitable that in a hybrid plant also identical factors are acting together in the production of constant forms. Since the different constant forms are produced in a single plant, even in just a single flower, it seems logical to conclude that in the ovaries of hybrids as many kinds of germinal cells (germinal vesicles), and in the anthers as many kinds of pollen cells are formed as there are possibilities for constant combination forms and that these germinal and pollen cells correspond in their internal make-up to the individual forms." (factors = genes; my added underlining identifies different levels of entities) (quoted from Mendel's essay in (Stern and Sherwood 1966)

Footnote 12 continued
mind-body problem. In G. Globus et al. (Ed.), *Consciousness and the brain*. New York: Plenum Press, pp. 205–267.

[13] One possibility that retains what Nagel called a correspondence rule interpretation of these connectability assumptions is to use a causal sequence interpretation of the logical empiricists' correspondence rules. For how this might be further analyzed see my Schaffner (1969). Correspondence rules. *Philosophy of Science 36*, 280–290. paper on correspondence rules and also Suppe's discussion of this view in his (1977) book, pp. 104–106.

A reduction of Mendel's laws and his process of factor segregation typically involve an appeal to entities intertwined from several levels of aggregation: cells, chromosomes, DNA strands and loci, and enzymes, so even this paradigm of reduction is also interlevel at the present time. The reduction is also partial because it is impossible (so far as I know) to account for *all* of the pea plant's phenotypes strictly in terms of molecular features and mechanisms, even in 2005.

(3) *The elements of a causal mechanical explanation model: field and preferred causal model system*. For this paper, I am going to restrict an account of my model of explanation to what might be called "local" explanations. This notion of "local" is intended to indicate I am referring to explanations within a time slice of a field that use a currently accepted theory or class of mechanisms. I distinguish this type of explanation from "global" explanations that capture explanations across successive historical periods of scientific change — of the sort that Kuhn described as revolutions involving major paradigm change.[14]

I want to argue that a satisfactory local explanation model, which I think can illuminate what occurs in partial reductions, has two main substantive components, with each substantive component having a closely related epistemological/logical aspect.[15] The first substantive component involves the scientific field, but more accurately (FE), and its epistemological aspect is a kind of inductive logic of comparative evaluation of plausible explanatory candidates, representing preliminary plausibility judgments. The second substantive component is the preferred (causal) model system (PCMS), which itself is an elaboration and extension of one of the plausible explanatory candidates of the first field element component. The epistemological aspect of the second component is a claim that the PCMS is a causal system representing a temporal process; such a system can be elaborated and tested using either deductive logic and/or statistical methodological logic. In a previous paper—(see Schaffner, 2000) I have called this the *field and focus* model, which itself was a renaming of an account of explanation I developed in my (1993a, Chap. 6). In the present paper, I have re-renamed it a "*field elements and preferred causal model system*" or FE- and PCMS account in an attempt to underscore the key constituent concepts involved in the explanation model. (For convenience, I suggest pronouncing FE and PCMS as "*fee-pems*.") Each component needs some additional discussion, and in the following section I provide an illustration that relates the model to partial reductions as follows:

(a) *The Field and Field Element Component.* I should begin by noting that the general sense of field used here is (probably) *not* the sense originally used in Darden and Maull's "interfield theory" approach to intertheoretic relations, including their different way of looking at reductions (Darden & Mull, 1977). I think a reading of their seminal 1977 paper suggests that each field is unilevel and that it is *interfield* explanations that are surrogates for (or alternatives to) reductions.

[14] This second type of (global) explanation involves what I call in Chap. 5 of my 1993a book "temporally extended theories" that allow for replacement in some circumstances. Using such temporally extend theories is too complex for a first cut at getting back to the explanatory roots of reductions. This global type of explanation also involves issues of "global evaluation" (trans-theoretical criteria) that needs to bracketed for another paper, though a list of those criteria and a Bayesian analysis of how they work can be found in Chap. 5 of my 1993a.

[15] I have debated whether this aspect should be best characterized as logical or epistemological. It seems to involve a logic of weighing and comparing, but the aspect also indicates varying strengths of warranted belief. Further below, I will describe subscribing to a type of causality as the "epistemological" aspect of the second substantive component.

That said, the approaches taken there and in my present paper may well be quite congruent, with any differences possibly more terminological than real. The field *elements* concept has certain analogies with Shapere's notion of a "domain," (Shapere, 1977) as a set of "items of information" having "an association," but I think the field elements notion differs in being broader, and at the same time clearer, in the sense that *particular research articles define those field elements by specifying them*. (For additional comments on the pros and cons of the domain concept, see my 1993a, p. 52.)

The substantive *field* component in my approach contains most of the basic generalizations, mechanisms, experiments, and theories, typically introduced in a standard textbook for the field, and field *elements draw from* the field. This makes the typical field (and usually the FE, which selects portions from the field) vigorously interlevel, as well as (typically and also paradoxically) *inter*disciplinary, in virtually all instances with which I am familiar.[16] A textbook may however draw on several pre-existing fields, e.g., neuroscience and molecular biology, which can usually roughly be distinguished by referring to consensus *classic* texts in those fields. This general field component has possibly been captured implicitly in explanation models in the philosophical literature by Railton's (and also Salmon's) notion of an ideal explanatory text. Concrete examples, of such texts in biology would be the Watson *Molecular Biology of the Gene* series of texts, or the Kandel and Schwartz *Principles of Neural Science* series of texts; in medicine this would be a standard medical textbook, such as *Harrison's Principles of Internal Medicine*. A more specific example is the field of *C. elegans* research, typified by what are known as the *Worms* I and II collections of essays. Although I do not believe anyone has ever done this, someone reasonably well acquainted with a field could make a list of major explanatory devices in a field by working through such a textbook. Some of these would go by the terms model, or mechanism, or law, or generalization, or theory, or hypothesis. And they would not be independent, nor representable in a simple hierarchy, since some would be partial components of others, and would reappear in slightly different forms multiple times. It is that richness and complexity that I believe, we find and also have to deal with in real science.

In the kind of partial reductions I want to explicate in the present paper, we should begin by considering a typical scientific journal research article (not a textbook nor collection of articles nor a review article, usually) in which an *explanation* is proffered. The typical article situates the phenomenon to be explained within a field (or sometimes in two and possibly more fields) and then presents a list of the classes of alternative possible explanations utilizing the field elements (possible explanations as seen by the authors as being viable in the field) for the phenomenon of interest. The alternatives are not exhaustive of other elements that can be found in the field as a whole, but are proposed, sometimes as a cluster,

[16] How to best define a field and a discipline are likely to require considerably more analysis that I provide in this paper, and there may be historical and sociological dimensions that need to be taken into account to provide an adequate characterization of these terms and their relations. That neuroscience is extraordinarily interdisciplinary is a point that has been stressed by several commentators, including Craver and Bickle—see Craver (2005). Beyond reduction: mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences*, Bickle, J. (2006a). Neuroscience. In *Encyclopedia of Philosophy 6*: 563–572.

or sometimes seriatim, in the article.[17] The possible explanations in a scientific article often are evaluated and roughly ranked as best, better, good, worse, and worst, which is a *logical aspect* of the first substantive (field) component.[18]

(b) *The PCMS component.* The second, and perhaps most salient substantive component of my model of explanation, is the designation of a PCMS, which implicitly or explicitly involves the laws or generalizations that are relevant to the particular problem or problems of interest to the investigator. Such a PCMS can be quite simple, as when one introduces a simple single-locus Mendelian model of a dominant/recessive gene pair, say as a Punnett-square representation, and then uses that model to explain the inheritance of Huntington's disease or cystic fibrosis. Alternatively, the PCMS can be more complex, as in an explanation of feeding behavior using specific mutants and neuron types in *C. elegans* which will be discussed below, a Kandel cartoon depicting presynaptic sensitization in *Aplysia*, or a Hodgkin and Huxley classic sodium action potential model.

There is no formal limit on the degree of complexity of a PCMS, though these are always idealized to a greater or lesser extent. A critically important aspect of a PCMS is that there is a list of general assumptions embedded in the preferred model system that describe the system under study, and which are believed to generalize to other like systems. These generalization(s), however, may have narrow or broad applicability: the generalization may be family—or population-limited, strain limited, or species limited, though possibly even broader, holding for all mammals, for example. The generalizations are typically *qualitative causal generalizations*, describing parts of mechanisms in a process, such as an inducer combining with repressor molecule in a lac operon model with the resultant loss of the repressor's affinity for the operator. In (fairly) rare cases, these generalizations will be mathematical formulas, such as a Nernst equation or a flux equation.[19] These generalizations that are instantiated in the models can be found in the text and especially in the figure legends of pictorial representations of models and mechanisms (and also referred to in the indexes) in standard biological textbooks, such as the Watson or Kandel series noted above. In my view, the explanatory elements in the biomedical sciences are a collection of (sometimes overlapping) model systems (PCMSs).

The epistemological aspect of this second substantive component of my explanation account is a claim about causality appealed to in the explanation.[20] Most explanations

[17] In medicine, an analogy to a list of alternative hypotheses is what is known as a "differential diagnosis" for the cause of an illness that afflicts a particular patient or population of patients; it is a list of possible diseases.

[18] The list of alternatives and a comparison of their strengths and weaknesses can include alternative possible states in which a mechanism might be can also be evaluated in this approach. The original suggestion for this evaluative dimension is due to van Fraassen (1980), who asked, e.g., why is this circuit off rather than on?; why is this patient sick rather than healthy?

[19] For an excellent example, of a model, which uses multiple equations (see Bogen's account of the classic Hodgkin and Huxley 1952 paper on action potentials) in Bogen, J. (2005). Regularities and Causality: Generalizations and Causal Explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences 36.*

[20] Again I term this aspect an "epistemological" aspect, though causal claims involve, in addition, logical dimensions (at least in the sense of types of conditionals) and also metaphysical dimensions (in the sense of ontic claims and process metaphysics). For details of my view on these dimensions see pp. 298–307 of my 1993a, and for a slightly later discussion of a manipulation interpretation of causation (see Schaffner, 1993b). Clinical trials and causation: Bayesian perspectives. *Statistical Medicine 12* (15–16), 1477–1494; discussion 1495–1499.

in basic science are causal mechanical, but they might involve a random probabilistic process, or even a human motivational account (in economics, or human psychology, for example).[21] Example of a causal mechanical explanation strategy can be found in many of Wesley Salmon's examples.[22] A related logical aspect of this epistemological aspect is closely related to the type of the causality assumed in the proposed PCMS studied: deterministic systems can easily be elaborated using deductive logic; probabilistic causality suggests the need for an inductive logic.[23] The Popper–Hempel (or perhaps more accurately Aristotle–Mill–Popper–Hempel) model of explanation falls into the first type, involving deductive logic.

5 How this explanation-partial reduction model is illustrated in practice

(1) *A recapitulation and overview of the explanation-partial reduction process.* A reasonably detailed illustration of how this two component explanation-partial reduction model works, especially in partial reductions, may help clarify it. I have selected the area of molecular behavioral genetics for my example, and as the specific case some recent work on two types of feeding behaviors of the worm, *C. elegans*. To reiterate the general process: first, a typical scientific or medical research article provides explanations, for example, of an organism's behaviors. But even in such focused research articles, the broader context of the problem(s) are sketched (however briefly) and assumptions are made that the reader is knowledgeable about the organism and familiar with the relevant parts of the field (the FEs) in neuroscience, or genetics, or molecular biology, etc. Within this broad framework, such a research article quickly zeros in on several well defined questions, and then proceeds to present answers to the questions in terms of the advances that are the rationale for the publication of the paper. Within the context of these answers, it is possible to pick out a focus (or foci), and ask what are the specific PCMSs used in the explanation. It is at this point with a focus on a specific PCMS that, we can usefully begin to appeal to the nature of the law(s), mechanisms, component parts, and pathways, as well as to scrutinize the nature of the inference (deductive, statistical) and ask whether this explanation is causal (or perhaps unificatory), and if causal, what type(s) of causal conditions are operative. This general pattern of explanation is found in many of the papers in the study of the nematode and other model organisms. A useful preface to my specific example may be to first, and very briefly, summarize some basic facts about the worm for the readers of this paper. In an important sense, the following section will introduce some of the FEs needed to characterize an explanation (and a partial reduction) in the case below.

[21] It is not possible in this paper to elaborate on the motive-cause distinction, nor on related narrative versus causal explanations. I say more about this in my forthcoming book from Oxford University press, tentatively titled *Behaving: What's Genetic and What's Not?*

[22] Some explanation may claim to be noncausal, e.g., unificatory, but I find this claim (by Kitcher) questionable—see my 1993a, Chap. 6, but for a possibly more positive view see Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press, Chap. 8.

[23] Probabilistic explanation using infinite classes is deductively elaboratable. The logic in some cases might even be abductive in some instances, maybe in "inference to the best explanation," though abductive inference (and logic) is even less well understood than inductive.

(2) *The anatomy, genetics, and neurology of C. elegans.* This model organism, which has attracted more than 1000 fulltime researchers worldwide, received additional recognition in 2002 when the Nobel Prize in biology and medicine went to the worm: i.e., to Brenner, Horvitz and Sulston for cell death work in *C. elegans*. The animal (yes, researchers do use that term) has been called "the reductionist's delight" (Cooke-Deegan 1994), but a review of the *C. elegans* literature indicates its behavior is much more complex than originally thought: most behavior types relate to genes influencing them in a *many-many relation* (for details see my 1998a). Nevertheless, a recent essay in *Nature Neuroscience* commented on the use of the worm in the following terms:

> With a circuitry of 302 invariant neurons, a quick generation time, and a plethora of genetic tools, *C. elegans* is an ideal model system for studying the interplay among genes, neurons, circuits, and behavior. (Potter and Luo 2003)

Some features of the worm's anatomy are presented in Figure 1.

This 1 mm long adult hermaphrodite has 959 somatic nuclei and the male (not pictured) 1,031 nuclei; there are about 2,000 germ cell nuclei (Hodgkin et. al 1995). The haploid genome contains 1×10^8 nucleotide pairs, organized into five autosomal and one sex chromosome (hermaphrodites are XX, males XO), comprising about 19,000 genes. The genes have all been sequenced. The organism can move itself forward and backward by graceful undulatory movements, and responds to touch and a number of chemical stimuli, of both attractive and repulsive or aversive forms, with simple reflexes. More complex behaviors include egg laying and mating between hermaphrodites and males (Wood, 1988, p. 14)—and the worm also learns—as studied by Rankin and others (Rankin, 2002).

The nervous system is the worm's largest organ, being comprised, in the hermaphrodite, of 302 neurons, subdividable into 118 subclasses, along with 56 glial and associated support cells; there are 95 muscle cells on which the neurons can synapse. The neurons have been fully described in terms of their location and synaptic connections.

These neurons are essentially identical from one individual in a strain to another (Sulston et al., 1983; White et al., 1986), and form approximately 5,000 synapses, 700 gap junctions, and 2,000 neuromuscular junctions (White et al. 1986). The synapses are typically "highly reproducible" (~85% the same) from one animal to another, but are not identical, due to "developmental noise." (For further details about this concept see my 1998a.)[24]

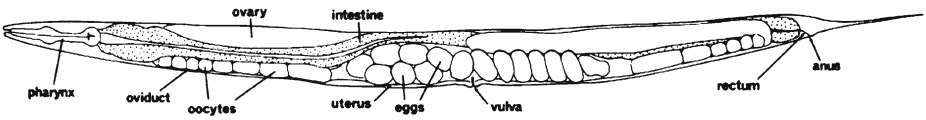


Fig. 1 Reprinted from *Developmental Biology*, Mar; 56(1):110-56. Sulston JE, Horvitz HR. "Post-embryonic cell lineages of the nematode, Caenorhabditis elegans." Pages: 110-56, Copyright (1977), with permission from Elsevier

[24] de Bono (personal communication) indicates that "There may also be more plasticity in synapse number/size than indicated in the mind of the worm—difficult to say as only 2–3 worms were sectioned in the John White's em (electron micrograph) studies."

(3) *An interesting "exception" to the many-many genes behavior relation?* In their 1998 essay in the prestigious scientific journal *Cell*, de Bono and Bargmann investigated the feeding behavior of two different strains of the worm, one of which engaged in solitary feeding, and the other in social feeding (aggregated in a crowd) (de Bono & Bargmann, 1998). A picture of the two types of strains showing these two contrasting behaviors is provided below in Fig 2.
De Bono and Bargmann summarized their 1998 results in an abstract in *Cell* (de Bono & Bargmann, 1998, p. 679), which I closely paraphrase here, interpolating just enough in the way of additional information that nonspecialists can follow the nearly original abstract text:

Natural subpopulations of *C. elegans* exhibit solitary or social feeding behavior. Solitary eaters move slowly across a surface rich in the bacteria (a bacterial "lawn") on which they feed and also disperse on that surface. Social eaters on the other hand move rapidly on the bacteria and bunch up together, often near the edge of a bacterial lawn. A knock-out ("loss of function") mutation in a gene known as *npr*-1 causes a solitary strain to take on social behavior. This gene

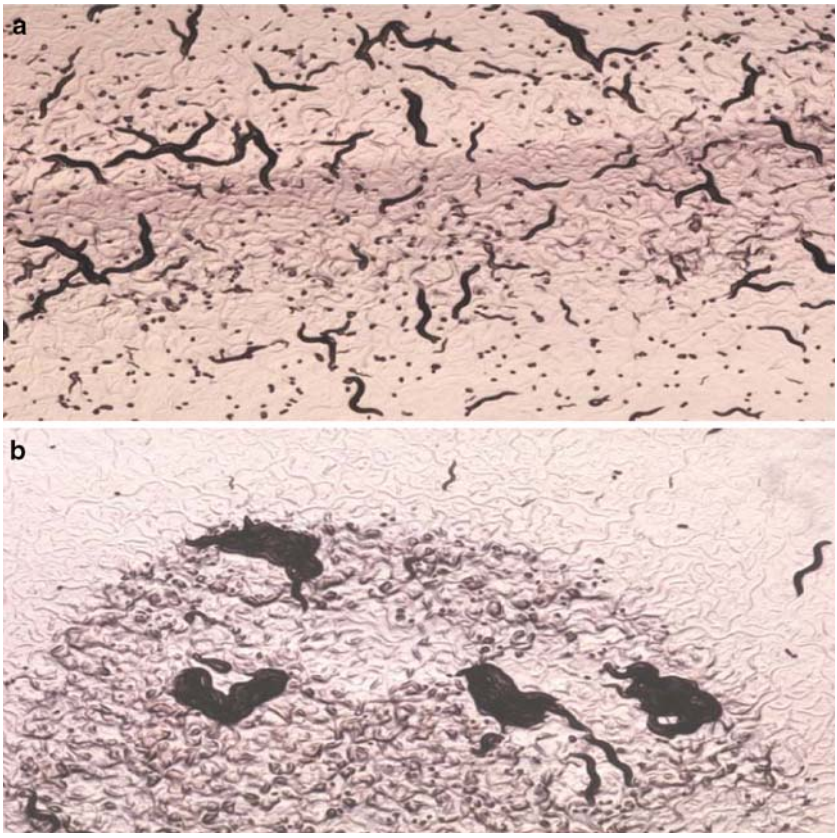


Fig. 2 Figure from Rankin (2002), based on de Bono's work. Reprinted by permission from Macmillan Publishers Ltd.: *Nature Reviews Genetics*, Catharine H. Rankin "From gene to identified neuron to behaviour in Caenorhabditis elegans," vol 3, pp. 622–630, copyright (2002). http://www.nature.com/nrg/

is known to encode a type of protein, here it is NPR-1, known as a G-protein coupled receptor, a protein that acts like a switch to open or close ion channels in nerve cells. This NPR-1 protein is similar to a family of proteins called Y receptors that are widely present in the nervous system of other organisms and relate to feeding and foraging behavior in other species. Two variants of the NPR-1 protein that differ only in a single amino acid (phenylalanine versus valine) occur naturally in the wild. One variant, termed NPR-1 215F (with phenylalanine, abbreviated as F) is found exclusively in social strains, while the other variant, NPR-1 215 V (with valine) is found only in solitary strains. The difference between the *F* and *V* variants are due to a single nucleotide difference in the gene's DNA sequence (T versus G). Inserting a gene that produces the *V* form of the protein can transform wild social strains into solitary ones. Thus these only slightly different proteins generate the two natural variants in *C. elegans'* feeding behavior.

This remarkable paper by de Bono and Bargmann made strong claims involving a genetic explanation of behavior. At the end of the introduction to this 1998 essay, the authors wrote that "we show that variation in responses to food and other animals in wild strains of *C. elegans is due to* natural variation in *npr*-1" (my emphasis) (1998, p. 679). The phenotype difference is actually somewhat more complex, and not just related to social or solitary feeding in the presence of sufficient amount of bacterial food supply. As already indicated, the social and solitary strains also differ in their speed of locomotion. Also, the two types differ in burrowing behavior in the agar jell surface on which the worms are studied in the laboratory. But de Bono and Bargmann contended that "a single gene mutation can give rise to all of the behavioral differences characteristic of wild and solitary strains" (1998, p. 680).

de Bono and Bargmann offered several "different models that could explain the diverse behavioral phenotypes of *npr*-1 mutants" (1998, p. 686), but added that "resolution of these models awaits identification of the cells, in which *npr*-1 acts, and the cells that are the source of the *npr*-1 [sic] ligands [those molecules that bind to and regulate this receptor]" (1998, p. 686).

(4) *Complications and an example of a PCMS.* This wonderfully "simple" story of one gene that influences one type of behavior in the worm was told in 1998 as just described. Since then, further work by de Bono and Bargmann, who did search for the cells in which *npr*-1 acts and for the source of the NPR-1 ligands, has indicated that the story is more complex. In follow-up work to determine how such feeding behavior is regulated, de Bono and Bargmann have proposed two so-far separate pathways (de Bono, Tobin, & Davis, 2002) (Coates & de Bono, 2002). One pathway suggests that there are modifying genes that restore social feeding to solitary feeders under conditions of external environmental stress. The other pathway is internal to the organism, and will be very briefly described at the conclusion of this section. (An accessible overview of the two pathways, and some possible very interesting connections with fly and honeybee foraging and feeding behaviors, can be found in (Sokolowski, 2002)'s editorial accompanying the publication of the two de Bono et al., 2002 papers).

The first 2002 paper by de Bono and Bargmann, also writing with Tobin, Davis, and Avery, indicates how a partially reductive explanation works, and also nicely illustrates the features of the FE and PCMS system approach discussed above. The explanandum

is to account for the difference in social versus solitary feeding patterns, as depicted in Fig. 2 above. The explanation (at a very abstract level) is contained in the title of the paper "Social feeding in *C. elegans* is induced by neurons that detect aversive stimuli." The specifics of the explanation appeal to the 1998 study as background, and look at *npr*-1 mutants, examining what *other* genes might prevent social feeding, thus restoring solitary feeding in *npr*-1 mutants. A search among various *npr*-1 *mutants* (these would be social feeders) indicated that mutations in the *osm*-9 and *ocr*-2 genes resulted in significantly more *solitary* feeding in those mutant animals. (Both of these genes code for *components* of a sensory transduction ion channel known as TRPV (transient receptor potential channel that in vertebrates responds to the "vanilloid" (V) compound capsaicin found in hot peppers). Both the *osm*-9 and *ocr*-2 genes are required for chemoattraction as well as aversive stimuli avoidance). Additionally, it was found that *odr*-4 and *odr*-8 gene mutations could disrupt social feeding in *npr*-1 mutants. The *odr*-4 and *odr*-8 genes are required to localize a group of olfactory receptors to olfactory cilia. Interestingly, a mutation in the *osm-3* gene, which is required for the development of 26 ciliated sensory neurons, *restores social* feeding in the *odr*-4 and *ocr*-2 mutants. (Readers who have followed the account of the genetic influences on ion channels, other genes, and neurons thus far are now entitled to a break.)

de Bono et al. present extensive data supporting these findings in the article. Typically the reasoning with the data examines the effects of screening for single, double, and even triple mutations that affect the phenotype of interest (feeding behaviors), as well as looking at the results of gene insertion or gene deletion. This reasoning essentially follows Mill's methods of difference and concomitant variation (the latter because graded rather than all-or-none results are often obtained), and is prototypical causal reasoning. Also of interest, are the results of the laser ablation of two neurons that were suggested to be involved in the feeding behaviors. These two neurons, known as ASH and ADL are implicated in the avoidance of noxious stimuli and toxic chemicals. Identification of the genes noted above (*osm*-9*, ocr*-2*, odr*-4, and *odr*-8) allowed the investigators to look at where those genes were expressed (by using Green Florescent Protein (GFP) tags). It turned out that ASH and ADL neurons were the expression sites. The investigators could then test the effects of laser beam ablation of those neurons, and showed that ablation of both of them restored a solitary feeding phenotype, but that the presence of either neuron would support social feeding.

The net result of the analysis is summarized in a "model for social feeding in C. elegans" on the following page [Fig. 5 in de Bono et al.; my Fig. 3].

The legend for the model reads as follows:

> Figure 5(c), A model for social feeding in *C. elegans*. The ASH and ADL nociceptive neurons are proposed to respond to aversive stimuli from food to promote social feeding. This function requires the putative OCR-2/OSM-9 ion channel. The ODR-4 protein may act in ADL to localize seven transmembrane domain chemoreceptors that respond to noxious stimuli. In the absence of ASH and ADL activity, an unidentified neuron (XXX) [involving *osm*-3] represses social feeding, perhaps in response to a different set of food stimuli. The photograph shows social feeding of a group of >30 npr-1 mutant animals on a lawn of *Escherichia coli*.

This model *is* the preferred causal model system for this *Nature* article. It is simplified and idealized, and uses causal language such as "respond to" and "represses."

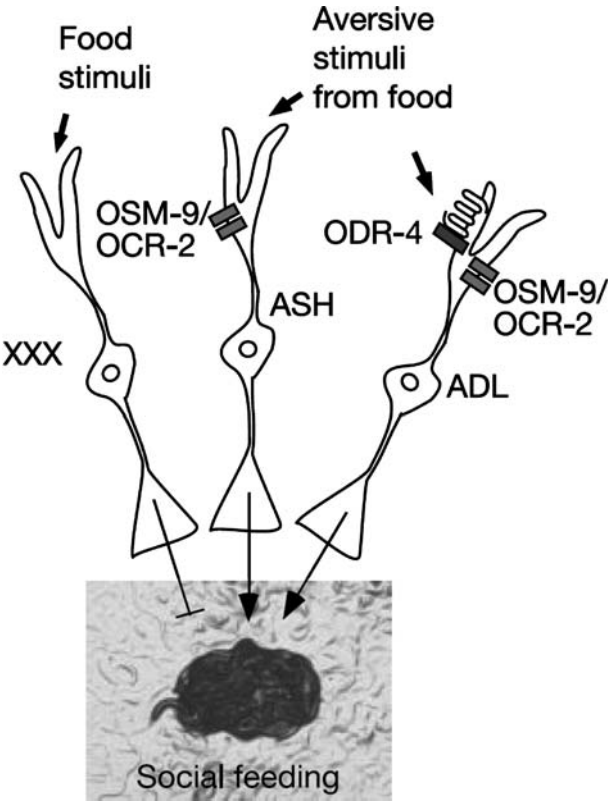


Fig. 3 From de Bono et al. (2002). Reprinted by permission from Macmillan Publishers Ltd: *Nature*, Mario de Bono, David M. Tobin, M. Wayne Davis, Leon Avery, Cornelia I. Bargmann, "Social feeding in Caenorhabditis elegans is induced by neurons that detect aversive stimuli," v. 419, pp. 899–903, copyright (2002). http://www.nature.com

(The causal verbs also contain the word "act," to which I return in Sect. 7, since much has been made in recent years in the philosophy of biology and neuroscience literature about "activities," as opposed to causation, which may be present in "mechanisms.") The PCMS is clearly interlevel. The fields on which the model draws are molecular genetics and neural science. Scattered throughout the article are occasional alternative but possible causal pathways (FE), which are evaluated as not as good an explanation as those provided in the preferred model system presented. (One example is the *dauer* pheromone explanation, discussed on p. 899 of de Bono et al. 2002; another is the "reducing stimuli production" versus "reducing stimuli detection" hypotheses on p. 900 of that article.)

The preparation or *experimental system* investigated in the laboratory (this may include several data runs of the "same" experimental system) is identified in its relevant aspects with the preferred model system. At the abstract or "philosophical" level, the explanation proceeds by identifying the laboratory experimental system with the theoretical system – the PCMS – and exhibiting the *explanandum* as the causal consequence of the system's behavior. The *explanans* here uses molecular biology and is mainly comparative rather than involving quantitative derivational reasoning, in the

sense that in this paper two qualitatively different end states—the solitary and the social states of the worms—are compared and contrasted.[25] The theoretical system (the PCMS) utilizes generalizations of varying scope, often having to appeal to similarity analyses among like systems (e.g. the use of TRPV channel *family*) to achieve the scope, as well as make the investigation of interest and relevance to other biologists (e.g., via analogies of the NPR-1 receptor to *Y* receptors and the internal worm circuit to cyclic GMP signaling pathways found in flies and bees that control foraging and feeding behavior—(see Sokolowski, 2002). For those concerned with philosophical rigor, the preferred model system and its relations to model-theoretic explanation can be made more philosophically precise (and technical), along the lines suggested in a "philosopher-speak" footnote below.[26]

The discussion sections of scientific papers are the usual place where larger issues are raised, and where extrapolations are frequently found. This is also the case in this de Bono et al. (2002) paper where the discussion section states that "food, food acquisition, and population density are important regulators of aggregation in a variety of species." (902). The paper concludes on an evolutionary note, tying the proximate cause model to a distal causal (i.e., evolutionary) advantage, where the authors write:

> The data in this paper and in the accompanying paper suggest that the regulation of social feeding behaviour in *C. elegans* is complex, involving several layers of positive and negative inputs. Such complexity may have evolved as a result of the tension between cooperation and competition that underlies social behaviour, and may be important to ensure that social behaviour is induced only when it offers a selective advantage.

Further work on the circuits that affect social and solitary feeding has been done in addition to what has just been described in detail above. Earlier I mentioned an essay

[25] A quantitative derivation of a path of *C. elegans* motion that agrees with the experimentally observed path can be computed based on neural theory, though the explanation quickly becomes extraordinarily complicated—see my summary of Lockery's results using this type of approach in my (2000).

[26] The following philosophically general account parallels the discussion in my 1993 book. It assumes an analysis of biological explainers as involving models representable as a collection of generalizations of variable scope instantiated in a series of overlapping mechanisms as developed in Chap. 3 of the 1993 book. We can, as described in that chapter, employ a generalization of Suppes' set-theoretic approach and also follow Giere (1984) in introducing the notion of a "theoretical model" as "a kind of system whose characteristics are specified by an explicit definition" (1984, p. 80). Here entities $\eta_1, \ldots, \eta_n$ will designate neurobiological objects such as neuropeptide receptors, the Φs such *causal* properties as "ligand binding" and "neurotransmitter secretion," and the scientific generalizations $\Sigma_1, \ldots, \Sigma_n$ will be of the type "This odorant activates a G-protein cascade opening an ion channel." Then $\Sigma_i(\Phi(\eta_1, \ldots, \eta_n))$ will represent the ith generalization, and

$$\prod_{i=1}^{n} [\Sigma_i(\Phi(\eta_1, \ldots, \eta_n))]$$

will be the conjunction of the assumptions (which we will call Π) constituting the preferred model system or PCMS. Any given system which is being investigated or appealed to as explanatory of some explanandum is such a PCMS if and only if it satisfies Π. We understand Π, then, as implicitly defining a kind of natural system, though there may not be any actual system that is a realization of the complex expression Π. To claim that some particular system satisfies Π is a theoretical hypothesis, which may or may not be true. If it is true, then the PCMS can serve as an explanandum of phenomena such as "social feeding." (If the PCMS is potentially true and well-confirmed, then it is a "potential and well-confirmed explanation" of the phenomena it entails or supports.)

that appeared simultaneously with the above paper in *Nature*. This second paper by Coates and de Bono (2002) described a regulatory circuit that sensed the internal fluid in the worm and controlled social versus solitary forms of behavior. It involved different neurons (AQR, PQR, and URX), and was affected by *tax* 2 and 4 gene mutations – genes which produce components of a cyclic GMP-gated ion channel.[27]

Also, in late 2003, de Bono's group was able to identify the ligands, which stimulate the NPR-1 receptor (Rogers, Reale, & Kim 2003). These are a class of neuropeptides known as "FMRFamide and related peptides" (FaRPs) that stimulate foraging receptors in other species. In the worm, the relevant FaRPs are encoded by 22 different *flp* genes that can potentially produce 59 FaRP peptides by alternative splicings. It was also reported in this paper that comparative sequencing of the two NPR-1 variants (the *F* and *V* forms) as well as three other species of *Caenorhabditis*, suggests that the *social* form of the receptor is ancestral, and that the behavior of solitary feeding arose later via a gain of function mutation. This is preliminary conclusion, and some insect researchers find it implausible, believing that social behaviors are likely to appear later than solitary activities (de Bono, personal communication). But that may depend on the different selection pressures experienced in different environments by different species. (More recent articles on The worm stress a pathway from oxygen (O_2) on NPR-1 via body cavity neurons. See, de Bono et al., 2005.)

6 This explanation is both reductive and non-reductive

The above example is typical of molecular biological explanations of behavior. Behavior is an organismic property, and in the example is actually a populational property (of aggregation), and the explanation appeals to entities that are *parts* of the organism, including molecularly characterized genes and molecular interactions such as ligand-receptor bindings and G-protein coupled receptor mechanisms—thus this is generally characterized as a *reductive* explanation. But it represents *partial* reduction—what I termed reduction of the *creeping* sort—and it differs from *sweeping* reductive explanations because of several important features as follows:

(1) It does not explain *all* cases of social versus solitary feeding; a different though somewhat related model (that of Coates & de Bono, 2002) is needed for the internal triggering of solitary behavior in *npr*-1 mutants. (Also compare de Bono et al., 2005.)
(2) Some of the key entities, such as the signal from bacteria that is noxious to the worms and the neuron represented by XXX, have not yet been identified.
(3) It utilizes what might be termed "middle-level" entities, such as neuronal cells, in addition to molecular entities.
(4) It is not a quantitative model that derives behavioral descriptions from rigorous general equations of state, but is causally qualitative and only roughly comparative.
(5) Interventions to set up, manipulate, and test the model are at higher aggregative levels than the molecular, such as selection of the worms by their organismic

[27] More recently, de Bono's lab showed that the internal circuit involves soluble guanylate cyclases in that pathway. See Cheung, Arellano-Carbajal, and Rybicki (2004). Soluble guanylate cyclases act in neurons exposed to the body fluid to promote *C. Elegans* aggregation behavior. *Curr Biology, 14*(12), 1105–1111. These appear to be activated by oxygen – (see Gray, Karow, & Lu, 2004). Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature, 430*(6997), 317–322.

properties (feeding behaviors), distributing the worms on an agar plate, and ablating the neurons with a laser.

The explanation does meet the three conditions delineated above on page 15, namely

(1) the explainers (here the preferred model system as shown in Fig. 3) are a partially decomposable microstructure in the organism/process of interest.
(2) the explanandum (the social or solitary feeding behavior) is a grosser (macro) typically aggregate property or end state.
(3) The CAs, sometimes called bridge laws or reduction functions, are involved, which permit the relation of macrodescriptions to microdescriptions. Sometimes these CAs are causal sequences as depicted in the model figure where the output of the neurons under one set of conditions causes clumping, but in critical cases the CAs are identities (such as social feeding = clumping, and aversive stimulus = (probably) bacterial odor).

Although reductive, the preferred model system explanation is not "ruthlessly reductive," to use Bickle's phrase, even though a classical organismic biologist would most likely term it strongly reductionistic in contrast to their favored nonreductive or even antireductionist cellular or organismic points of view. It is a *partial* reduction.

7 Will "mechanism language" suffice?

One recent philosophical alternative to classical models of theory reduction can be found in what Bickle calls, in this collection of essays, "the recently revived *mechanistic* philosophy of science." (Bickle, 2006b, this volume) This revival dates to the seminal article by Machamer, Darden, and Craver (2000) that stressed the importance of the "mechanism" concept as an alternative to law-based approaches to explanation and to reduction. In this approach, a mechanism is "a collection of entities and activities organized in the production of regular changes from start or set up conditions to finish or termination conditions" (Machamer, Darden, & Craver, 2000 p. 3). The analysis has been applied to examples in the neurosciences and molecular biology, and recognizes that mechanisms need not be molecular, but can be multi-level see (Craver, 2004 submitted). In some of its variants, the approach wishes to eschew causal language, causal generalizations, and any appeals to standard counterfactual analyses, that are typically developed as elucidations of causation (compare Schaffner, 1993a, pp. 296–312; Glennan, 1996; Woodward, 2003 with Tabery (2004) and Bogen (2004).

An appeal to mechanisms, as a contrast with an emphasis on high-level general theories, is a viable approach. In biology there are few such general theories (with component laws) that are broadly accepted, though population genetics is a notable exception. An early commitment to theories such as population genetics as representing the best examples of biological theory (see Ruse, 1973) is one, as I argued, in my 1980 and again 1993a, Chap. 3, that skewed the appreciation of philosophers of biology away from better or more representative alternative approaches to theory structure and explanation. And in that 1980 article and in (1993a) Chap.3 as well as in Chaps. 6 and 9, I frequently utilized references to "mechanisms" as another way to describe the "models" that are so widely found in biology, and which function broadly as surrogates for theories in the biomedical sciences.

But the *strong* form of appeals to mechanisms, as in early arguments by Wimsatt (1976b) seemed to aim at avoiding any discussion of generalizations and laws of

working of a mechanism, an avoidance which appeared both philosophically incomplete (see my 1993a, pp. 494–495 for specifics), as well as contradicted by the way biologists present their own models. A paradigm case of how generalizations are articulated to form a model can be found in Jacob and Monod's classic paper on the operon model.[28] In their concluding section they write that "a convenient way of summarizing the conclusions derived in the preceding sections of this paper will be to organize them into a model designed to embody the main elements, which we were led to recognize as playing a specific role in the control of protein synthesis; namely the structural, regulator and operator genes, the operon, and the cytoplasmic repressor." Jacob and Monod then state the generalizations, which constitued the model.[29] Similar generalizations can be found in the figure legend from de Bono et al. (2002) quoted above on p. 32.

This avoidance of generalizations by the revived mechanistic tradition is even more evident in the recent essays by Tabery (2004) and also in Darden (2004; 2005) and especially in Bogen (2004; 2005), which also seems to me to try to replace the admittedly still problematic concept of causation with appeals to "activities"—a notion that I find much more opaque than causation. (In those places in scientific articles where terms like "acts" appear, I think a good case can be made that what is being referred to is plain old-fashioned causal action.)

But in a weaker form, such as in (Glennan, 1996) and in most of Machamer et al. (2000), the revived mechanistic philosophy of science appears to me to be an important complement to the account of explanation developed in the present paper, as well as to my 1993a and 2000 essays. I had noted in my 1993a that appeals to mechanisms that eschewed generalizations (such as Wimsatt's 1976b) were problematic for a number of reasons, a chief one of which was that earlier writers in this tradition appeared to take "mechanism" as a largely unanalyzed term and place a very heavy burden on that term. The new mechanistic philosophy of science remedies that problem by articulating a complex analysis of the terminology involved in appeals to mechanisms, but some of the stronger theses, such as those replacing causation by activities, seem to me to move in a less promising direction.

8 Summary and conclusion

In this paper, I began by proposing two theses, and then examined what the consequences of those theses were for reduction and emergence. The first thesis was that what have traditionally been seen as robust reductions of one theory or one branch of science by another more fundamental one are a largely a myth, though some rare instances of them can be found in physics. On closer inspection, and particularly in biology, these reductions seem to fade away, like the body of the famous Cheshire

[28] Another paradigmatic example of how generalizations, and even simplified "laws" are involved in the articulation of a model or mechanism can be found in Hodgkin and Huxley's classic article on the action potential in the giant squid axon: Hodgkin and Huxley (1952). A quantitative description of membrane current and Its application to conduction and excitation in nerve. *Journal of Physiology, 117*, 500–544. Bogen (2005) analyzes Hodgkin and Huxley's model construction as not supporting a typical generalization account, but I read their paper differently.

[29] A full quotation of the statement of the operon model from Jacob and Monod (1961), Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*, 318–356. can be found on 158–159 of my 1993a.

cat, leaving only a smile…. The second thesis was that the "smiles" are fragmentary patchy explanations, and often partial reductions, and though patchy and fragmentary, they are very important, potentially Nobel-prize winning advances.

To get the best grasp of them, I argued that we needed to return to the roots of discussions and analyses of scientific explanation more generally, and not focus mainly on reduction models, though three conditions based on earlier reduction models are retained in the present analysis. This led us through a brief history of explanation and its relation to reduction models, such as Nagel's, and through an account of my own evolving views in this area. Although the account of scientific explanation, I presented above is one I have discussed before, in this paper I tried to simplify it, and characterized it as involving and abbreviated as FE and PCMS. This FE and PCMS account was then applied to a recent set of neurogenetic papers on two kinds of worm foraging behaviors: solitary and social feeding. One of the preferred model systems from a 2002 *Nature* paper was used to illustrate the FE and PCMS analysis in detail, and was characterized as a partial reduction.

The paper closed with a brief discussion of how this FE and PCMS approach partially differed from and partially was congruent with Bickle's "ruthless reductionism" (Bickle 2003) and the recently revived mechanistic philosophy of science of Machamer, Darden, Craver. In that section I could only very briefly indicate some parallels of these approaches with the one developed in the present paper. Clearly further discussion will continue on these topics for some time to come, and should deepen our appreciation of both the power and the limits of reductive explanations.

Acknowledgements Presented at the workshop on reduction and emergence in Paris in November 2003 as part of a research project funded by the French national research foundation (CNRS), and organized by the Institut Jean Nicod; workshop essays to be submitted to *Synthese.* Special thanks to Mario de Bono, Ken Kendler, members of the DCHPB, and participants at the University of Oslo's reduction workshop for comments on an earlier draft. This material is based upon work performed while a Visiting Fellow at the Center for Philosophy of Science at the University of Pittsburgh, and was supported by the National Science Foundation under Grant No. 0324367. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

References

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer.

Bickle, J. (2006a). Neuroscience. In D.M. Borchert (Ed.), *Encyclopedia of Philosophy.* Farmington Hills, MI: Macmillan Reference USA. Volume 6, pp. 563–572.

Bickle, J. (2006b) *Synthese* (this volume).

Bogen, J. (2004). Analyzing causality: The opposite of counterfactual is factual. *International Studies in the Philosophy of Science, 18*(1), 3–26.

Bogen, J. (2005). Regularities and causality: Generalizations and causal explanations. *Studies in the History and Philosophy of Biology and Biomedical Science*, 36. [H-H paper]

Cheung, B. H., Arellano-Carbajal, F., & Rybicki, I. et al. (2004). Soluble guanylate cyclases act in neurons exposed to the body fluid to promote *C. Elegans* aggregation behavior. *Current Biology, 14*(12), 1105–1111.

Coates, J. C., & de Bono, M. (2002). Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *C. Elegans*. *Nature, 419*(6910), 925–929.

Cooke-Deegan, R. (1994). *Gene Wars.* New York: Norton.

Craver, C.F. (2005). Beyond reduction: Mechanisms, multifield integration, and the unity of science. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36:373–396.

Culp, S., & Kitcher, P. (1989). Theory structure and theory change in molecular biology. *British Journal for the Philosophy of Science, 40*, 459–483.

Darden, L. (2005). Relations among fields: Mendelian, cytological and molecular mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36: 349–371.
Darden, L. (2006). *Reasoning in biological discoveries: essays on mechanisms, interfield relations, and anomaly resolution.* Cambridge, UK: Cambridge University Press.
Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science, 44*, 43–64.
de Bono, M., & Bargmann C. I. (1998). Natural variation in a neuropeptide *Y* receptor homolog modifies social behavior and food response in *C. Elegans. Cell, 94*(5), 679–689.
de Bono, M., Tobin, D. M., & Davis, M. W. et al. (2002). Social feeding in *C. Elegans* is induced by neurons that detect aversive stimuli. *Nature, 419*(6910), 899–903.
de Bono, M., & Villu Maricq, A. (2005). Neuronal substrates of complex behaviours in *C. Elegans. Annual Review of Neuroscience, 28,* 451–501.
Feyerabend, P. (1962). Explanation, reduction, and empiricism. In H. M. Feigl, & G. Minnesota (Eds.), *Minnesota studies in the philosophy of science* (Vol. 3, pp. 28–97). Minneapolis, MN: University of Minnesota Press.
Giere, R. (1984). *Understanding scientific reasoning* (2nd ed.) New York: Holt, Reinhart and Winston.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntis, 44*, 49–71.
Gray, J. M., Karow, D. S., & Lu, H. et al. (2004). Oxygen sensation and social feeding mediated by a *C. Elegans* guanylate cyclase homologue. *Nature, 430*(6997), 317–322.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science, 15*, 135–175.
Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology, 117*, 500–544.
Hodgkin, J., Plasterk, R. H., & Waterston, R. H. (1995). The nematode Caenorhabditis elegans and its genome. *Science.* Oct 20; *270* (5235), 410–414.
Hull, D. L. (1974). *Philosophy of biological science*. Englewood Cliffs, NJ: Prentice-Hall.
Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology, 3*, 318–356.
Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill, Health Professions Division.
Kemeny, J., & Oppenheim, P. (1956). On reduction. *Philosophical Studies, 7*, 6–17.
Kitcher, P., & Salmon, W. C. (1989). *Scientific explanation*. Minneapolis, MN: University of Minnesota Press.
Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos, & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge: Cambridge University Press.
Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth*. Berkeley: University of California Press.
Machamer, P., Darden, L., & Craver C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Belknap Press.
Nagel, E. (1961). *The structure of science; Problems in the logic of scientific explanation*. New York: Harcourt.
Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.
Potter, C. J. & Luo, L. (2003) Food for thought: a receptor finds its ligand. *Nature Neuroscience, 6* Nov, 1119.
Railton, P. (1980). *Explaining explanation*. Ph.D. Dissertation, Princeton: Princeton University.
Rankin, C. H. (2002). From gene to identified neuron to behaviour in *C. Elegans*. *Natural Review Genetics, 3*(8), 622–630.
Rogers, C., Reale, V., & Kim, K. (2003). Inhibition of *C. Elegans* social feeding by fmrfamide-related peptide activation of *Npr*-1. *Natural Neuroscience, 6*(11), 1178–1185.
Ruse, M. (1973). *Philosophy of biology*. London: Hutchinson.
Schaffner, K. F. (1967). Approaches to reduction. *Philosophy of Science, 34*, 137–147.
Schaffner, K. F. (1969). Correspondence rules. *Philosophy of Science, 36*, 280–290.
Schaffner, K. F. (1972). *Nineteenth century Aether theories*. Oxford: Pergamon Press.
Schaffner, K. F. (1974a). The peripherality of reductionism in the development of molecular biology. *Journal of the History of Biology, 7*, 111–139.
Schaffner, K. F. (1974b). Logic of discovery and justification in regulatory genetics. *Studies in History and Philosophy of Science, 4*, 349–385.

Schaffner, K. F. (1977). Reduction, reductionism, values, and progress in the biomedical sciences. In R. Colodny (Ed.), *Logic, laws, and life* (Vol. 6, pp. 143–171). Pittsburgh: University of Pittsburgh Press.
Schaffner, K. F. (1980). Theory structure in the biomedical sciences. *The Journal of Medicine and Philosophy, 5*, 57–97.
Schaffner, K. F. (1993a). *Discovery and explanation in biology and medicine*. Chicago: University of Chicago Press.
Schaffner, K. F. (1993b). Clinical trials and causation: Bayesian perspectives. *Statistical Medicine, 12*(15–16), 1477–1494 (discussion 1495–1499).
Schaffner, K. F. (1998a). Genes, behavior, and developmental emergentism: one process, indivisible? *Philosophy of Science, 65*(June), 209–252.
Schaffner, K. F. (1999). Complexity and research strategies in behavioral and psychiatric genetics. In A. Ronald, R. Carson, & A. Mark, (Eds.), *Behavioral genetics: The clash of culture and biology* (pp. 61–88). Baltimore: Johns Hopkins University Press.
Schaffner, K. F. (2000). Behavior at the organismal and molecular levels: the case of *C. Elegans. Philosophy of Science, 67*, S273–S278.
Schaffner, K. F. (2001d). Nature and nurture. *Current Opinion in Psychiatry, 14*(September), 486–490.
Schaffner, K. F. (2002a). Neuroethics: reductionism, emergence, and decision-making capacities. In *Neuroethics: Mapping the Field: Conference Proceedings, May 13–14, 2002*. San Francisco, California: Steven Marcus, New York: Dana Press, p. 367.
Schaffner, K. F. (2002b). Reductionism, complexity and molecular medicine: genetic chips and the 'globalization' of the genome. In M. H. V. Hull, & D. Van Regenmortel, (Eds.), *Promises & Limits of reductionism in the biomedical sciences* (pp. 323–347). London: John Wylie.
Schilpp, P. A., & Einstein, E. (1949). *Albert Einstein, Philosopher-scientist*. Evanston, IL: Library of Living Philosophers.
Shapere, D. (1977). Scientific theories and their domain. In F. Suppe (Ed.), *The structure of scientific theories* (pp. 518–565). Illinois university Press, Urbana, IL.
Simon, H. (1981). *The sciences of the artificial*. Cambridge, MA: MIT Press.
Sokolowski, M. B. (2002). Neurobiology: social eating for stress. *Nature, 419*(6910), 893–894.
Sommerfeld, A. (1950a). *Lectures on theoretical physics: Electrodynamics*. New York: Academic Press.
Sommerfeld, A. (1950b). *Lectures on theoretical physics: Optics*. New York: Academic Press.
Sterelny, K., & Griffiths, P. E. (1999). *Sex and death: An introduction to philosophy of biology*. Chicago, IL: University of Chicago Press.
Stern, C., & Sherwood, E. R. (1966). *The Origin of genetics; a Mendel source book*. San Francisco: W. H. Freeman.
Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. (1983), The embryonic cell lineage of the nematode. *Caenorhabditis Elegans, Developmental Biology, 100*, 64–119.
Sulston, J. E., & Horvitz, H. R. (1977) Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Developmental Biology, Mar; 56* (1):110–156.
Tabery, J. (2004). Activities and interactions. *Philosophy of Science, 71*, 1–15.
Van Fraassen, B. C. (1980). *The scientific image*. Oxford, New York: Clarendon Press, Oxford University Press.
Van Regenmortel, M., & Hull, D., (Eds.), (2002). *Promises and limits of reductionism in the biomedical sciences*. London: John Wylie Ltd.
Waters, C. K. (1990). Why the antireductionist consensus would not survive the case of classical genetics. *Proceedings of the Philosophy of Science Association, 1*, 125–139.
Watson, J. D. (1987). *Molecular biology of the gene*. Menlo Park, CA: Benjamin/Cummings.
White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Phil. Trans. Roy. Soc. London Ser. B, 314, 1–340.
Wimsatt, W. (1976a). Reductionism, levels of organization, and the mind-body problem. In G. Globus et al. (Eds.), *Consciousness and the brain* (pp. 205–267). New York: Plenum Press.
Wimsatt, W. (1976b). Reductive explanation: a functional account. In R. S. Cohen et al. (Ed.), *Proceedings of the 1974 Philosophy of Science Association*. Dordrecht: Reidel, pp. 671–710.
Wood, W. (Ed.), (1988). *The Nematode*: Caenorhabditis elegans. Cold Spring Harbor: Cold Spring Harbor Press.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Michael D. Silberstein

Statement

and

Readings

When Super-Theories Collide: A Brief History of the Emergence/Reduction Battles between Particle Physics and Condensed Matter Theory

Abstract

In the last few decades one of the most publicized controversies in fundamental physics has been the argument between condensed matter theory (CMT) physicists such as P.W. Anderson, Robert Laughlin and David Pines on the one hand, and particle physicists such as Steven Weinberg and Leonard Susskind on the other over which theoretical framework is in the best position to unify physics and lead it into the twenty first century. For reasons that will be made clear, CMT has been branded as the purveyor of emergence and particle physics considered the champion of reduction in this struggle. This battle still rages today in a volley of books and articles such as Laughlin's *A Different Universe: Reinventing Physics from the Bottom Down* (2005) and Susskind's *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design* (2005). The key events in this fight will be detailed, from the publication of P.W Anderson's classic *More is Different* article (1972) to the protracted debates about whether or not to fund the Superconducting Super Collider (SSC) and on up to the skirmishes of the present day. The historical significance of these machinations can only be fully appreciated when it is clear exactly what is at issue philosophically, methodologically and empirically between these two warring factions of fundamental physics. Thus by way of conceptual analysis, a taxonomy of various critical notions of emergence and reduction will be provided and the

combatant's claims properly situated therein. Though as we shall see, this is no easy task as both sides equivocate madly in their use of the terms "emergence" and "reduction." In addition to raising profound ontological questions about the structure of the world such as the true nature of interlevel relations, epistemic questions about fundamental scientific explanation and intertheoretic relations, our history lesson suggests that theoretical physics (especially quantum gravity) may well be in a revolutionary Kuhnian state. We will find that there is fundamental disagreement over what is in fact fundamental and disagreement over how, if at all, the physical sciences and the world can be unified.

"Reduction, Emergence, and Explanation" The Blackwell Guide to the Philosophy of Science. Chapter 5, pp. 80-107.

Chapter 5

Reduction, Emergence and Explanation

Michael Silberstein

Introduction: The Problem of Emergence and Reduction

Can everything be reduced to the fundamental constituents of the world? Or can there be, and are there, non-reducible, or emergent entities, properties and laws? What exactly do we mean by "reduction" and "emergent" when we ask such questions? For example, if everything can be reduced to the fundamental constituents of the world, does that preclude the existence of emergent entities, properties or laws? Obviously, the answers to many of these questions depend on what is meant by the terms "reduction" and "emergence." These terms are used in a variety of ways in the literature, none of which is uniquely privileged or uniform. Therefore, clarity is crucial to avoid confusion and equivocation. The first task of this chapter is to sort out and schematize the main versions of reduction and emergence, and then to turn to the current debates. The current state of the reductionism vs. emergentism debate is examined and the Final section looks toward future debate.

Historically, there are two main construals of the problem of reduction and emergence: ontological and epistemological; see Stephan (1992), McLaughlin (1992) and Kim (1999) for historical background.

- The ontological construal: is there some robust sense in which everything in the world can be said to be *nothing but* the fundamental constituents of reality (such as super-strings) or at the very least, *determined by* those constituents?
- The epistemological construal: is there some robust sense in which our scientific theories/schemas (and our common-sense experiential conceptions) about the macroscopic features of the world can be *reduced to* or *identified with* our scientific theories about the most fundamental features of the world?

Yet, these two construals are inextricably related. For example, it seems impossible to justify ontological claims (such as the cross-theoretic identity of conscious mental processes with neurochemical processes) without appealing to epistemological claims (such as the attempted intertheoretic reduction of folk psychology to neuroscientific theories of mind) and vice versa. We would like to believe that the unity of the world will be described in our scientific theories and, in turn, the success of those theories will provide evidence for the ultimate unity and simplicity of the world; things are rarely so straightforward.

Historically "reductionism" is the "ism" that stands for the widely held belief that both ontological and epistemological reductionism are more or less true. Reductionism is the view that the best understanding of a complex system should be sought at the level of the structure, behavior and laws of its component parts plus their relations. However, according to mereological reductionism, *the relations* between basic parts are themselves reducible to the intrinsic properties of the relata (see below). The ontological assumption implicit is that the most fundamental physical level, whatever that turns out to be, is ultimately the "real" ontology of the world, and anything else that is to keep the status of real must somehow be able to be 'mapped onto' or 'built out of' those elements of the fundamental ontology. Relatedly, fundamental theory, *in principle*, is deeper and more inclusive in its truths, has greater predictive and explanatory power, and so provides a deeper understanding of the world.

"Emergentism", historically opposed to reductionism, is the "ism" according to which both ontological and epistemological emergentism are more or less true, where ontological and epistemological emergence are just the negation of their reductive counterparts. Emergentism claims that a whole is "something more than the sum of its parts", or has properties that cannot be understood in terms of the properties of the parts. Thus, emergentism rejects the idea that there is any fundamental level of ontology. It holds that the best understanding of complex systems must be sought at the level of the structure, behavior and laws of the whole system and that science may require a plurality of theories (different theories for different domains) to acquire the greatest predictive/explanatory power and the deepest understanding.

The problem of reduction and emergence is (and has been) of great interest and importance in philosophy and scientific disciplines from physics to psychology; see *Philosophical Studies*, Vol. 95, 1999 and Beckermann et al., (1992), Blazer et al., (1984) and Sarkar (1998). It is always possible to divide claims about reductionism and emergentism. One may accept ontological reductionism but reject epistemological reductionism, and vice-versa, likewise for ontological emergentism and epistemological emergentism. Further, one may restrict the question of reductionism and emergentism to particular domains of discourse. For example, one might accept reductionism (epistemic and/or ontic) for the case of classical mechanics and quantum mechanics, but reject it (epistemic and/or ontic) for the case of folk psychology and theories from neuroscience.

The Varieties of Reductionism: Ontological and Epistemological

The basic idea of reduction is conveyed by the "nothing more than . . ." cliché. If *X*s reduce to *Y*s, then we would seem to be justified in saying or believing things such as "*X*s are nothing other (or more) than *Y*s," or "*X*s are just special sorts, combinations or complexes of *Y*s." However, once beyond cliches, the notion of reduction is ambiguous along two principal dimensions: the types of items that are reductively linked and the nature of the link involved. To define a specific notion of reduction, we need to answer two questions:

- Question of the relata: Reduction is a relation, but *what types of things* may be related?
- Question of the link: *In what way(s)* must the items be linked to count as a reduction?

Let us first consider the question of the relata. The things that may be related have been viewed either as:

- real world items – entities, events, properties, etc. – which is the *Ontological* form of *Reduction*, or
- representational items – theories, concepts, models, frameworks, schemas, regularities, etc. – which is the *Epistemological* form of *Reduction*.

Thus, the first step in our taxonomy subdivides into two types of reduction. Each type further subdivides based on the specific kinds of relata in question. Ontological subdivisions include: parts and wholes; properties; events/processes; and causal capacities. Epistemological subdivisions include: concepts; laws (epistemically construed); theories; and models. (These lists are not intended to be exhaustive, but merely representative.)

The second question about the link was *in what way(s)* must the items be *linked* to count as a case of reduction? Again, there are a variety of answers on both the ontological and the epistemological side.

Question of the *ontological link*: How must things be related for one to ontologically reduce to the other? At least four major answers have been championed:

- Elimination
- Identity
- Mereological supervenience (includes "composition", "realization" and other related weaker versions of this kind of determination relation)
- Nomological supervenience/determination

The relative merits of competing claims have been extensively debated, but for present purposes it suffices to say a brief bit about each and give a general sense of the range of options.

Elimination

One of the three forms of reduction listed by Kemeny and Oppenheim in their classic paper on reduction (1956) was replacement, i.e., cases in which we come to recognize that what we thought were *X*s are really just *Y*s. *X*s are eliminated from our ontology, e.g., claims of demonic possession (Rorty, 1970; Churchland, 1981; Dennett, 1988; Wilkes, 1988, 1995).

Identity

Identity involves cases in which we continue to accept the existence of *X*s but come to see that they are identical with *Y*s (or with special sorts of *Y*s). *X*s reduce to *Y*s in the strictest sense of being the same thing as *Y*s. This may happen when a later *Y*-theory reveals the true nature of *X* to us. For example, we have come to see that heat is just kinetic molecular energy and that genes are just functionally active DNA sequences. However, the identity does not require elimination or deny the existence of the prior items, rather we see that two distinct theories have described or referred to the same entities/properties.

Mereological supervenience

Reductionism pertaining to parts and wholes goes by several names: "mereological supervenience," "Humean supervenience" and "part/whole reductionism." Mereological supervenience says that the properties of a whole are determined by the properties of its parts (Lewis, 1986, p. 320).

More specifically, mereological supervenience holds that all the properties of the whole are determined by the qualitative intrinsic properties of the most fundamental parts. Intrinsic properties being non-relational properties had by the parts which these bear in and of themselves, without regard to relationships with any other objects or relationships with the whole. Sometimes, philosophers say that intrinsic properties are properties that an object would have even in a possible world in which it alone exists. Paradigmatic examples include mass, charge, and spin. Further, intrinsic properties are much like the older primary qualities. It is notoriously difficult to define the notion of an intrinsic property or a relational property in a non-circular and non-question begging manner; nonetheless, philosophers and physicists rely heavily on this distinction (Lewis, 1986).

Nomological supervenience/determination

Fundamental physical laws (*ontologically construed*), governing the most basic level of reality, determine or necessitate all the higher-level laws in the universe. Mereological supervenience, on the one hand, says that the intrinsic properties of the most basic parts *determine* all the properties of the whole – this is a claim about part-whole determination (purely physical necessity). Nomological supervenience is about *nomic necessity*, the most fundamental laws of physics ultimately necessitate all the special science laws, and therefore these fundamental laws determine everything that happens (in conjunction with initial or boundary conditions). Thus, if two worlds are wholly alike in terms of their most fundamental laws and in terms of initial/boundary conditions, then we should expect them to be the same in all other respects.

In *epistemological reduction* one set of representational items is reduced to another. These representational items are all human constructions and often taken to be linguistic or linguistic surrogates, though this need not be the case. It was noted above that reduction relations might hold among at least four different kinds of representational items.

Concerning the *epistemological links (or relations) that do the reducing*, a diversity of claims have been made. Some relations, such as derivability, make sense as a relation between theories seen as sets of propositions but not among models or concepts. However, certain commonalities run through the family of epistemological-reductive relations. Most of the specific variants of epistemological reduction fall into one of four general categories:

- Replacement
- Theoretical-derivational (logical empiricist)
- Semantic/model-theoretic/structuralist analysis
- Pragmatic

Replacement

The analogue of elimination on the epistemological side would be replacement. Our prior ways of describing and conceptualizing the world might drop out of use and be superseded by newer more adequate ways of representing reality. For example, many of our folk psychological concepts might turn out not to do a good job of characterizing the aspects of the world at which they were directed, as happened with such concepts as demonic possession (Feyerabend, 1962).

Theoretical-derivational

The classic notion of intertheoretic reduction in terms of theoretical derivation, as found in Kemeny and Oppenheim (1956) or in Ernest Nagel's classic treatment (1961), descends from the logical empiricist view of theories as interpreted formal calculi statable as sets of propositions of symbolic logic. Intertheoretic reduction is the derivation of one theory from another; and so constitutes an *explanation* of the reduced theory by the reducing theory. This model treats intertheoretic reduction as deductive, and as a special case of deductive-nomological explanation. Thus if one such theory T_1 could be logically derived from another T_2, then everything T_1 says about the world would be captured by T_2. Because the theory to be reduced T_1 normally contains terms and predicates that do not occur in the reducing theory T_2, the derivation also requires some bridge laws or bridge principles to connect the vocabularies of the two theories. These may take the form of strict biconditionals linking terms in the two theories, and when they do such biconditionals may underwrite an ontological identity claim. However, the relevant bridge principles need not be strict biconditionals. All that is required is enough of a link between the vocabularies of the two theories to support the necessary derivation.

One caveat is in order. Strictly speaking, in most cases what is derived is not the original reduced theory but an image of that theory within the reducing theory, and that image is typically only a close approximation of the original rather than a precise analogue (Feyerabend, 1977; Churchland, 1985).

Nagel's account (1961) of intertheoretic reduction has become a standard for this type, and all alternative accounts are in one way or another amendments to it or reactions against it. So, let us look at it a little more closely, and see how problems for this account have arisen. Nagel distinguishes two types of reductions on the basis of whether or not the vocabulary of the reduced theory is a subset of the reducing theory. If it is – that is, if the reduced theory T_1 contains no descriptive terms not contained in the reducing theory T_2, and the terms of T_1 are understood to have approximately the same meanings that they have in T_2, then Nagel calls the reduction of T_1 by T_2 "homogeneous" (Nagel, 1961, p. 339).

From a historical perspective, this attitude is somewhat naïve (Sklar, 1967, pp. 110–11). The number of actual cases in the history of science where a genuine homogeneous reduction takes place are few and far between. One escape for the proponent of Nagel-type reductions is to distinguish explaining a theory (or explaining the laws of a given theory) from explaining it away (Sklar, 1967, pp. 112–13). Thus, we may still speak of reduction if the derivation of the approximations to the reduced theory's laws serves to account for why the reduced theory works as well as it does in its (perhaps more limited) domain of applicability.

The task of characterizing reduction is more involved when the reduction is heterogeneous, that is, when the reduced theory contains terms or concepts that do not appear in the reducing theory. Nagel takes as a paradigm example the (apparent) reduction of thermodynamics, or at least some parts of thermodynamics, to

statistical mechanics. For instance, thermodynamics contains the concept of temperature (among others) that is lacking in the reducing theory of statistical mechanics. Nagel notes that "if the laws of the secondary science (the reduced theory) contain terms that do not occur in the theoretical assumptions of the primary discipline (the reducing theory) the logical derivation of the former from the latter is *prima facie* impossible" (Nagel, 1961, pp. 352–4). As a consequence, Nagel introduces two "necessary formal conditions" required for reduction to take place known as *connectability* and *derivability*. Connectability has to do with the bridge laws that relate the sets of terms from the theories in question. The consideration of certain examples lends plausibility to the idea that the bridge laws should be considered to express some kind of identity relation. For instance, Sklar notes that the reduction of the "theory" of physical optics to the theory of electromagnetic radiation proceeds by *identifying* one class of entities – light waves – with (part of) another class – electromagnetic radiation (Sklar, 1967, p. 120). In fact, if something like Nagelian reduction is going to work, it is generally accepted that the bridge laws should reflect the existence of some kind of synthetic identity.

One problem facing the theoretical-derivational account of intertheoretic reduction was forcefully presented by Feyerabend in "Explanation, Reduction, and Empiricism" (Feyerabend, 1962). Consider the term "temperature" as it functions in classical thermodynamics. This term is defined in terms of Carnot cycles and is related to the strict, nonstatistical zeroth law as it appears in that theory. The so-called reduction of classical thermodynamics to statistical mechanics, however, fails to identify or associate *nonstatistical* features in the reducing theory, statistical mechanics, with the nonstatistical concept of temperature as it appears in the reduced theory. How can one have a genuine reduction, if terms with their meanings fixed by the role they play in the reduced theory are identified with terms having entirely different meanings? Classical thermodynamics is not a statistical theory. The very possibility of finding a reduction function or bridge law that captures the concept of temperature and the strict, nonstatistical role it plays in the thermodynamics seems impossible (Takesaki, 1970; Primas, 1998).

Many physicists, now, would accept the idea that our concept of temperature and our conception of other exact terms that appear in classical thermodynamics such as "entropy," need to be reformulated in light of the alleged reduction to statistical mechanics. Textbooks, in fact, typically speak of the theory of "statistical thermodynamics."

Because of the problem mentioned above, as well as others, many philosophers of science felt that the theoretical-derivational model (Nagel, 1961) did not realistically capture the actual process of intertheoretic reduction. As Primas puts it, "there exists not a single physically well-founded and nontrivial example for theory reduction in the sense of Nagel (1961). The link between fundamental and higher-level theories is far more complex than presumed by most philosophers" (1998, p. 83). Therefore, alternative models of intertheoretic reduction abandon one or more *ontological assumptions* made by the theoretical-derivational account (i.e., the logical empiricist account):

1 Property/kind cross-theoretic (ontological) identities are to be determined solely by formal criteria such as successful intertheoretic reduction, e.g., smooth intertheoretic reduction is both necessary and sufficient for cross-theoretic identity.
2 Realism, scientific theories are more than mere "computational devices."

and/or one or more *epistemological assumptions*:

1 Philosophy of science is prescriptive rather than descriptive, e.g., philosophy of science should seek a grand, universal account of intertheoretic reduction.
2 Scientific theories are axiomatic systems.
3 Reduction = logical deduction, or at least deduction of a structure specified within the vocabulary and framework of the reduced theory or some corrected version of it.
4 Necessity of bridge laws or some other equally strong cross-theoretic connecting principles to establish synthetic identities.
5 Symbolic logic is the appropriate formalism for constructing scientific theories.
6 Scientific theories are linguistic entities.
7 Hardcore explanatory unification. Reduction is proof of displacement (in principle) showing that the more comprehensive reducing theory contains explanatory and predictive resources equaling or exceeding those of the reduced theory.
8 Intertheoretic reductions are an all or nothing *synchronic* affair as in the case of "microreductions" (Oppenheim and Putnam, 1958; Causey, 1977): the lower-level theory and its ontology reduce the higher-level theory and its ontology. Ontological levels are mapped one-to-one onto levels of theory which are mapped one-to-one onto fields of science.
9 The architecture of science is a layered edifice of analytical levels (Wimsatt, 1976).

Alternatives to the Nagel (1961) model are deemed more or less radical (by comparison) depending on which of the preceding tenets are abandoned. On the more conservative side, many alternative accounts of intertheoretic reduction merely modify (3) by moving to logico-mathematical deduction, but reject (4). For example, the requirement of bridge laws gets replaced by notions such as: "analog relation" – an ordered pair of terms from each theory (Hooker, 1981; Bickle, 1998), "complex mimicry" (Paul Churchland, 1989) or "equipotent image" (Patricia Churchland, 1986), to name a few. Many of these comparatively conservative accounts also reject (8), preferring to talk about a range of reductions, from replacement on one end of the continuum to identity on the other. More radical alternatives to the Nagel models are as follows.

Semantic/model-theoretic/structuralist analysis

This approach (the "semantic" approach for short), is regarded by some as comparatively radical because it rejects the conception of scientific theories as formal calculi formalizable in first-order logic and (partially) interpretable by connecting principles such as bridge laws. The semantic approach makes the following assumptions:

(i) Scientific theories are not essentially linguistic entities (sets of sentences), but are terms or families of their *mathematical models or mathematical structures.*
(ii) The formal explication of the structure of scientific theories is not properly carried out with first-order logic and metamathematics, but with *mathematics*, though the choice of mathematical formalisms will differ depending on who you read (Giere, 1988; Bickle, 1998; Batterman, 2000).

The semantic approach minimally rejects epistemological assumptions (2)–(6) and (8), i.e., rejects the derivation of laws and abandons truth preservation (everything the reduced theory asserts is also asserted by the reducing theory). On the semantic approach, the reduction relation might be conceived of as some kind of "isomorphism" or "expressive equivalence" between models (Bickle, 1998). However, as we shall shortly see, more radical versions of the semantic approach reject all the preceding epistemological assumptions held by the logical empiricist account of intertheoretic reduction.

Pragmatic

Success in real world representation is, in large part, a practical matter of whether and how fully one's attempted representation provides *practical causal* and *epistemic access* to the intended representational target. A good theory or model succeeds as a representation if it affords reliable avenues for *predicting*, *manipulating* and *causally* interacting with the items it aims to represent. It is the practical access that the model affords in its context of application that justifies viewing it as having the representational content that it does (Van Fraassen, 1989; Kitcher, 1989). If a lower-level theory about a specific domain provides superior *real-world explanatory* and *predictive* value compared to a higher-level theory representing the same domain, then the lower-level theory has met the ultimate test of successful intertheoretic reduction. Note that this contextual, pragmatic account of intertheoretic reduction is also highly *particularist*; it advocates adjudicating on a case-by-case basis; no universal theory of reduction is sought. This account rejects at least assumptions (1)–(6) in the epistemological category, and assumption (1) in

the ontological category (Patricia Churchland, 1986). More radical versions reject all nine of the preceding epistemological assumptions.

Whereas the theoretical-derivational account (i.e., the logical empiricist account) of intertheoretic reduction (and its variants) only makes sense if you presuppose nomological and mereological supervenience; *in principle*, both the semantic and the pragmatic accounts of intertheoretic reduction are compatible with the failure of mereological supervenience and perhaps even nomological supervenience. We shall encounter specific versions of such accounts of reduction shortly.

While there are certainly mutually exclusive and competing accounts of intertheoretic reduction that represent each of our four *types*, there is no principled reason why the four types could not be synthesized into a single account. Schaffner's "generalized replacement-reduction" (GRR) model of intertheoretic reduction is one such attempt (Schaffner, 1992, 1998, 2000).

Though much more could be said about the many varieties of ontological and epistemological reduction and their respective faults and merits, the main versions may be graphically summarized (figure 5.1):

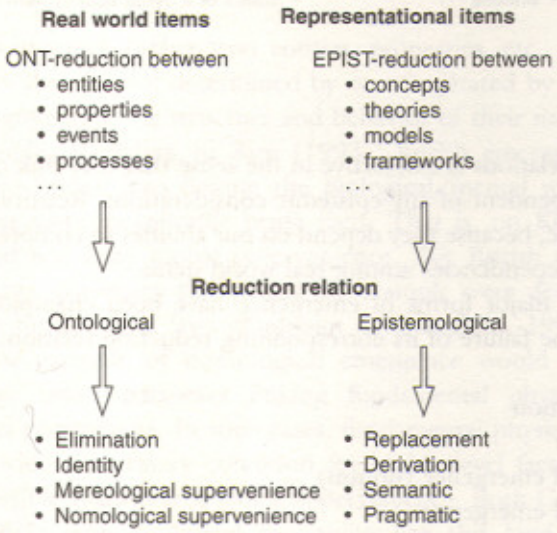


Figure 5.1

The Varieties of Emergence: Ontological and Epistemological

Emergence, like reduction, is interpreted in diverse ways (Silberstein and McGeever, 1999). Again, my aim is to survey the main variants.

The basic idea of emergence is roughly the converse of reduction. Though the

emergent features of a whole or complex are not completely independent of those of its parts since they "emerge from" those parts, the notion of emergence nonetheless implies that, in some significant way, they *go beyond* the features of those parts. There are many senses in which a system's features might be said to emerge, some of which are relatively modest (Rueger, 2000a,b; Batterman, 2000; Bedau, 1997) and others which are more controversial (Humphreys, 1997; Silberstein, 1998).

The varieties of emergence can be divided into several groups along lines similar to those divisions between the types of reduction (figure 5.2).

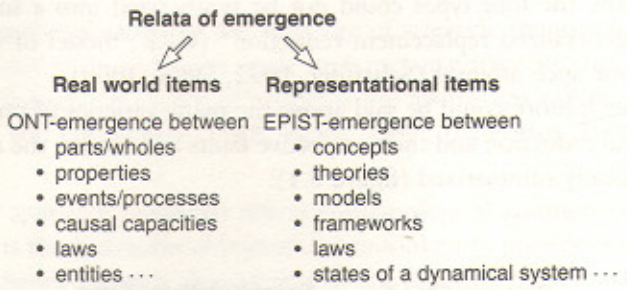


Figure 5.2

Ontological relations are objective in the sense that they link ontic items, e.g., properties, independent of any epistemic considerations. Relations of the second sort are epistemic, because they depend on our abilities to comprehend the nature of the links or dependencies among real world items.

At least four major forms of emergence have been championed; each is an elaboration of the failure of its corresponding reduction relation:

- Non-elimination
- Non-identity
- Mereological emergence (holism)
- Nomological emergence

Non-elimination

If a property, entity, causal capacity, kind or type cannot be eliminated from our ontology, then one must be a realist about said item. Obviously, this leaves open the question of what the criteria ought to be for non-elimination in any given case; but they will almost certainly be epistemological/explanatory in nature.

Non-identity

If a property, type or a kind cannot be ultimately identified with a physical (or lower-level) property, type or kind then one must accept that said item is a distinct non-physical (or higher-level) property, type or kind. Again, this leaves open the criteria for non-identifiability and again, such criteria are generally epistemological/explanatory in nature.

Mereological emergence (holism)

These are cases in which objects have properties that are not determined by the intrinsic (non-relational) physical properties of their most basic physical parts. Or, cases in which objects are not even wholly composed of basic (physical) parts at all. British (classical) emergentism held that mereological emergence is true of chemical, biological and mental phenomena (McLaughlin, 1992).

Nomological emergence

These are cases in which higher-level entities, properties, etc., are governed by higher-level laws that are not determined by or necessitated by the fundamental laws of physics governing the structure and behavior of their most basic physical parts. For example, according to Kim (1993), British emergentism held that while there were bridge laws linking the biological/mental with the physical, such bridge laws were inexplicable brute facts. That is, on Kim's view British emergentism did not deny global supervenience. But British emergentism did deny that the laws governing the mental for example were determined (or explained) by the fundamental laws of physics (McLaughlin, 1992; Kim, 1993). A more extreme example of nomological emergence would be where there were no bridge laws whatsoever linking fundamental physical phenomena with higher-level phenomena. In such cases, fundamental physical facts and laws would only provide a necessary condition for higher-level facts and laws. This would imply possible violations of global supervenience. Both Cartwright (1999) and Dupré (1993) *seem* to defend something like this kind of nomological emergence. An even more extreme example is found in cases in which either fundamental physical phenomena or higher-level phenomena are not law-governed at all. This would amount to eliminativism or antirealism regarding nomological or physical necessity; see Van Fraassen (1989) for a defense of this view. It is important to note that in all cases of *nomological emergence*, it is *in principle* impossible to derive or predict the higher-level phenomena on the basis of the lower-level phenomena.

The *epistemological link* must describe how things are related such that one

epistemologically emerges from another. At least two major views have been championed:

- Predictive/explanatory emergence
- Representational/cognitive emergence

Predictive/explanatory emergence

Wholes (systems) have features that cannot *in practice* be explained or predicted from the features of their parts, their mode of combination, and the laws governing their behavior. In short, X bears predictive/explanatory emergence with respect to Υ if Υ cannot (reductively) predict/explain X. More specifically, in terms of types of intertheoretic reduction, X bears predictive/explanatory emergence with respect to Υ: if Υ cannot *replace* X, if X cannot be *derived* from Υ, or if Υ cannot be shown to be *isomorphic* to X. A lower-level theory Υ (description, regularity, model, schema, etc.), for purely epistemological reasons (conceptual, cognitive or computational limits), can fail to predict or explain a higher-level theory X. If X is predictive/explanatory emergent with respect to Υ for *all possible cognizers in practice*, then we might say that X is *incommensurable* with respect to Υ. A paradigmatic and notorious example of predictive/explanatory emergence is chaotic, non-linear dynamical systems (Silberstein and McGeever, 1999). The emergence in chaotic systems (or models of non-linear systems exhibiting chaos) follows from their sensitivity to initial conditions, plus the fact that physical properties can only be specified to finite precision; infinite precision would be necessary to perform the required "reduction", given said sensitivity. It does not follow, however, that chaotic systems provide evidence of violations of mereological supervenience or nomological supervenience (Kellert, 1993, pp. 62, 90), e.g., dynamical systems have attractors as high-level emergent features only in the sense that you cannot deduce them from equations for the system. McGinn (1999) and other mysterians hold that folk psychology is predictive/explanatory emergent with respect to the theories of neuroscience.

Representational/cognitive emergence

Wholes (systems) exhibit features, patterns or regularities that cannot be fully represented (understood) using the theoretical and representational resources adequate for describing and understanding the features and regularities of their more basic parts and the relations between those more basic parts. X bears representational/cognitive emergence with respect to Υ, if X does *not* bear predictive/explanatory emergence with respect to Υ, but nonetheless X represents higher-level patterns or non-analytically guaranteed regularities that cannot be

fully, properly or easily represented or understood from the perspective of the lower-level Υ. As long as X retains a significant *pragmatic* advantage over Υ with respect to understanding the phenomena in question, then X is representational/cognitive emergent with respect to Υ. Nonreductive physicalism holds that folk psychology is representational/cognitive emergent with respect to the theories of neuroscience (Antony, 1999).

The Reduction and Emergence Debate Today: Specific Cases Seeming to Warrant the Label of Ontological or Epistemological Emergence

Not since the first half of the twentieth century have emergence and reduction enjoyed so much critical attention. Claims involving emergence are now rife in discussions of philosophy of mind, philosophy of physics, various branches of physics itself including quantum mechanics, condensed matter theory, nonlinear dynamical systems theory (especially so-called chaos theory), cognitive-neuroscience (including connectionist/neural network modeling and consciousness studies) and so-called complexity studies (Silberstein and McGeever, 1999). To quote Kim:

> we are now seeing an increasing and unapologetic use of expressions like "emergent," "emergent property," and "emergent phenomenon" . . . not only in serious philosophical literature but in the writings in psychology, cognitive science, systems theory, and the like (1998, pp. 8–9).

Kim also says that

> the return of emergentism is seldom noticed, and much less openly celebrated; it is clear, however, that the fortunes of reductionism correlate inversely with those of emergentism . . . It is no undue exaggeration to say that we have been under the reign of emergentism since the early 1970s (1999, p. 5).

There are two primary reasons for the return of emergentism. First, regarding nomological emergence, a growing body of literature focusing on actual scientific practice suggests that there really are not many cases of successful intertheoretic reduction in the empiricist tradition of *demonstrating* nomological supervenience.

> Our scientific understanding of the world is a patchwork of vast scope; it covers the intricate chemistry of life, the sociology of animal communities, the gigantic wheeling galaxies, and the dances of elusive elementary particles. But it is a patchwork nevertheless, and the different areas do not fit well together (Berry, 2000, p. 3).

Focus on actual scientific practice suggests that either there really are not many cases of successful epistemological (intertheoretic) reduction or that most philosophical accounts of reduction bear little relevance to the way reduction in science actually works. Most working scientists would probably opt for the latter claim.

Often discussed cases of failed or incomplete intertheoretic reduction in the literature include:

1 the reduction of thermodynamics to statistical mechanics (Primas, 1991, 1998; Sklar, 1999)
2 the reduction of thermodynamics/statistical mechanics to quantum mechanics (Hellman, 1999)
3 the reduction of chemistry to quantum mechanics (Cartwright, 1997; Primas, 1983)
4 the reduction of classical mechanics to quantum mechanics (such as the worry that quantum mechanics cannot recover classical chaos) (Belot and Earman, 1997).

Take the case of chemistry and its alleged reduction to quantum mechanics. Currently chemists do not use fundamental quantum mechanics (Hamiltonians and Schrödinger's equation) to do their science. Quantum chemistry cannot be deduced directly from Schrödinger's equation due to multiple factors that include the many-body problem (Hendry, 1998). Quantum mechanical wave functions are not well-suited to represent chemical systems or support key inferences essential to chemistry (Woody, 2000). It is still an open question as to whether quantum mechanics can describe or represent a molecule (Berry, 2000). Indeed, little of current chemistry can be represented by pure quantum mechanical calculations (Primas, 1983; Scerri, 1994; Ramsey, 1997). Chemistry uses idealized models whose relationship to fundamental quantum mechanics is questionable (Primas, 1983; Hendry, 1999). As Cartwright (1997, p. 163) puts it:

> Notoriously, we have nothing like a real reduction of the relevant bits of physical chemistry to physics – whether quantum or classical. Quantum mechanics is important for explaining aspects of chemical phenomena but always quantum concepts are used alongside of sui generis – that is, unreduced-concepts from other fields. They do not explain the phenomena on their own.

Another well-known example is the case of thermodynamics and statistical mechanics. First, there is a variety of distinct concepts of both temperature and entropy that figure in both statistical mechanics and classical thermodynamics. Second, thermodynamics can be applied to a number of very differently constituted microphysical systems. Thermodynamics can be applied to gases, electromagnetic radiation, magnets, chemical reactions, star clusters and black holes. As Sklar (1993, p. 334) puts it:

> The alleged reduction of thermodynamics to statistical mechanics is another one of those cases where the more you explore the details of what actually goes on, the more convinced you become that no simple, general account of reduction can do justice to all the special cases in mind.

Third, the status of the probability assumptions that are required to recover thermodynamic's principles within statistical mechanics are themselves problematic or *ad hoc*. For example, the assumption that the micro-canonical ensemble is to be assigned the standard, invariant, probability distribution. Fourth, perhaps the thorniest problem of all, statistical mechanics is time symmetric and thermodynamics possesses time asymmetry.

These are especially important examples because they involve difficulties between different levels of explanation *within physical science*. Some of the four (e.g., the reduction of thermodynamics to statistical mechanics) were once thought of as successes for philosophical accounts of intertheoretic reduction (Sklar, 1993).

Perhaps the most highly advertised case of failed intertheoretic reduction is the attempt to reduce folk psychology to theories of neuroscience. Presently a popular *ontological* version of the mind/body problem goes by the name of "the hard problem of phenomenal consciousness": how and why are brain states conscious? (Chalmers, 1996). As Kim (1998, pp. 102–3) puts it:

> We are not capable of designing, through theoretical reasoning, a wholly new kind of structure that we can predict will be conscious; I don't think we even know how to begin, or indeed how to measure our success . . . In any case it seems to me that if emergentism is correct about anything, it is more likely to be correct about qualia than about anything else.

For more on the problems of phenomenal consciousness and emergence see Silberstein (2001).

In this spirit, philosophers of science and mind have made a cottage industry of collecting many of the cases of incomplete intertheoretic reduction, calling them all "emergence"; see, for example, *Special Issue: Reduction and Emergence*, *Philosophical Studies*, 95 (1–2), August 1999 and Beckermann et al. (1992). The essays in both volumes span psychology, biology and physics. Each of the essays is an examination of an attempted intertheoretic reduction that is currently having grave difficulties. Taken in toto, these cases seem a barometer of the prospects for unifying the sciences, and therefore *indicative* of the prospects of epistemological and ontological reductionism. There is a movement afoot devoted to arguing this point. The movement is known as the "disunity of science movement" or the "anti-fundamentalism movement" (Dupré, 1993; Cartwright, 1999). However, an *indication* is not an argument, so each case deserves to be examined in its own right.

There is no doubt danger in lumping all these cases together. It is clear, for example, that thermodynamics is predictive/explanatory emergent with respect to

statistical mechanics. As of yet, few are ready to conclude that thermodynamical phenomena are, for example, nomologically or mereologically emergent with respect to statistical mechanical phenomena. By way of contrast, when Kim talks about phenomenal consciousness being emergent, he seems to be making a claim about emergent phenomenal consciousness which goes beyond a function of ignorance interpretation (Kim, 1998, 1999). It is not uncommon for such equivocations on the term "emergence" to appear in the same volume.

This brings me to the second major reason for the return of emergence. There are some people who allege that quantum mechanics *itself* provides examples of mereological emergence:

> In quantum theory, then, the physical state of a complex whole cannot always be reduced to those of its parts, or to those of its parts together with their spatiotemporal relations, even when the parts inhabit distinct regions of space. Modern science, and modern physics in particular, can hardly be accused of holding reductionism as a central premise, given that the result of the most intensive scientific investigations in history is a theory that contains an ineliminable holism (Maudlin, 1998, p. 55).

By and large, a system in classical physics can be analyzed into parts, whose states and properties determine those of the whole they compose. But the state of a system in quantum mechanics resists such analysis. The quantum state of a system gives a specification of its probabilistic dispositions to display various properties on its measurement. Quantum mechanics' most complete such specification is given by what is called a pure state. Even when a compound system has a pure state, its subsystems generally do not have their own pure states. Schrödinger, emphasizing this characteristic of quantum mechanics, described such component subsystems as "entangled." Such entanglement of systems demonstrates nonseparability – the state of the whole is not constituted by the states of its parts. State assignments in quantum mechanics have been taken to violate state separability in two ways: the subsystems may simply not be assigned any pure states of their own, or else the states they are assigned may fail to completely determine the state of the system they compose.

The quantum state of a system may be either pure or mixed. A pure state is represented by a vector in the system's Hilbert space. It is commonly understood that any entangled quantum systems violate state separability in so far as the vector representing the state of the system they compose does not factorize into a vector in the Hilbert space of each individual subsystem that could be taken to represent its pure state. A set of entangled quantum systems compose a system whose quantum state is represented quantum mechanically by a tensor-product state-vector which does not factorize into a vector in the Hilbert space of each individual system:

$$\Psi_{1,2,\ldots,R} \neq \Psi_1 \otimes \Psi_2 \otimes \cdots \otimes \Psi_R$$

Now in such a case each subsystem $1, 2, \ldots, n$ may be uniquely assigned what is called a mixed state (represented in its Hilbert space not by a vector but by a

so-called von Neumann density operator). But then state separability fails for a different reason: the subsystem mixed states do not uniquely determine the compound system's state.

On the basis of nonseparability, many people have argued that quantum mechanics provides us with examples of systems that have properties that do not always reduce to the intrinsic properties of the most basic parts, i.e., quantum mechanical systems exhibit mereological emergence (Healey, 1991; Hawthorne and Silberstein, 1995; Humphreys, 1997). Such entangled systems appear to have novel properties of their own. Quantum systems that are in superpositions of possible states are behaviorally distinct from systems that are in mixtures of these states and individual systems can be become entangled and thus form a new unified system which is not the sum of its intrinsic parts. From this, some further infer that: "the state of the compound [quantum] system determines the state of the constituents, but not vice versa. This last fact is exactly the reverse of what [mereological] supervenience requires" (Humphreys, 1997, p. 16). The opinion of a growing number of philosophers of physics is expressed by Maudlin (1998, pp. 58–60):

> Quantum holism ought to give some metaphysicians pause. As has already been noted, one popular "Humean" thesis holds that all global matters of fact supervene on local matters of fact, thus allowing a certain ontological parsimony. Once the local facts have been determined, all one needs to do is distribute them throughout all of space-time to generate a complete physical universe. Quantum holism suggests that our world just doesn't work like that. The whole has physical states that are not determined by, or derivable from, the states of the parts. Indeed, in many cases, the parts fail to have physical states at all. The world is not just a set of separately existing localized objects, externally related only by space and time. Something deeper, and more mysterious, knits together the fabric of the world. We have only just come to the moment in the development of physics that we can begin to contemplate what that might be.

At any rate, quantum nonseparability is not restricted to settings such as twin-slit experiments and EPR (non-locality) experiments. Superpositions and entangled states are required to explain certain chemical and physical phenomena such as phase transitions that give rise to superconductivity, superfluidity, paramagnetism, ferromagnetism; see Anderson (1994), Auyang (1998) and Cornell and Wieman (1998).

Some interpretations of quantum mechanics such as Bohr (1934) and Bohm and Hiley (1993) imply mereological emergence (holism) with respect to *entities*: there are physical objects that are not wholly composed of basic (physical) parts. On Bohr's interpretation one can meaningfully ascribe properties such as position or momentum to a quantum system only in the context of some well-defined experimental arrangement suitable for measuring the corresponding property. Although a quantum system is purely physical on this view, it is not composed of distinct happenings involving independently characterizable physical objects such

as the quantum system on the one hand, and the classical apparatus on the other. On Bohm's interpretation, it is not just quantum object and apparatus that are holistically connected, but any collection of quantum objects by themselves constitute an indivisible whole. A complete specification of the state of the "undivided universe" requires not only a listing of all its constituent particles and their positions, but also of a field associated with the wave-function that guides their trajectories. If one assumes that the basic physical parts of the universe are just the particles it contains, then this establishes ontological holism in the context of Bohm's interpretation.

For the purposes of this discussion, what is most important is not whether or not quantum mechanics actually does provide cases of mereological emergence, but that the belief that it does, in part, fuels emergentism. Though it must be said, there are some philosophers who are still skeptical about the reality, coherence or importance of quantum holism (Lewis, 1986; Dickson, 1998). Not everyone acknowledges that nonseparability implies mereological emergence. For example, Healey argues that whether or not nonseparability implies mereological emergence is a matter of interpretation (1989, pp. 142–5). Healey's own modal interpretation (1989) does imply mereological emergence, however he stipulates that the formalism of quantum mechanics is open to interpretations that do not. He argues (Healey, 1991) that nonseparability in general and so-called non-locality are best explained by positing mereological emergence.

Questions for Future Research

Recall that the best reason for believing in reductionism is an acceptance of mereological and/or nomological supervenience based in large part on successful intertheoretic reduction (or epistemological reduction). Do the preceding examples of epistemological and ontological emergence indicate emergentism is true? At this juncture, may we even say whether emergentism or reductionism is more probable? What does the current state of disunity within any given science and across the various sciences imply about emergence? Regarding the ultimate fate of mereological and nomological emergence respectively, there are two general possibilities. Either these respective forms of emergence are merely a function of our ignorance or they are real facts about the world. If they are real facts about the world then they may be either universally true or restricted to a particular domain such as microphysics. Of course, the ultimate fate of mereological emergence might be different from that of nomological emergence and vice-versa. For example, the possibilities for nomological emergence are as follows:

There are four *reductive outcomes*:

1 Any claimed emergence is due to philosophical ignorance. A better, more appropriate *philosophical* theory of intertheoretic reduction needs to be con-

structed that will show that the lower-level theory does reductively explain the higher-level theory in question. It is possible (if not probable) that different cases will require different accounts of intertheoretic reduction for their resolution.

2 Any case of emergence is due to *empirical* or *experimental* ignorance. Future discoveries will allow us to see how the lower-level theory does in fact reductively explain the higher-level theory in question.

3 Any claim to emergence relies on lower-level theories that are false or incomplete, and such theories will be replaced or supplemented by correct lower-level theories in order to reductively explain the higher-level theory.

Outcomes 1–3 would all be unqualified wins for epistemological reductionism if not ontological reductionism.

4 The higher-level theory will cease to be predictive/explanatory emergent with respect to the lower-level theory, but for some (indeterminate) length of time the higher-level theory will be representational/cognitive emergent with respect to the lower-level theory.

This is more or less a win for epistemological (if not ontological) reductionism. There are then two *emergent outcomes*:

5 The higher-level theory is predictive/explanatory emergent with respect to the lower-level theory and for whatever reason, due to whatever *epistemological limits*, the lower-level theory and its successors will never be able to reductively explain the higher-level theory. This is a win for epistemological emergence only.

6 The higher-level theory is predictive/explanatory emergent with respect to the lower-level theory (and its successors) *because* the phenomena/laws represented by the higher-level theory are *nomologically emergent* with respect to the phenomena/laws represented by the lower-level theory. The lower-level phenomena only provide a *necessary (but not sufficient) condition* for the *emergence* of the higher-level phenomena. This would be an unqualified loss for *both* epistemological and ontological reductionism.

One important question for the future is to determine, in each specific instance of incomplete intertheoretic reduction (such as the cases discussed earlier), which of these six possibilities actually obtains. However it should be clear that emergentism and reductionism might form a *continuum* and not a dichotomy. This is true in several respects. First, even if mereological emergence is real it does not necessarily imply nomological emergence. Even if the quantum is mereologically emergent, it could still be the case that all higher-level phenomena nomologically supervenes upon it. Second, both mereological and nomological emergence might be restricted to certain domains. For example, mereological emergence might be

limited to the quantum and nomological emergence limited to the mental. Third, for any given case we can always divide the question for ontological and epistemological emergence. Or more generally, it could turn out, for example, that epistemological emergence is inescapable while ontological emergence is rare or nonexistent. Of course, given the former, it is an open question how we would ever discover the latter.

Recent accounts of *intertheoretic reduction*, the more radical versions of the semantic and pragmatic models mentioned earlier, such as *GRR* (Schaffner, 1992, 1998) and the more explicitly pragmatic and ontic *causal mechanical* model (Machamer et al., 2000), explicitly reject microreduction, in part because of the problematic cases mentioned earlier. Such alternative accounts of intertheoretic reduction, in their rejection of microreduction, explicitly acknowledge the continuum between reduction and emergence. For example, the causal mechanical model of intertheoretic reduction focuses on explanations as characterizing complex (nested and inter-connected) causal mechanisms and pathways, such as we find in molecular biology and neuroscience. The emphasis in this model is on causal/mechanical processes as opposed to nomological patterns of explanation. More importantly for our purposes, this model admits of *multilevel* descriptions of causal mechanisms that mix different levels of aggregation from cell to organ back to molecule.

Take the following example from behavioral genetics:

> there is no *simple* [reductive] explanatory model for behavior even in simple organisms. What *C. elegans* [a simple worm] presents us with is a tangled network of influences [causal mechanisms] at genetic, biochemical, intracellular, neuronal, muscle cell, and environmental levels (Schaffner, 1998, p. 237).

This kind of reductive explanation focuses on interlevel causal processes and emphasizes the limits and rarity of logical empiricist accounts of intertheoretic reduction. This approach to reduction is diachronic, emphasizing the gradual, partial and fragmentary nature of many real world cases. This model clearly views intertheoretic reduction as a continuum and not a dichotomy.

One can also find similar web-like and bushy cases of intertheoretic reduction within physics. For example, cases in which two domains (such as quantum mechanics and chemistry) are related by an asymptotic series often require appeal to an intermediate theory (Berry, 1994; Primas, 1998; Batterman, 2000, 2001). In the asymptotic borderlands between such theories, phenomena emerge that are not fully explainable in terms of either the lower-level or the higher-level theory, but require both theories or an intermediary (Batterman, 2000, 2001). Examples of this phenomena can be found in the borders between: quantum mechanics and chemistry, as well as thermodynamics and statistical mechanics (Berry and Howls, 1993; Berry, 1994; 2000; Batterman, 2000). Batterman speaks of the "asymptotic emergence of the upper level properties" in such cases, and he goes on to suggest that "it may be best, in this context, to give up on the various philosophical models

of reduction which require the connection of kind predicates in the reduced theory with kind predicates in the reducing theory. Perhaps a more fruitful approach is to investigate asymptotic relations between the different theory pairs. Such asymptotic methods often allow for the understanding of emergent structures which dominate observably repeatable behavior in the limiting domain between the theory pairs" (2000, pp. 136–7).

Intertheoretic reduction *à la* singular asymptotic expansions is not easy to characterize, though it is fair to say that it falls within the semantic approach to intertheoretic reduction. Examples of intertheoretic relations involving singular asymptotic expansions include: Maxwell's electrodynamics and geometrical optics; molecular chemistry and quantum mechanics and; classical mechanics and quantum mechanics (Primas, 1998; Berry, 2000).

There are several things worth noticing about both the preceding models of intertheoretic reduction. Such reductions are not universally valid, they can only be considered on a case-by-case basis. Such reductions require specification of context, the new description or higher-level theory cannot be derived from the lower-level theory. Indeed, such reductions generally start with the higher-level theory/context and work back to the more fundamental theory (Berry, 1994). The lower-level theory (the reducing theory) is not, as a rule, more powerful or universal in its predictive/explanatory value than the higher-level theory (the reduced theory). Indeed, the new ontology and topology generated by the higher-level description cannot be replaced or eliminated precisely because of its more universal explanatory power; and the intertheoretic reductions on such accounts show why this must be the case. Contrary to the standard view, failure of reduction need not imply failure of explanation. A more fundamental theory can explain a higher-level theory ("from below" as it were) without providing a reduction of that theory in the standard senses of the term. Emergent phenomena need not be inexplicable brute facts contrary to classical emergentism. Given such accounts of intertheoretic reduction, there is good reason to think that contra the dreams of the unity of science movement, that unification of scientific theories will be local at best.

Such alternative accounts of intertheoretic reduction suggest that the relationship between "higher-level" and "lower-level" scientific theories is a nested hierarchy as opposed to a pyramid structure. And if we think such accounts of reduction reflect the actual ontology of the world, they suggest that the relationship between the various "levels" (subatomic, atomic, molecular, etc.) is also a nested hierarchy. An even more radical speculation along these lines is that the relationship between higher-level and lower-level scientific theories as well as between the various ontic "levels" themselves looks more like non-Boolean lattices (Primas, 1991). The various domains will have overlapping areas or unions, but they will not be co-extensional. So properties in one domain may be necessary for properties in another domain to emerge, but not sufficient. Such alternative accounts of intertheoretic reduction do not obviously imply or demand either mereological or nomological supervenience.

Humphreys suggests (1997) if there were widespread mereological emergence or nonseparability then lower-level property instances would often "merge" in the formation of higher-level properties such that they no longer exist as separate subvenient entities. Widespread mereological emergence calls into question the very picture of reality as divided into a "discrete hierarchy of levels"; rather it is

> more likely that even if the ordering on the complexity of structures ranging from those of elementary physics to those of astrophysics and neurophysiology is discrete, the interactions between such structures will be so entangled that any separation into levels will be quite arbitrary (Humphreys, 1997, p. 15).

Given widespread mereological emergence, the standard divisions and hierarchies between phenomena that are considered fundamental and emergent, simple and aggregate, kinematic and dynamic, and perhaps even what is considered physical, biological and mental are redrawn and redefined. Such divisions will be dependent on what question is being put to nature and what scale of phenomena is being probed.

But on the face of it, one can embrace these alternative models of intertheoretic reduction while maintaining that all apparent emergence is just a function of ignorance. For example, Schaffner strongly suggests that nothing about such tangled causal processes warrants any claims for either mereological emergence or nomological emergence (such as vital or configurational forces). Rather, at worst, such systems provide us with cases of predictive/explanatory emergence or representational/cognitive emergence (Schaffner, 1998, pp. 242–5).

At present, both the emergentist and reductionist feel that, so far, things are going their way. The emergentist points to failures of ontological and methodological reductionism, and the reductionist points to successes. Regarding the problematic cases of intertheoretic reduction, the perennial reductionist reply is to claim that the future will bring success, just as in the past; emergentists likewise feel that they will be redeemed by the future just as they are by the present. This much is true I think, given the examples of both epistemological and ontological emergence canvassed, there is no reason why the burden of proof should continue to lie exclusively with emergentism. At this juncture, neither view is irrational in light of the evidence and neither view is conclusive. Ultimately, emergentists and reductionists are divided by a deeply held philosophical or aesthetic preference that neither will relinquish easily. For example, many philosophers persist in assuming that nomological and mereological reductionism are true in spite of the actual state of unification within science and in spite of the fact that fundamental physics itself might prove a counter-example to mereological supervenience. Do the past successes of reductionism warrant those assumptions on their part or is the assumption based largely on faith?

We know what questions need to be answered to resolve the debate between emergentism and reductionism, but is it possible to ever answer them? How will we know when we have answered them? It is no doubt prudent to remain agnostic

while patiently awaiting the outcome of each "crucial question" for the debate. But unfortunately, not all the problems and questions are empirical. Given that progress on the *ontological* questions of reductionism/emergentism is inextricably bound with progress on the *epistemological* questions of reductionism/emergentism, and vice-versa, there still remains a deeper *conceptual* or *philosophical* problem about how to ultimately adjudicate the evidence at any given point in time. For example, the problem with reducing chemistry to quantum mechanics is not just a computational or calculational one. The explanatory success of chemistry requires both a new *ontology* and a new *topology* (e.g., molecules) beyond that of quantum mechanics (Primas, 1998; Hendry, 1999). Can we therefore conclude that chemical phenomena are ontologically emergent in some important respects? But trying to answer this seemingly straight ontological question will immediately raise the specter of trying to cut the Gordian knot of ontology (e.g., cross-theoretic identities) and epistemology (e.g., intertheoretic reduction). Any answer to the question will require falling back on *philosophical* criteria that are not easily justified. Perhaps the point here is that, in any given case, deciding on the means of intertheoretic reduction (formal or otherwise) and deciding whether or not the attempted reduction is successful (the criteria for successful reduction), is inescapably normative.

For example, is it smooth intertheoretic reduction that motivate and sustain claims of cross-theoretic property identity or the other way around? Likewise, is it the failure of smooth intertheoretic reduction that motivates and sustains claims for failures of cross-theoretic identities, or the other way around? Is there any fact of the matter regarding such questions or are such questions largely normative?

Whichever way we choose, it seems to either lead in circles or raise new and equally hairy problems. If we hold that ontological concerns such as the question of identifying the mental and the physical for example should be completely subordinated to the epistemological question of whether or not the theory of folk psychology can be intertheoretically reduced to some theory of neuroscience, then we need an acceptable and agreed upon account of intertheoretic reduction. As Patricia Churchland puts it, "By making theories the fundamental relata [of the reduction relation], much of the metaphysical bewilderment and dottiness concerning how entities or properties could be reduced simply vanishes" (as quoted in Bickle, 1998, p. 44). But this, of course, brings us back full circle to our problematic cases of intertheoretic reduction. Exactly what we lack at the moment is an acceptable and agreed upon account, method or criteria of intertheoretic reduction in many problematic cases.

Take the case of nonreductive physicalism versus reductive physicalism for example. Both accounts of the mental accept mereological and nomological supervenience, yet the former denies that the mental can be cross-identified with the physical. This is because nonreductive physicalism denies that successful intertheoretic reduction is, in principle or in practice, sufficient for ontological identification of properties (Antony, 1999, pp. 37–43). On this view, mental properties are ontologically distinct while being explicable and predictable in principle from their physical basis. Nonreductive physicalism holds that the identification of one

property with another is not a function of successful intertheoretic reduction, but whether or not the higher-level property figures in patterns or causal relations in non-analytically-guaranteed regularities. An entity/property is ontologically non-identifiable if it participates essentially in regularities that are novel from the point of view of the alleged reducing base – a situation not precluded by successful intertheoretic reduction. Truths discovered that are not true by definition about higher-level properties are irreducible to lower-level truths. As Antony (1999) puts it, nonreductive physicalism is "a non-ontologically-reductive materialism, coupled with an insistence on explanatory reduction" (p. 43). Thus, the only thing that really separates reductive from nonreductive physicalism then, is their respective *philosophical* criterion for identifying one natural kind/property with another; there is no disagreement here about the basic ontological and scientific facts. According to nonreductive physicalism the fact that folk psychology is representational/cognitive emergent with respect to neuroscientfic theories of mind, is sufficient to block the cross-theoretic identity of mental properties with physical properties. According to reductive physicalism on the other hand, if folk psychology can in principle be intertheoretically reduced to some theory of neuroscience then that is sufficient for cross-identification of mental properties with physical properties. The question is this: Is there any objective fact of the matter about who is right in such a dispute? In the long run, it is important to try to separate out the normative from the more empirical aspects of the debate between emergentism and reductionism.

References

Anderson, P. W. (1994): *A Career in Theoretical Physics*. Singapore: World Scientific Publishing.

Antony, L. (1999): "Making Room for the Mental. Comments on Kim's 'Making Sense of Emergence,'" *Philosophical Studies*, 95(2), 37–43.

Auyang, S. (1998): *Foundations of Complex-System Theories*. Cambridge: Cambridge University Press.

Batterman, R. W. (2000): "Multiple Realizability and Universality," *British Journal of the Philosophy of Science*, 51, 115–45.

Batterman, R. W. (2001): *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford: Oxford University Press.

Beckermann, A., Flohr, H. and Kim, J. (eds.) (1992): *Emergence or Reduction? Essays on the Prospects for Nonreductive Physicalism*. Berlin: DeGruyter.

Bedau, M. (1997): "Weak Emergence," in J. E. Tomberlin (ed.), *Philosophical Perspectives (11): Mind, Causation, and World*, Boston: Blackwell, 374–99.

Belot, G. and Earman, J. (1997): "Chaos out of Order: Quantum Mechanics, the Correspondence Principle and Chaos," *Studies in History and Philosophy of Modern Physics*, 2, 147–82.

Berry, M. V. (1991): "Asymptotics, Singularities, and the Reduction of Theories," in D. Prawiz, B. Skyrms and D. Westerståhl (eds.), *Logic, Methodology, and Philosophy of Science IX: Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy*

of Science, Uppsala, Sweden, August 7–14, 1991, volume 134 of *Studies in Logic and Foundations of Mathematics*, Amsterdam: Elsevier Science B. V, 597–607.

Berry, M. V. (1994): "Singularities in Waves and Rays," in R. Balian, M. Kléman and J. P. Poirier (eds.), *Physics of Defects (Les Houches, Session XXXV, 1980)*, Amsterdam: North Holland, 453–543.

Berry, M. V. (2000): "Chaos and the Semiclassical Limit of Quantum Mechanics (is the Moon there when somebody looks?)," *Proceedings of the Vatican Conference on Quantum Mechanics and Quantum Field Theory*, 2–26.

Berry, M. V. and Howls, C. (1993): "Infinity Interpreted," *Physics World*, 12, 35–9.

Bickle, J. (1998): *Psychoneuronal Reduction: The New Wave*. Cambridge, MA: MIT Press.

Blazer, W., Pearce, D. A. and Schmidt, H.-J. (1984): *Reduction in Science: Structure Examples and Philosophical Problems*. Dordrecht: D. Reidel Publishing Company.

Bohm, D. and Hiley, B. J. (1993): *The Undivided Universe*. London: Routledge.

Bohr, N. (1934): *Atomic Theory and the Descriptive of Nature*. Cambridge: Cambridge University Press.

Cartwright, N. (1997): "Why Physics?" in R. Penrose, A. Shimony, N. Cartwright and S. Hawking (eds.), *The Large, the Small and the Human Mind*, Cambridge: Cambridge University Press, 161–8.

Cartwright, N. (1999): *The Dappled World: A Study of the Boundaries of Science*. New York: Cambridge University Press.

Causey, R. L. (1977): *Unity of Science*. Dordrecht: Reidel.

Chalmers, D. (1996): *The Conscious Mind*. Oxford: Oxford University Press.

Churchland, P. M. (1981): "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy*, 78, 67–90.

Churchland, P. M. (1985): "Reduction, Qualia, and the Direct Introspection of Brain States," *Journal of Philosophy*, 82, 8–28.

Churchland, P. M. (1989): *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. Cambridge, MA: MIT Press.

Churchland, P. S. (1986): *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.

Cornell, E. A. and Wieman, C. E. (1998): "The Bose-Einstein Condensate," *Scientific American*, 278(3), 40–5.

Dennett, D. (1988): "Quining qualia," in A. J. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science*, Oxford: Clarendon Press, 240–78.

Dickson, W. M. (1998): *Quantum Chance and Non-Locality*. Cambridge: Cambridge University Press.

Dupré, J. (1993): *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Boston: Harvard University Press.

Feyerabend, P. K. (1962): "Explanation, Reduction and Empiricism," in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science III*. Minnesota: University of Minnesota Press, 231–72.

Feyerabend, P. K. (1977): "Changing Patterns of Reconstruction," *British Journal for the Philosophy of Science*, 28, 351–82.

Giere, R. (1988): *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Hawthorne, J. and Silberstein, M. (1995): "For Whom the Bell Arguments Toll," *Synthese*, 102, 99–138.

Healey, R. A. (1989): *The Philosophy of Quantum Mechanics: An Interactive Interpretation.* Cambridge: Cambridge University Press.

Healey, R. A. (1991): "Holism and Nonseparability," *Journal of Philosophy*, 88, 393–421.

Hellman, G. (1999): "Reduction (?) to What (?) Comments on L. Sklar's 'The Reduction (?) of Thermodynamics to Statistical Mechanics'," *Philosophical Studies*, 95(1–2), 200–13.

Hendry, R. (1998): "Models and Approximations in Quantum Chemistry," in N. Shanks (ed.), *Idealization in Contemporary Physics*, Amsterdam: Rodopi, 123–42.

Hendry, R. (1999): "Molecular Models and the Question of Physicalism," *Hyle*, 5(2), 143–60.

Hooker, C. A. (1981): "Towards a General Theory of Reduction. Part I: Historical and Scientific Setting. Part II: Identity in Reduction. Part III: Cross-Categorical Reduction," *Dialogue*, 20, 38–59, 201–36, 496–529.

Humphreys, P. (1997): "How Properties Emerge," *Philosophy of Science*, 64, 1–17.

Kellert, S. (1993): *In the Wake of Chaos.* Chicago: University of Chicago Press.

Kemeny, J. and Oppenheim, P. (1956): "On Reduction," *Philosophical Studies*, 7, 6–17.

Kim, J. (1993): "Multiple Realization and the Metaphysics of Reduction," in J. Kim (ed.), *Supervenience and Mind*, Cambridge: Cambridge University Press, 309–35.

Kim, J. (1998): *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation.* Cambridge, MA: MIT Press.

Kim, J. (1999): "Making Sense of Emergence," *Philosophical Studies*, 95(1–2), 3–36.

Kitcher, P. (1989): "Explanatory Unification and the Causal Structure of the World," in P. Kitcher and W. C. Salmon (eds.), *Minnesota Studies in the Philosophy of Science, vol. 13, Scientific Explanation*, Minneapolis: University of Minneapolis Press, 410–505.

Lewis, D. (1986): "Causation," in D. Lewis (ed.), *Philosophical Papers, Volume II.* Oxford: Oxford University Press, 159–213.

McGinn, C. (1999): *The Mysterious Flame: Conscious Minds in a Material World.* New York: Basic Books.

McLaughlin, B. (1992): "Rise and Fall of British Emergentism," in A. Beckermann, H. Flohr and J. Kim (eds.), *Emergence or Reduction? Essays on the Prospects for Nonreductive Physicalism*, Berlin: DeGruyter, 30–67.

Machamer, P., Darden, L. and Craver, C. F. (2000): "Thinking About Mechanisms," *Philosophy of Science*, 67, 1–25.

Maudlin, T. (1998): "Part and Whole in Quantum Mechanics," in E. Castellani (ed.), *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, Princeton: Princeton University Press, 46–60.

Nagel, E. (1961): *The Structure of Science.* New York: Harcourt, Brace.

Oppenheim, P. and Putnam, H. (1958): "The Unity of Science as a Working Hypothesis," in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science. Philosophical Studies, vol. 95*, Minneapolis, Minnesota: University of Minnesota, 45–90.

Primas, H. (1983): *Chemistry, Quantum Mechanics, and Reductionism.* Berlin: Springer-Verlag.

Primas, H. (1991): "Reductionism: Palaver Without Precedent," in E. Agazzi (ed.), *The Problem of Reductionsm in Science*, Dordrecht: Kluwer, 161–72.

Primas, H. (1998): "Emergence in the Exact Sciences," *Acta Polytechnica Scandinavica*, 91, 83–98.

Ramsey, J. L. (1997): "Molecular Shape, Reduction, Explanation and Approximate Concepts," *Synthese*, 111, 233–51.

Rorty, R. (1970): "In Defense of Eliminative Materialism," *The Review of Metaphysics*, 24, 112–21.

Rueger, A. (2000a): "Physical Emergence, Diachronic and Synchronic," *Synthese*, 124, 297–322.

Rueger, A. (2000b): "Robust Supervenience and Emergence," *Philosophy of Science*, 67, 466–89.

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

Scerri, E. (1994): "Has Chemistry Been at Least Approximately Reduced to Quantum Mechanics?" in D. Hull, M. Forbes and R. Burian (eds.), *PSA 1994, vol. 1*. East Lansing, MI: Philosophy of Science Association, 160–70.

Schaffner, K. F. (1992): "Philosophy of Medicine," in M. H. Salmon, J. Earman, C. Glymour, J. G. Lennox, P. Machamer, J. E. McGuire, J. D. Norton, W. C. Salmon and K. F. Schaffner (eds.), *Introduction to the Philosophy of Science*, Englewood Cliffs: Prentice Hall, 310–45.

Schaffner, K. F. (1998): "Genes, Behavior, and Developmental Emergentism: One Process, Indivisible?" *Philosophy of Science*, 65, 209–52.

Schaffner, K. F. (2000): "Behavior at the Organismal and Molecular Levels: The Case of *C. elegans*," *Philosophy of Science*, supplement to 67(3), D. A. Howard (ed.), *Part II: Symposia Papers*, S273–S288.

Silberstein, M. (1998): "Emergence and the Mind/Body Problem," *Journal of Consciousness Studies*, 5, 464–82.

Silberstein, M. (2001): "Converging on Emergence: Consciousness, Causation and Explanation", *Journal of Consciousness Studies*, 8(9)(10), 61–98.

Silberstein, M. and McGeever, J. (1999): "The Search for Ontological Emergence," *The Philosophical Quarterly*, 49, 182–200.

Sklar, L. (1967): "Types of Inter-theoretic Reduction," *The British Journal of the Philosophy of Science*, 18, 109–24.

Sklar, L. (1993): *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Sklar, L. (1999): "The Reduction (?) of Thermodynamics to Statistical Mechanics," *Philosophical Studies*, 95(1–2), 187–99.

Stephan, A. (1992): "Emergence – A Systematic View on its Historical Facets," in A. Beckermann, H. Flor and J. Kim (eds.), *Emergence or Reduction?: Essays on the Prospects of Nonreductive Physicalism*, New York: DeGruyter, 68–90.

Takesaki, M. (1970): "Disjointness of the kms States of Different Temperatures," *Communications in Mathematical Physics*, 17, 33–41.

Van Fraassen, B. C. (1989): *Laws and Symmetry*. Oxford: Oxford University Press.

Wilkes, K. (1988): "Yishi, Duh, Um and Consciousness," in A. J. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science*, Oxford: Clarendon Press, 148–77.

Wilkes, K. (1995): "Losing Consciousness," in T. Metzinger (ed.), *Conscious Experience*, Thorverton: Imprint Academic, 35–89.

Wimsatt, W. C. (1976): "Reductive Explanation: A Functional Account," in R. S. Cohen, C. A. Hooker, A. C. Michalos and G. van Evra (eds.), *Philosophy of Science Association*, Dordrecht: Reidel, 671–710.

Woody, A. I. (2000): "Putting Quantum Mechanics to Work in Chemistry: The Power of Diagrammatic Representation," *Philosophy of Science*, supplement 67(3), D. A. Howard (ed.), *Part II: Symposia Papers*, S612–S627.

The Theory of Everything

R. B. Laughlin* and David Pines†‡§

*Department of Physics, Stanford University, Stanford, CA 94305; †Institute for Complex Adaptive Matter, University of California Office of the President, Oakland, CA 94607; ‡Science and Technology Center for Superconductivity, University of Illinois, Urbana, IL 61801; and §Los Alamos Neutron Science Center Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Contributed by David Pines, November 18, 1999

We discuss recent developments in our understanding of matter, broadly construed, and their implications for contemporary research in fundamental physics.

The Theory of Everything is a term for the ultimate theory of the universe—a set of equations capable of describing all phenomena that have been observed, or that will ever be observed (1). It is the modern incarnation of the reductionist ideal of the ancient Greeks, an approach to the natural world that has been fabulously successful in bettering the lot of mankind and continues in many people's minds to be the central paradigm of physics. A special case of this idea, and also a beautiful instance of it, is the equation of conventional nonrelativistic quantum mechanics, which describes the everyday world of human beings—air, water, rocks, fire, people, and so forth. The details of this equation are less important than the fact that it can be written down simply and is completely specified by a handful of known quantities: the charge and mass of the electron, the charges and masses of the atomic nuclei, and Planck's constant. For experts we write

$$i\hbar\,\frac{\partial}{\partial t}\,|\Psi> \;=\; \mathcal{H}|\Psi> \qquad [1]$$

where

$$\mathcal{H} = -\sum_{j}^{N_e}\frac{\hbar^2}{2m}\nabla_j^2 - \sum_{\alpha}^{N_i}\frac{\hbar^2}{2M_\alpha}\nabla_\alpha^2 - \sum_{j}^{N_e}\sum_{\alpha}^{N_i}\frac{Z_\alpha e^2}{|\vec{r}_j-\vec{R}_\alpha|} + \sum_{j\ll k}^{N_e}\frac{e^2}{|\vec{r}_j-\vec{r}_k|} + \sum_{\alpha\ll\beta}^{N_j}\frac{Z_\alpha Z_\beta e^2}{|\vec{R}_\alpha-\vec{r}_\beta|}. \qquad [2]$$

The symbols $Z\alpha$ and $M\alpha$ are the atomic number and mass of the α^{th} nucleus, $R\alpha$ is the location of this nucleus, e and m are the electron charge and mass, r_j is the location of the j^{th} electron, and $\hbar$ is Planck's constant.

Less immediate things in the universe, such as the planet Jupiter, nuclear fission, the sun, or isotopic abundances of elements in space are not described by this equation, because important elements such as gravity and nuclear interactions are missing. But except for light, which is easily included, and possibly gravity, these missing parts are irrelevant to people-scale phenomena. Eqs. **1** and **2** are, for all practical purposes, the Theory of Everything for our everyday world.

However, it is obvious glancing through this list that the Theory of Everything is not even remotely a theory of every thing (2). We know this equation is correct because it has been solved accurately for small numbers of particles (isolated atoms and small molecules) and found to agree in minute detail with experiment (3–5). However, it cannot be solved accurately when the number of particles exceeds about 10. No computer existing, or that will ever exist, can break this barrier because it is a catastrophe of dimension. If the amount of computer memory required to represent the quantum wavefunction of one particle is N then the amount required to represent the wavefunction of k particles is N^k. It is possible to perform approximate calculations for larger systems, and it is through such calculations that we have learned why atoms have the size they do, why chemical bonds have the length and strength they do, why solid matter has the elastic properties it does, why some things are transparent while others reflect or absorb light (6). With a little more experimental input for guidance it is even possible to predict atomic conformations of small molecules, simple chemical reaction rates, structural phase transitions, ferromagnetism, and sometimes even superconducting transition temperatures (7). But the schemes for approximating are not first-principles deductions but are rather art keyed to experiment, and thus tend to be the least reliable precisely when reliability is most needed, i.e., when experimental information is scarce, the physical behavior has no precedent, and the key questions have not yet been identified. There are many notorious failures of alleged *ab initio* computation methods, including the phase diagram of liquid ^{3}He and the entire phenomenonology of high-temperature superconductors (8–10). Predicting protein functionality or the behavior of the human brain from these equations is patently absurd. So the triumph of the reductionism of the Greeks is a pyrrhic victory: We have succeeded in reducing all of ordinary physical behavior to a simple, correct Theory of Everything only to discover that it has revealed exactly nothing about many things of great importance.

In light of this fact it strikes a thinking person as odd that the parameters e, $\hbar$, and m appearing in these equations may be measured accurately in laboratory experiments involving large numbers of particles. The electron charge, for example, may be accurately measured by passing current through an electrochemical cell, plating out metal atoms, and measuring the mass deposited, the separation of the atoms in the crystal being known from x-ray diffraction (11). Simple electrical measurements performed on superconducting rings determine to high accuracy the quantity the quantum of magnetic flux $hc/2e$ (11). A version of this phenomenon also is seen in superfluid helium, where coupling to electromagnetism is irrelevant (12). Four-point conductance measurements on semiconductors in the quantum Hall regime accurately determine the quantity e^2/h (13). The magnetic field generated by a superconductor that is mechanically rotated measures e/mc (14, 15). These things are clearly true, yet they cannot be deduced by direct calculation from the Theory of Everything, for exact results cannot be predicted by approximate calculations. This point is still not understood by many professional physicists, who find it easier to believe that a deductive link exists and has only to be discovered than to face the truth that there is no link. But it is true nonetheless. Experiments of this kind work because there are higher organizing principles in nature that make them work. The Josephson quantum is exact because of the principle of continuous symmetry breaking (16). The quantum Hall effect is exact because of localization (17). Neither of these things can be deduced from microscopics, and both are transcendent, in that they would continue to be true and to lead to exact results even if the Theory of Everything were changed. Thus the existence of these effects is profoundly important, for it shows us that for at least some

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

fundamental things in nature the Theory of Everything is irrelevant. P. W. Anderson's famous and apt description of this state of affairs is "more is different" (2).

The emergent physical phenomena regulated by higher organizing principles have a property, namely their insensitivity to microscopics, that is directly relevant to the broad question of what is knowable in the deepest sense of the term. The low-energy excitation spectrum of a conventional superconductor, for example, is completely generic and is characterized by a handful of parameters that may be determined experimentally but cannot, in general, be computed from first principles. An even more trivial example is the low-energy excitation spectrum of a conventional crystalline insulator, which consists of transverse and longitudinal sound and nothing else, regardless of details. It is rather obvious that one does not need to prove the existence of sound in a solid, for it follows from the existence of elastic moduli at long length scales, which in turn follows from the spontaneous breaking of translational and rotational symmetry characteristic of the crystalline state (16). Conversely, one therefore learns little about the atomic structure of a crystalline solid by measuring its acoustics.

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. There are many of these, the classic prototype being the Landau fermi liquid, the state of matter represented by conventional metals and normal ^{3}He (18). Landau realized that the existence of well-defined fermionic quasiparticles at a fermi surface was a universal property of such systems independent of microscopic details, and he eventually abstracted this to the more general idea that low-energy elementary excitation spectra were generic and characteristic of distinct stable states of matter. Other important quantum protectorates include superfluidity in Bose liquids such as ^{4}He and the newly discovered atomic condensates (19–21), superconductivity (22, 23), band insulation (24), ferromagnetism (25), antiferromagnetism (26), and the quantum Hall states (27). The low-energy excited quantum states of these systems are particles in exactly the same sense that the electron in the vacuum of quantum electrodynamics is a particle: They carry momentum, energy, spin, and charge, scatter off one another according to simple rules, obey fermi or bose statistics depending on their nature, and in some cases are even "relativistic," in the sense of being described quantitively by Dirac or Klein-Gordon equations at low energy scales. Yet they are not elementary, and, as in the case of sound, simply do not exist outside the context of the stable state of matter in which they live. These quantum protectorates, with their associated emergent behavior, provide us with explicit demonstrations that the underlying microscopic theory can easily have no measurable consequences whatsoever at low energies. The nature of the underlying theory is unknowable until one raises the energy scale sufficiently to escape protection.

Thus far we have addressed the behavior of matter at comparatively low energies. But why should the universe be any different? The vacuum of space-time has a number of properties (relativity, renormalizability, gauge forces, fractional quantum numbers) that ordinary matter does not possess, and this state of affairs is alleged to be something extraordinary distinguishing the matter making up the universe from the matter we see in the laboratory (28). But this is incorrect. It has been known since the early 1970s that renormalizability is an emergent property of ordinary matter either in stable quantum phases, such as the superconducting state, or at particular zero-temperature phase transitions between such states called quantum critical points (29, 30). In either case the low-energy excitation spectrum becomes more and more generic and less and less sensitive to microscopic details as the energy scale of the measurement is lowered, until in the extreme limit of low energy all evidence of the microscopic equations vanishes away. The emergent renormalizability of quantum critical points is formally equivalent to that postulated in the standard model of elementary particles right down to the specific phrase "relevant direction" used to describe measurable quantities surviving renormalization. At least in some cases there is thought to be an emergent relativity principle in the bargain (29, 30). The rest of the strange agents in the standard model also have laboratory analogues. Particles carrying fractional quantum numbers and gauge forces between these particles occur as emergent phenomena in the fractional quantum Hall effect (17). The Higgs mechanism is nothing but superconductivity with a few technical modifications (31). Dirac fermions, spontaneous breaking of CP, and topological defects all occur in the low-energy spectrum of superfluid ^{3}He (32–34).

Whether the universe is near a quantum critical point is not known one way or the other, for the physics of renormalization blinds one to the underlying microscopics as a matter of principle when only low-energy measurements are available. But that is exactly the point. The belief on the part of many that the renormalizability of the universe is a constraint on an underlying microscopic Theory of Everything rather than an emergent property is nothing but an unfalsifiable article of faith. But if proximity to a quantum critical point turns out to be responsible for this behavior, then just as it is impossible to infer the atomic structure of a solid by measuring long-wavelength sound, so might it be impossible to determine the true microscopic basis of the universe with the experimental tools presently at our disposal. The standard model and models based conceptually on it would be nothing but mathematically elegant phenomenological descriptions of low-energy behavior, from which, until experiments or observations could be carried out that fall outside the its region of validity, very little could be inferred about the underlying microscopic Theory of Everything. Big Bang cosmology is vulnerable to the same criticism. No one familiar with violent high-temperature phenomena would dare to infer anything about Eqs. **1** and **2** by studying explosions, for they are unstable and quite unpredictable one experiment to the next (35, 36). The assumption that the early universe should be exempt from this problem is not justified by anything except wishful thinking. It could very well turn out that the Big Bang is the ultimate emergent phenomenon, for it is impossible to miss the similarity between the large-scale structure recently discovered in the density of galaxies and the structure of styrofoam, popcorn, or puffed cereals (37, 38).

Self-organization and protection are not inherently quantum phenomena. They occur equally well in systems with temperatures or frequency scales of measurement so high that quantum effects are unobservable. Indeed the first experimental measurements of critical exponents were made on classical fluids near their liquid-vapor critical points (39). Good examples would be the spontaneous crystallization exhibited by ball bearings placed in a shallow bowl, the emission of vortices by an airplane wing (40), finite-temperature ferromagnetism, ordering phenomena in liquid crystals (41), or the spontaneous formation of micelle membranes (42). To this day the best experimental confirmations of the renormalization group come from measurements of finite-temperature critical points (43). As is the case in quantum systems, these classical ones have low-frequency dynamic properties that are regulated by principles and independent of microscopic details (44, 45). The existence of classical protectorates raises the possibility that such principles might even be at work in biology (46).

What do we learn from a closer examination of quantum and classical protectorates? First, that these are governed by emergent rules. This means, in practice, that if you are locked in a room with the system Hamiltonian, you can't figure the rules out in the absence of experiment, and hand-shaking between theory

and experiment. Second, one can follow each of the ideas that explain the behavior of the protectorates we have mentioned as it evolved historically. In solid-state physics, the experimental tools available were mainly long-wavelength, so that one needed to exploit the atomic perfection of crystal lattices to infer the rules. Imperfection is always present, but time and again it was found that fundamental understanding of the emergent rules had to wait until the materials became sufficiently free of imperfection. Conventional superconductors, for which nonmagnetic impurities do not interfere appreciably with superconductivity, provide an interesting counterexample. In general it took a long time to establish that there really were higher organizing principles leading to quantum protectorates. The reason was partly materials, but also the indirectness of the information provided by experiment and the difficulty in consolidating that information, including throwing out the results of experiments that have been perfectly executed, but provide information on minute details of a particular sample, rather than on global principles that apply to all samples.

Some protectorates have prototypes for which the logical path to microscopics is at least discernable. This helped in establishing the viability of their assignment as protectorates. But we now understand that this is not always the case. For example, superfluid ^{3}He, heavy-fermion metals, and cuprate superconductors appear to be systems in which all vestiges of this link have disappeared, and one is left with nothing but the low-energy principle itself. This problem is exacerbated when the principles of self-organization responsible for emergent behavior compete. When more than one kind of ordering is possible the system decides what to do based on subtleties that are often beyond our ken. How can one distinguish between such competition, as exists for example, in the cuprate superconductors, and a "mess"? The history of physics has shown that higher organizing principles are best identified in the limiting case in which the competition is turned off, and the key breakthroughs are almost always associated with the serendipitous discovery of such limits. Indeed, one could ask whether the laws of quantum mechanics would ever have been discovered if there had been no hydrogen atom. The laws are just as true in the methane molecule and are equally simple, but their manifestations are complicated.

The fact that the essential role played by higher organizing principles in determining emergent behavior continues to be disavowed by so many physical scientists is a poignant comment on the nature of modern science. To solid-state physicists and chemists, who are schooled in quantum mechanics and deal with it every day in the context of unpredictable electronic phenomena such as organogels (47), Kondo insulators (48), or cuprate superconductivity, the existence of these principles is so obvious that it is a cliché not discussed in polite company. However, to other kinds of scientist the idea is considered dangerous and ludicrous, for it is fundamentally at odds with the reductionist beliefs central to much of physics. But the safety that comes from acknowledging only the facts one likes is fundamentally incompatible with science. Sooner or later it must be swept away by the forces of history.

For the biologist, evolution and emergence are part of daily life. For many physicists, on the other hand, the transition from a reductionist approach may not be easy, but should, in the long run, prove highly satisfying. Living with emergence means, among other things, focusing on what experiment tells us about candidate scenarios for the way a given system might behave before attempting to explore the consequences of any specific model. This contrasts sharply with the imperative of reductionism, which requires us never to use experiment, as its objective is to construct a deductive path from the ultimate equations to the experiment without cheating. But this is unreasonable when the behavior in question is emergent, for the higher organizing principles—the core physical ideas on which the model is based—would have to be deduced from the underlying equations, and this is, in general, impossible. Repudiation of this physically unreasonable constraint is the first step down the road to fundamental discovery. No problem in physics in our time has received more attention, and with less in the way of concrete success, than that of the behavior of the cuprate superconductors, whose superconductivity was discovered serendipitously, and whose properties, especially in the underdoped region, continue to surprise (49, 50). As the high-T_c community has learned to its sorrow, deduction from microscopics has not explained, and probably cannot explain as a matter of principle, the wealth of crossover behavior discovered in the normal state of the underdoped systems, much less the remarkably high superconducting transition temperatures measured at optimal doping. Paradoxically high-T_c continues to be the most important problem in solid-state physics, and perhaps physics generally, because this very richness of behavior strongly suggests the presence of a fundamentally new and unprecedented kind of quantum emergence.

In his book "The End of Science" John Horgan (51) argues that our civilization is now facing barriers to the acquisition of knowledge so fundamental that the Golden Age of Science must be thought of as over. It is an instructive and humbling experience to attempt explaining this idea to a child. The outcome is always the same. The child eventually stops listening, smiles politely, and then runs off to explore the countless infinities of new things in his or her world. Horgan's book might more properly have been called the End of Reductionism, for it is actually a call to those of us concerned with the health of physical science to face the truth that in most respects the reductionist ideal has reached its limits as a guiding principle. Rather than a Theory of Everything we appear to face a hierarchy of Theories of Things, each emerging from its parent and evolving into its children as the energy scale is lowered. The end of reductionism is, however, not the end of science, or even the end of theoretical physics. How do proteins work their wonders? Why do magnetic insulators superconduct? Why is ^{3}He a superfluid? Why is the electron mass in some metals stupendously large? Why do turbulent fluids display patterns? Why does black hole formation so resemble a quantum phase transition? Why do galaxies emit such enormous jets? The list is endless, and it does not include the most important questions of all, namely those raised by discoveries yet to come. The central task of theoretical physics in our time is no longer to write down the ultimate equations but rather to catalogue and understand emergent behavior in its many guises, including potentially life itself. We call this physics of the next century the study of complex adaptive matter. For better or worse we are now witnessing a transition from the science of the past, so intimately linked to reductionism, to the study of complex adaptive matter, firmly based in experiment, with its hope for providing a jumping-off point for new discoveries, new concepts, and new wisdom.

We thank E. Abrahams, P. W. Anderson, G. Baym, S. Chakravarty, G. Volovik, and P. Nozières for thoughtful criticism of the manuscript. D.P. thanks the Aspen Institute for Physics, where part of the manuscript was written, for its hospitality. This work was supported primarily by the National Science Foundation under Grant DMR-9813899 and the Department of Energy. Additional support was provided by National Aeronautics and Space Administration Collaborative Agreement 974-9801 and the Ministry of International Trade and Industry of Japan through the New Energy and Industrial Technology Development Organization, and the Science and Technology Center for Superconductivity under National Science Foundation Grant No. DMR 91-2000.

1. Gribbin, G. R. (1999) *The Search for Superstrings, Symmetry, and the Theory of Everything* (Little Brown, New York).
2. Anderson, P. W. (1972) *Science* **177,** 393–396.
3. Graham, R. L., Yeager, D. L., Olsen, J., Jorgensen, P., Harrison, R., Zarrabian,

S. & Bartlett, R. (1986) *J. Chem. Phys*. **85,** 6544–6549.
4. Wolnicwicz, L. (1995) *J. Chem. Phys*. **103,** 1792.
5. Pople, J. (2000) *Rev. Mod. Phys*., in press.
6. Slater, J. C. (1968) *Quantum Theory of Matter* (McGraw–Hill, New York).
7. Chang, K. J., Dacorogna, M. M., Cohen, M. L., Mignot, J. M., Chouteau, G. & Martinez, G. (1985) *Phys. Rev. Lett*. **54,** 2375–2378.
8. Vollhardt, D. & Wölfle, P. (1990) *The Superfluid Phases of Helium 3* (Taylor and Francis, London).
9. Osheroff, D. D. (1997) *Rev. Mod. Phys*. **69,** 667–682.
10. Bok, J., Deutscher, G., Pavuna, D. & Wolf, S. A., eds. (1998) *The Gap Symmetry and Fluctuations in High-T_c Superconductors* (Plenum, New York).
11. Taylor, B. N., Parker, W. H. & Langenberg, D. N. (1969) *Rev. Mod. Phys*. **41,** 375 and references therein.
12. Pereversev, S. V., Loshak, A., Backhaus, S., Davis, J. C. & Packard, R. E. (1997) *Nature (London)* **385,** 449–451.
13. von Klitzing, K., Dorda, G. & Pepper, M. (1980) *Phys. Rev. Lett*. **45,** 494.
14. Liu, M. (1998) *Phys. Rev. Lett*. **81,** 3223–3226.
15. Tate, J., Felch, S. B. & Cabrera, B. (1990) *Phys. Rev. B* **42,** 7885–7893.
16. Anderson, P. W. (1976) *Basic Notions of Condensed Matter Physics* (Benjamin, Menlo Park, CA).
17. Laughlin, R. B. (1999) *Rev. Mod. Phys*. **71,** 863–874.
18. Pines, D. & Nozieres, P. (1966) *The Theory of Quantum Liquids* (Benjamin, New York).
19. Anderson, M. H., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. (1995) *Science* **269,** 198–201.
20. Bradley, C. C., Sackett, C. A., Tollett, J. J. & Hulet, R. G. (1995) *Phys. Rev. Lett*. **75,** 1687–1690.
21. Davis, K. B., Mewes, M. O., Andrew, M. R., Vandruten, N. J., Durfee, D. S., Kurn, D. M. & Ketterle, W. (1995) *Phys. Rev. Lett*. **75,** 3969–3973.
22. Schrieffer, J. R. (1983) *Theory of Superconductivity* (Benjamin, New York).
23. de Gennes, P. G. (1966) *Superconductivity of Metals and Alloys* (Benjamin, New York).
24. Sze, S. M. (1981) *Physics of Semiconductor Devices* (Wiley, New York).
25. Herring, C. (1966) in *Magnetism*, eds. Rado, G. T. & Suhl, H. (Academic, New York).
26. Kittel, C. (1963) *Quantum Theory of Solids* (Wiley, New York).
27. Prange, R. E. & Girvin, S. M. (1987) *The Quantum Hall Effect* (Springer, Heidelberg).
28. Peskin, M. E. & Schroeder, D. V. (1995) *Introduction to Quantum Field Theory* (Addison–Wesley, Reading, MA).
29. Wilson, K. G. (1983) *Rev. Mod. Phys*. **55,** 583.
30. Fisher, M. E. (1974) *Rev. Mod. Phys*. **46,** 597.
31. Anderson, P. W. (1963) *Phys. Rev*. **130,** 439.
32. Volovik, G. E. (1992) *Exotic Properties of Superfluid* ^{3}He (World Scientific, Singapore).
33. Volovik, G. E. (1998) *Physica B* **255,** 86.
34. Volovik, G. E. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 6042–6047.
35. Zeldovich, Y. B. & Raizer, Y. P. (1966) *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Academic, New York).
36. Sturtevant, B., Shepard, J. E. & Hornung, H. G., eds. (1995) *Shock Waves* (World Scientific, Singapore).
37. Huchra, J. P., Geller, M. J., Delapparent, V. & Corwin, H. G. (1990) *Astrophys. J.* Suppl., **72,** 433–470.
38. Shechtman, S. A., Landy, S. D., Oemler, A., Tucker, D. L., Lin, H., Kirshner, R. P. & Schechter, P. L. (1996) *Astrophys. J.* **470,** 172–188.
39. Levett Sengers, J. M. H. (1974) *Physica* **73,** 73.
40. Saffman, P. G. (1992) *Vortex Dynamics* (Cambridge Univ. Press, Cambridge).
41. Lubensky, T. C., Harris, A. B., Kamien, R. D. & Yan, G. (1998) *Ferroelectrics* **212,** 1–20.
42. Safran, S. (1994) *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes* (Addison–Wesley, Reading, MA).
43. Stanley, H. E. (1987) *Introduction to Phase Transitions and Critical Phenomena* (Oxford, New York).
44. Riste, T., Samuelsen, E. J., Otnes, K. & Feder, J. (1971) *Solid State Comm*. **9,** 1455–1458.
45. Courtery, E. & Gammon, R. W. (1979) *Phys. Rev. Lett*. **43,** 1026.
46. Doniach, S. (1994) *Statistical Mechanics, Protein Structure, and Protein-Substrate Interactions* (Plenum, New York).
47. Geiger, C., Stanescu, M., Chen, L. H. & Whitten, D. G. (1999) *Langmuir* **15,** 2241–2245.
48. Aeppli, G. & Fisk, Z. (1992) *Comments Condens. Matter Phys*. **16,** 155.
49. Pines, D. (1997) *Z. Phys*. **103,** 129.
50. Pines, D. (1998) in *The Gap Symmetry and Fluctuations in High-T_c Superconductors*, eds. Bok, J., Deutscher, G., Pavuna, D. & Wolf, S. A. (Plenum, New York), pp. 111–142.
51. Horgan, J. (1997) *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age* (Addison–Wesley, Reading, MA).

Manfred D. Laublichler

Statement

and

Readings

The emergence of evolutionary novelties: social insects as a models system for evolutionary developmental biology

One of the remaining challenges of evolutionary biology is to mechanistically explain the origin of complex novel structures and behaviors. Darwin already struggled with this problem in the Origin of Species and recent advances in comparative genomics clearly demonstrate that novel features are thus not just a consequence of new genes or even new versions of old genes. What then accounts for the obvious phenotypic differences between groups of organisms and for the emergence of novel structures in the course of evolution? The short answer to this question is that changes in the developmental systems of these organisms and in most cases changes in the regulatory networks of genes are responsible for these differences. Intuitively this is a rather obvious conclusion as all phenotypic differences both morphological and behavioral first emerge during the development of individual organisms. Changes in developmental processes will thus always be the immediate or proximate causes of phenotypic variation. Still, several questions remain: Exactly how do developmental mechanisms contribute to phenotypic changes and also how can developmental explanations be integrated into the theoretical framework of evolutionary biology, that wants to explain cladogenesis and adaptation? And, more practically, what are the best model systems to study these questions experimentally and comparatively?

Studies of evolutionary novelties still face many difficulties. Part of the problem can be attributed to the focus on traditional model organisms, which tend to concentrate on major morphological transformations, such as the fin-limb transition in early tetrapod evolution. Here, I will introdcue new model systems – social insects - that are well suited to address questions about the origin of evolutionary novelties experimentally as well as theoretically. Social insects display a remarkable diversity in social behavior and structure even between closely related species, thus allowing the repeated and direct study of the evolution of social novelties. Social novelties include morphological, physiological and behavioral innovations like worker polymorphism, cooption of hormone regulation for division of labor, or the bee dance. For many social species we know a good deal about their phylogenetic relationships, developmental mechanisms (larval and adult maturation) as well as their physiological and behavioral repertoire. Many social insects can also be manipulated experimentally in the lab, in several cases actually inducing novel types of social behavior. Since 2006 the honey bee is the first social insect with a fully annotated genome and it also has now the adequate tools to manipulate expression levels of specific genes. Social insects are unique in that they provide a system in which individuals can express phenotypes that are detrimental or maladaptive for the individual if they are solitary but are highly adaptive in the context of a colony and that the expression of a certain phenotype is context dependent. Social insects are therefore ideal model systems for the study of evolutionary novelties and the role of development, environment and epigenetics and the interaction of these three factors during the evolution of a novel trait.

In my presentation I will discuss some experimental strategies and theoretical implications of “Social Insect Evo Devo” and discuss the implications of this work for larger questions of emergence.

LETTERS

Complex social behaviour derived from maternal reproductive traits

Gro V. Amdam[1,2], Angela Csondes[3], M. Kim Fondrk[1] & Robert E. Page Jr[1]

A fundamental goal of sociobiology is to explain how complex social behaviour evolves[1], especially in social insects, the exemplars of social living. Although still the subject of much controversy[2], recent theoretical explanations have focused on the evolutionary origins of worker behaviour (assistance from daughters that remain in the nest and help their mother to reproduce) through expression of maternal care behaviour towards siblings[3,4]. A key prediction of this evolutionary model is that traits involved in maternal care have been co-opted through heterochronous expression of maternal genes[5] to result in sib-care, the hallmark of highly evolved social life in insects[6]. A coupling of maternal behaviour to reproductive status evolved in solitary insects, and was a ready substrate for the evolution of worker-containing societies[3,4,7,8]. Here we show that division of foraging labour among worker honey bees (*Apis mellifera*) is linked to the reproductive status of facultatively sterile females. We thereby identify the evolutionary origin of a widely expressed social-insect behavioural syndrome[1,5,7,9], and provide a direct demonstration of how variation in maternal reproductive traits gives rise to complex social behaviour in non-reproductive helpers.

Worker honey bees change the tasks that they perform with age[10]. This behaviour results in a division of labour that is age-associated[11]. Workers usually make a transition from working in the nest to foraging in their second or third week of life[12], and foragers often specialize in collecting nectar or pollen. Recent studies have identified a suite of traits that differ between nectar and pollen foragers[9]. These traits are affected by a pleiotropic genetic network[13], and it has been suggested that this pleiotropy can be explained if a reproductive regulatory network was co-opted by natural selection to differentiate the foraging behaviour of the facultatively sterile workers[7]. This hypothesis emerged from studies of honey bees that were selected to collect and store high (the high-hoarding strain) or low (the low-hoarding strain) amounts of pollen[14]. Traits of the strains diverge, so that high pollen-hoarding bees switch from nest tasks to foraging earlier in life, and are more likely to collect pollen and carry larger pollen loads. Bees from the high pollen-hoarding strain are more likely than bees from the low pollen-hoarding strain to collect water and nectar with low sugar concentration, and at emergence they have higher haemolymph (blood) levels of juvenile hormone and vitellogenin protein[7]. Pollen foraging is a maternal reproductive behaviour in solitary bees, and non-reproductive females feed mainly on nectar[15]. Elevated juvenile hormone levels cause physiological and behavioural changes during the reproductive maturation of many insects[7,16,17], and vitellogenin is a conserved yolk precursor synthesized by most oviparous females[18]. Therefore, the evidence from pollen-hoarding strains suggests that nectar-foraging bees display a non-reproductive phenotype, whereas pollen foragers display the ancestral maternal character state of solitary species[7]. As a consequence, the foraging division of labour between worker bees would be derived from variation in maternal reproductive traits. Validation of this hypothesis, however, requires the demonstration of a relationship between the reproductive status and the foraging behaviour of honey bee workers[7].

We addressed the debate on the origin of complex social behaviour by first inspecting the number of ovarioles (egg-forming filaments in the ovary) in newly emerged workers from the previously examined[7] high and low pollen-hoarding strains. Developmental differentiation of ovariole number[19] is influenced by endocrine regulatory networks that during the adult stage are responsible for modulation of maternal reproductive behaviour in insects[7,20,21]. Ovariole number is, moreover, a recognized marker of reproductive potential in the honey bee[22], as well as in the well-studied solitary insect *Drosophila*[21,23]. We found that high pollen-hoarding strain workers had more ovarioles than those from the low pollen-hoarding strain (factorial analysis of variance (ANOVA), $P < 0.005$). This difference was independent (factorial ANOVA, $P = 0.72$) of whether the workers were co-fostered (mean $\pm$ s.e.m., 5.56 ± 0.42 and 2.96 ± 0.31 ovarioles for the high and low pollen-hoarding strains, respectively; $n = 25$ per strain) or reared by their native colony (5.88 ± 0.41 and 2.88 ± 0.19 ovarioles; $n = 25$). Furthermore, bees with eight or more ovarioles were exclusively found in the high pollen-hoarding strain, where they represented 26% of the sample population (Supplementary Table S1). We also observed that this higher number of ovary filaments was associated with a swelling of the ovarioles (Supplementary Table S1), which is an established indicator of previtellogenic ovarian activation[24,25]. These results demonstrate that a regulatory system that affects female reproductive morphology, physiology, and behaviour[7,20,21,26] is differentially tuned during the development of honey bees characterized by different levels of pollen hoarding.

To verify that the observed variation in ovariole number translates into functional differences in adult reproductive potential, we next introduced high and low pollen-hoarding bees into host colonies with or without a queen (the presence of a queen inhibits worker oogenesis[27]). The experimental design also controlled for rearing environment by using workers that were co-fostered and workers that were reared in their native high or low pollen-hoarding strain colony. The bees were examined after 10–21 days. In colonies with a queen ($n = 6$ colonies), we found that $29.5 \pm 3.6\%$ of the bees from the high pollen-hoarding strain ($n = 201$) had activated previtellogenic ovaries, compared with $2.6 \pm 1.8\%$ of the workers from the low pollen-hoarding strain ($n = 201$) (Supplementary Table S2). This divergence (factorial ANOVA, $P < 0.005$) was independent of whether the bees were co-fostered or reared by their native colony (factorial ANOVA, $P = 0.42$). The effect of hoarding strain on the proportion of individuals with non-activated ovaries versus previtellogenic ovaries was significant in all hives (V-square test,

[1]Arizona State University, School of Life Sciences, Tempe, Arizona 85287, USA. [2]University of Life Sciences, Department of Animal and Aquacultural Sciences, 1432 Aas, Norway. [3]University of California, Department of Entomology, Davis, California 95616, USA.

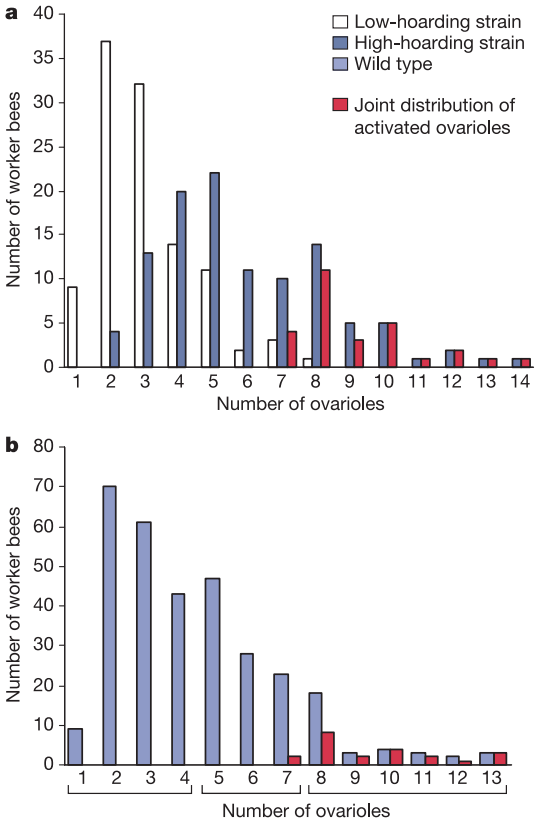


Figure 1 | Distributions of ovariole number and patterns of previtellogenic ovarian activation in worker bees. **a**, Ovariole number in mature 10- to 21-day-old bees from strains selected for high or low levels of pollen hoarding ($n = 109$ bees per strain). **b**, Samples from wild-type bees collected at presumably their first foraging flight ($n = 314$). The mean numbers of ovarioles (±s.e.m.) for groups with 1–4, 5–7 and 8 or more ovarioles are 2.75 ± 0.06, 5.76 ± 0.08 and 9.30 ± 0.30, respectively. The joint distributions of ovarian activation are superimposed on the original densities and refer, therefore, to bees within the genotype-specific data sets.

$P < 0.05$). Also, previtellogenic ovarian activation was exclusively found in workers with seven or more ovarioles (Supplementary Table S2). These results from mature workers (Fig. 1a) correspond with the data from newly emerged bees (Supplementary Table S1), suggesting that a sizable proportion of worker bees selected to collect and store high amounts of pollen emerge with an active ovarian phenotype that persists for several weeks in the presence of a fully functional queen.

In colonies without a queen ($n = 6$ colonies), 75.8 ± 0.1% of the high pollen-hoarding workers ($n = 212$) had active ovaries that were previtellogenic, vitellogenic with developing oocytes, or vitellogenic with eggs (Supplementary Table S2). In comparison, 42.0 ± 0.1% of bees from the low pollen-hoarding strain had active ovaries ($n = 212$). This difference between strains (factorial ANOVA, $P < 0.05$) was independent of rearing environment (factorial ANOVA, $P = 0.94$). The effect of hoarding strain on the proportion of workers with non-activated ovaries versus previtellogenic ovaries was significant in all but one hive (V-square test, $P < 0.05$), and out of the 48 bees with eggs, 36 were from the high pollen-hoarding strain (Supplementary Table S2). Eggs were found in bees with five or more ovarioles (Supplementary Table S2). These results demonstrate that workers selected for a high level of pollen hoarding have a functional phenotype that more frequently achieves an advanced reproductive state.

Finally, we used workers from 'wild-type' colonies (not selected for pollen hoarding) to test whether the trait-associations that characterize the high and low pollen-hoarding strains are present in the general population. Wild-type bees were marked at adult eclosion and later captured at presumably their first foraging flight ($n = 551$). The nectar- and pollen-loads of the workers were quantified, and ovariole number was determined by dissection of those bees ($n = 314$) that carried measurable amounts of nectar or pollen (more than 0.0005 g).

We first investigated whether an association between ovariole number and previtellogenic ovarian activation was present. Activation occurred exclusively in bees with seven or more ovarioles (Fig. 1b), confirming our findings from the selected strains. On the basis of ovariole number, we then divided the data from the 314 workers into three groups. The first group had a mean ovariole number similar to the low pollen-hoarding strain (1–4 ovarioles, $n = 184$), the next had a mean ovariole number comparable to the high strain (5–7 ovarioles, $n = 97$), and the last group consisted of bees with eight or more ovarioles ($n = 33$) (Fig. 1b). Subsequent analysis of the data set showed that ovariole number correlated with the adult age of bees at their first foraging flight, the probability of being a pollen forager, and the nectar concentration collected by the workers (multivariate ANOVA; $P < 0.00001$). Worker bees with 5–7 and 8 or more ovarioles initiated foraging at younger ages than bees with 1–4 ovarioles (Fig. 2a). Workers with 5–7 and 8 or more ovarioles were also more likely to forage for pollen (Fig. 2b). In addition, the bees with 8 or more ovarioles collected lower nectar concentrations than workers with only 1–4 ovary filaments (Fig. 2c). Consequently, the trait-associations of wild-type bees with the greatest number of ovary filaments corresponded precisely with those shown for the strain selected to collect and store high amounts of pollen[7,9].

We conclude that division of foraging labour in the advanced

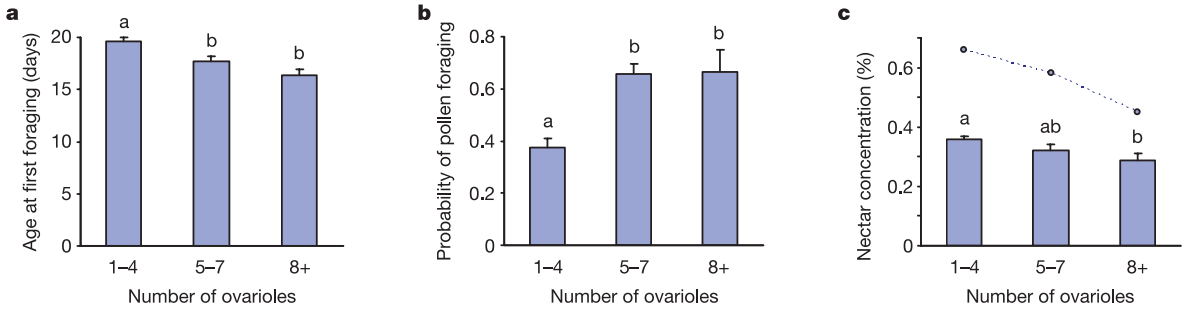


Figure 2 | Correlations between ovariole number and the social behaviour of wild-type bees. **a**, Honey bee age at presumably the first foraging flight. **b**, The probability of being a pollen forager. **c**, The sugar concentration of nectar collected by the worker bees. Data show mean ± s.e.m. Different letters (a, b) refer to groups that were different according to a Fisher's post-hoc test ($P < 0.05$). Points connected by a dotted line in **c** denote the highest nectar concentration collected by any single bee in the respective ovariole groups.

eusocial honey bee emerges from variation in maternal care behaviour. This finding illustrates how the behavioural mechanisms of division of labour evolve from solitary ancestry, and provides an experimental demonstration of the origins of sib-care behaviour from maternal reproductive traits[3–5,7]. The evolution of sib-care from maternal care is a critical step towards the evolution of eusociality in insects, and remains a point of substantial debate[5,8,28,29].

METHODS

Bees selected for high or low levels of pollen hoarding. Larvae from six high and six low pollen-hoarding strain queens were reared together in common wild-type nurse colonies. For workers reared by their native colony, frames with mature pupae were obtained from the same 12 sources. Newly emerged bees were collected for ovarian analysis or marked on the thorax with a spot of paint (Testors Enamel) for identification of strain and age. Marked workers were added to host colonies with or without a queen.

Wild-type bees. Newly emerged bees from four unrelated and unselected source/host colonies were mixed together to obtain a worker pool with high phenotypic variance. The bees were marked (see above) for identification of age, and each source/host colony received 400 workers from the mix. Starting five days later, the hive entrances were monitored between 9:00 in the morning and 14:00 in the afternoon, and marked bees that returned from flight were collected.

Foraging load measurements. Bees were treated with CO_2 until immobile to enable quantification of pollen weight, nectar weight and nectar sugar concentration, as reported previously[30].

Quantification of ovariole number and ovarian physiology. Bees were dissected under a stereomicroscope at ×40 magnification. Incisions were made dorsally, and the number of ovarioles in the right-side ovary[24] was determined at ×100 magnification. The extent of ovarian activation was determined using a relative scale as described previously[24]: 1, non-activated ovary; 2, previtellogenic activated ovary; 3, vitellogenic ovary with developing oocytes; 4, mature ovary with at least one egg.

Data analysis. Ovariole number and ovarian activation in bees selected for high or low levels of pollen hoarding were analysed using factorial ANOVA. Analyses were combined with Fisher's post-hoc and non-parametric *V*-square tests to examine the effect of strain. Foraging data from wild-type bees were analysed with multivariate ANOVA and Fisher's post-hoc test. Ovariole number (coded by group: 1–4, 5–7, and 8 or more ovarioles) and host colony were the categorical factors. The effect of host colony was used to control error variance. Pollen load was coded as a binary variable. Statistica 6.0 software was used.

Received 10 September; accepted 19 October 2005.

1. Robinson, G. E., Grozinger, C. M. & Whitfield, C. W. Sociogenomics: social life in molecular terms. *Nature Rev. Genet.* **6,** 257–270 (2005).
2. Wilson, E. O. & Hölldobler, B. Eusociality: origin and consequences. *Proc. Natl Acad. Sci. USA* **102,** 13367–13371 (2005).
3. West-Eberhard, M. J. in *Animal Societies: Theories and Fact* (eds Itô, Y., Brown, J. L. & Kikkawa, J.) 35–51 (Japan Sci. Soc. Press, Tokyo, 1987).
4. West-Eberhard, M. J. in *Natural History and Evolution of Paper Wasp* (eds Turillazzi, S. & West-Eberhard, M. J.) 290–317 (Oxford Univ. Press, New York, 1996).
5. Linksvayer, T. A. & Wade, M. J. The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: Maternal effects, sib-social effects, and heterochrony. *Q. Rev. Biol.* **80,** 317–336 (2005).
6. West-Eberhard, M. J. *Developmental Plasticity and Evolution* (Oxford Univ. Press, New York, 2003).
7. Amdam, G. V., Norberg, K., Fondrk, M. K. & Page, R. E. Reproductive ground plan may mediate colony-level selection effects on individual foraging behaviour in honey bees. *Proc. Natl Acad. Sci. USA* **101,** 11350–11355 (2004).
8. Hunt, J. H. & Amdam, G. V. Bivoltinism as an antecedent to eusociality in the paper wasp genus Polistes. *Science* **308,** 264–267 (2005).
9. Page, R. E. & Erber, J. Levels of behavioural organization and the evolution of division of labour. *Naturwissenschaften* **89,** 91–106 (2002).
10. Seeley, T. D. *The Wisdom of the Hive* (Harvard Univ. Press, Cambridge, Massachusetts, 1995).
11. Robinson, G. E. Regulation of division of labour in insect societies. *Annu. Rev. Entomol.* **37,** 637–665 (1992).
12. Winston, M. L. *The Biology of the Honey Bee* (Harvard Univ. Press, Cambridge, Massachusetts, 1987).
13. Rueppell, O., Pankiw, T. & Page, R. E. Pleiotropy, epistasis and new QTL: The genetic architecture of honey bee foraging behaviour. *J. Hered.* **95,** 481–491 (2004).
14. Page, R. E. & Fondrk, M. K. The effects of colony-level selection on the social organization of honey bee (*Apis mellifera* L.) colonies: colony-level components of pollen hoarding. *Behav. Ecol. Sociobiol.* **36,** 135–144 (1995).
15. Dunn, T. & Richards, M. H. When to bee social: interactions among environmental constraints, incentives, guarding, and relatedness in a facultatively social carpenter bee. *Behav. Ecol.* **14,** 417–424 (2003).
16. Simonet, G. *et al.* Neuroendocrinological and molecular aspects of insect reproduction. *J. Neuroendocrinol.* **16,** 649–659 (2004).
17. Min, K. J., Taub-Montemayor, T. E., Linse, K. D., Kent, J. W. & Rankin, M. A. Relationship of adipokinetic hormone I and II to migratory propensity in the grasshopper, *Melanoplus sanguinipes*. *Arch. Insect Biochem. Physiol.* **55,** 33–42 (2004).
18. Spieth, J., Nettleton, M., Zuckeraprison, E., Lea, K. & Blumenthal, T. Vitellogenin motifs conserved in nematodes and vertebrates. *J. Mol. Evol.* **32,** 429–438 (1991).
19. Capella, I. C. S. & Hartfelder, K. Juvenile hormone effect on DNA synthesis and apoptosis in caste-specific differentiation of the larval honey bee (*Apis mellifera* L.) ovary. *J. Insect Physiol.* **44,** 385–391 (1998).
20. Tatar, M. & Yin, C. M. Slow aging during insect reproductive diapause: why butterflies, grasshoppers and flies are like worms. *Exp. Gerontol.* **36,** 723–738 (2001).
21. Tu, M. P. & Tatar, M. Juvenile diet restriction and the aging and reproduction of adult *Drosophila melanogaster*. *Aging Cell* **2,** 327–333 (2003).
22. Tanaka, E. D. & Hartfelder, K. The initial stages of oogenesis and their relation to differential fertility in the honey bee (*Apis mellifera*) castes. *Arthropod Struct. Dev.* **33,** 431–442 (2004).
23. Hodin, J. & Riddiford, L. M. Different mechanisms underlie phenotypic plasticity and interspecific variation for a reproductive character in drosophilids (Insecta: Diptera). *Evolution* **54,** 1638–1653 (2000).
24. Hartfelder, K., Bitondi, M. M. G., Santana, W. C. & Simões, Z. L. P. Ecdysteroid titer and reproduction in queens and workers of the honey bee and of a stingless bee: loss of ecdysteroid function at increasing levels of sociality? *Insect Biochem. Mol. Biol.* **32,** 211–216 (2002).
25. Maurizio, A. Pollenernahrung und Lebensvorgange bei der Honigbiene (*Apis mellifera* L.). *Landwirtsch. Jahrb. Schweiz.* **245,** 115–182 (1954).
26. Hartfelder, K., Köstlin, K. & Hepperle, C. Ecdysteroid-dependent protein synthesis in caste-specific development of the larval honey bee ovary. *Rouxs Arch. Dev. Biol.* **205,** 73–80 (1995).
27. Butler, C. G. The control of ovary development in worker honeybees (*Apis mellifera*). *Experientia* **13,** 256–257 (1957).
28. Bloch, G., Wheeler, D. & Robinson, G. E. in *Hormones, Brain and Behavior* (eds Pfaff, D., Arnold, A. P., Etgen, A. M., Fahrbach, S. E. & Rubin, R. T.) 195–235 (Academic, San Diego, 2002).
29. Robinson, G. E. & Ben-Shahar, Y. Social behaviour and comparative genomics: new genes or new gene regulation? *Genes Brain Behav.* **1,** 197–203 (2002).
30. Gary, N. E. & Lorenzen, K. A method for collecting the honey-sac content from honeybees. *J. Apic. Res.* **15,** 73–79 (1976).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A.L.O.T. Aase for assistance with dissections, and K. Hartfelder and P. Kukuk for comments. The project was supported by grants from the Norwegian Research Council to G.V.A, and from the National Institute on Aging and the National Research Initiative of the USDA Cooperative State Research, Education and Extension Service to R.E.P.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.V.A. (Gro.Amdam@asu.edu) or R.E.P. (Robert.Page@asu.edu).

JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 304B:610–618 (2005)

Phenotypic Accommodation: Adaptive Innovation Due to Developmental Plasticity

MARY JANE WEST-EBERHARD*
Smithsonian Tropical Research Institute, Escuela de Biología, Universidad de Costa Rica, Costa Rica, Centroamerica

ABSTRACT Phenotypic accommodation is adaptive adjustment, without genetic change, of variable aspects of the phenotype following a novel input during development. Phenotypic accommodation can facilitate the evolution of novel morphology by alleviating the negative effects of change, and by giving a head start to adaptive evolution in a new direction. Whether induced by a mutation or a novel environmental factor, innovative morphological form comes from ancestral developmental responses, not from the novel inducing factor itself. Phenotypic accommodation is the result of adaptive developmental responses, so the novel morphologies that result are not ''random'' variants, but to some degree reflect past functionality. Phenotypic accommodation is the first step in a process of Darwinian adaptive evolution, or evolution by natural selection, where fitness differences among genetically variable developmental variants cause phenotype-frequency change due to gene-frequency change. *J. Exp. Zool. (Mol. Dev. Evol.) 304B:610–618, 2005.* © 2005 Wiley-Liss, Inc.

Adaptive responsiveness to conditions in the external and internal environment has long been considered a universal property of living things. Large subdisciplines of the biological sciences, including physiology, endocrinology, neurobiology, ethology, embryology, cell biology, and the molecular biology of gene expression, study the mechanisms of adaptive responsiveness, but its significance for the process of evolution has not been extensively explored.

Elsewhere I have argued that developmental plasticity, or responsiveness to external and internal environments whether adaptive or not, can play an important role in evolution (West-Eberhard, 2003). Here I summarize one aspect of that argument, namely, that adaptive flexibility, or phenotypic accommodation, can facilitate the origin and evolution of morphological novelties. *Phenotypic accommodation* is adaptive mutual adjustment, without genetic change, among variable aspects of the phenotype, following a novel or unusual input during development (West-Eberhard, '98, 2003).

The role of flexibility in facilitating evolutionary change has been noted by many authors, including most prominently Baldwin (1896, '02), whose concept of ''organic selection'' meant fitness enhancement due to phenotypic accommodation; Schmalhausen ('49 ['86]), who saw individual adaptability as a stabilizing force that promotes the origin and evolution of morphological novelties; Goldschmidt ('40 ['82]), who discussed how the ''regulative ability'' of developmental mechanisms could facilitate and exaggerate change; Frazzetta ('75), who referred to phenotypic ''compensation''; Müller ('90) on ''ontogenetic buffering''; and Kirschner ('92); and Gerhart and Kirschner ('97), who consider the mechanisms of phenotypic accommodation within cells and during embryogenesis an aspect of ''evolvability.''

PHENOTYPIC ACCOMMODATION IN MORPHOLOGY: THE TWO-LEGGED-GOAT EFFECT

Phenotypic accommodation can include adaptive plasticity in all aspects of the phenotype, including not only morphology, but also physiology and behavior. And it can involve developmental plasticity at more than one level of organization. For example, behavioral accommodation may involve flexible responses of many organs (e.g., heart, brain, and limbs) and mechanisms that operate at multiple levels within them (i.e., tissues, cells, and their component

*Correspondence to: Mary Jane West-Eberhard, Escuela de Biología, Universidad de Costa Rica, Costa Rica, Centroamerica. E-mail: mjwe@sent.com
Received 25 November 2004; Accepted 15 July 2005
Published online 13 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21071.

organelles) (see West-Eberhard, 2003). There are subdisciplines of biology that deal with adaptive accommodation in physiology and behavior, but there is no comparable field devoted primarily to adaptive responses in morphology. Adaptive morphological plasticity is nonetheless well documented, for example in studies of vertebrate muscle and bone (reviews in Slijper, '42a, b; Frazzetta, '75; Wimberger, '94); invertebrate body size and form (e.g., see Bernays, '86; Strathmann et al., '92); and in plants, perhaps the best studied group of organisms with respect to morphological plasticity (reviews in Bradshaw, '65; Schlichting, '86; Sultan, '87, 2000).

A handicapped goat studied by Slijper ('42a, b) can serve to illustrate the phenomenon of morphological phenotypic accommodation. Slijper's goat was born with congenital paralysis of its front legs, so that it could not walk on all fours. It managed to get around by hopping on its hind legs, an example of behavioral accommodation that led to dramatic morphological accommodation as well. When the goat died an accidental death at the age of 1 year, Slijper dissected it and published a description of its altered morphology, which included changes in the bones of the hind legs, the shape of the thoracic skeleton and sternum, changes in the shape and strength of the pelvis, which developed an unusually long ischium. Changes in the pelvic muscles included a greatly elongated and thickened gluteal tongue whose attachment to the bone was reinforced by a novel trait, a set of numerous long, flat tendons.

This example of phenotype accommodation shows how developmental responses can mold the form of a morphological novelty. In Slijper's goat, novel morphology came not from a series of mutational changes, but from reorganized expression of capacities that were already present. In the remainder of this article, I show how such immediate responses can be converted to evolutionary change and facilitate the origin of adaptive novelties.

PHENOTYPIC ACCOMMODATION AND THE ORIGINS OF NOVELTY

A morphological innovation can be defined as an aspect of morphology that was not present in the immediate ancestors of a species, in a given life stage and sex. Mayr ('59, p 89) defined an evolutionary novelty as ''any newly arisen character, structural or otherwise, that differs more than quantitatively from the character that gave rise to it.'' Müller and Wagner ('91, p 243) define morphological novelty as ''a structure that is neither homologous to any structure in the ancestral species nor homonomous to any other structure of the same organism.'' But this definition is impossible to apply given the reorganizational nature of evolutionary change. Unless ''homologous'' means identical, many novelties must have recognizable homologs in ancestral species which have given rise to them through ontogenetic repatterning (Wake and Roth, '89; developmental recombination of West-Eberhard, 2003, 2005). Some innovations, such as those derived via heterotopy (change in the location of expression of an ancestral trait), may exist alongside the ancestral morphology as homonymous structures in the descendent species. [For a more extensive discussion of the homology concept as related to developmental plasticity and evolution, see West-Eberhard (2003) and references therein.]

The two-legged goat is an instructive example for anyone interested in morphological innovation. It does not matter, for the form taken by the morphological change, whether the pivotal change (inability to walk on the front legs) was induced by a mutation or by an environmental accident. The particular characteristics of the novel morphology, that is, the novel features of the bones, muscles and tendons, arose via mechanisms of developmental plasticity, not owing to the particular genetic or environmental change that may have induced them. Any number of mutations or environmental factors could have triggered a defect in the front legs. Whatever the trigger, it acted as a kind of switch mechanism that controlled a whole suite of morphological changes—a complex, coordinated morphological novelty, a new modular ''trait'' whose developmental independence of others is defined by the integrated response of the phenotype to a new input.

A second important point is that the morphological change was mediated by behavior. Behavior is, of course, a common mediator of normal skeletal and muscle development because it is especially flexible in response to environmental contingencies. It follows that behavior must often be an important mechanism in the origins of novel morphological traits. So we have to list behavior and its neuroendocrinological underpinnings, alongside genomic changes, as among the primary developmental causes of morphological novelty.

Two-legged goats are unjustly maligned if treated as mere freaks with no evolutionary

importance. Slijper ('42a, b) compared the morphology of the two-legged goat with that of other bipedal mammals such as humans and kangaroos. Some of its novel morphological features closely resemble the evolutionary novelties of bipedal mammals: its compressed thorax and its elongate ischium resemble those of kangaroos; and the wide sternum resembles that of an orangutan, a bipedal primate that, like the two-legged goat, lacks a tail for support. A bipedal baboon filmed by William H. Hamilton III showed similar behavioral accommodation to disabled front legs (see West-Eberhard, 2003). The baboon's deformity is thought to have been caused by polio, for an epidemic had affected its troop.

Could phenotypic accommodation have played a key role in the morphological innovations of bipedal primates and kangaroos? As far as I know this question has never been answered, probably because it is seldom asked, although it was certainly suggested by the comparisons discussed by Slijper. It seems probable that plasticity has at least played a role, judging by the readiness with which mammals accommodate morphologically to behavioral alterations and extremes, as in human body builders and in potential osteoporosis patients, where weight-bearing exercise and a calcium-rich diet can have marked effects on the size and density of bone. Some of the changes described by Bramble and Lieberman (2004) as associated with the origin of a running gait in humans, including enlargement of the gluteus maximus muscle and elongation of certain bones of the legs, modification of the pelvis, and elongation of the Achilles tendon, could have appeared and then spread rapidly. Given natural selection (in whatever context) for increased running behavior in a human population of highly social adults and their imitative young, changes like those produced in the two-legged goat could come to characterize an entire population in a single generation (Slijper, '42a, b). Head stabilization and energetic efficiency, mentioned by Bramble and Lieberman (2004) as special problems during the evolution of hominoid running, increases in monkeys (Japanese macaques) trained over a period of years to walk upright (Hirasaki et al., 2004).

It is easy to see how a phenotypic accommodation could become a regularly occurring developmental pathway. To give just one example that involves an established trait of a natural population, the skulls of adult spotted hyenas (*Crocuta crocuta*) have a striking medial saggital crest and other attachment sites (cheekbones and forehead) for the powerful muscles used in chewing the bones and tough meat of the hyenas' prey (after Holekamp and Smale, '98; Holekamp, personal communication). The extreme modifications of the head are absent in infants of the same species, and they also fail to develop in captive individuals fed on soft diets. The full development of the exaggerated crest evidently requires years of chewing tough food. That is, normal species-specific adult morphology requires a particular kind of environmental factor—a tough diet—and the exercise that this promotes. It is also possible that, in addition, the skull bones have evolved under genetic accommodation of the response, to be especially responsive to such exercise to exaggerate special features of the skull—this is not known. But one thing is certain: a particular environmental factor (hard diet) and behavioral response (intense exercise of the jaw muscles during chewing) contributes to the normal development of the species-typical morphology.

Novel morphology that involves adaptive phenotypic accommodation is not ''random'' variation, for it begins with an adaptive phenotypic change. Phenotypic accommodation gives a head start to adaptive evolution by producing novel phenotypes likely to be favored by natural selection. In this respect, a theory of adaptive evolution that recognizes the role of phenotypic accommodation differs from one that views selection as operating on random variation due to mutation alone.

In sum, phenotypic accommodation facilitates adaptive evolution in two ways: (1) it provides a head start in adaptation. The new trait is produced by an already organized, adaptively flexible phenotype whose responses have been subjected to natural selection in the past. And (2) being a new developmental pathway associated with a developmental switch (the mutational or environmental inducer), the new trait is modular in nature. That is, it is somewhat independently expressed relative to other traits and therefore independently subject to selection (see West-Eberhard, '92, 2003). How adaptive evolution proceeds from this initial step of phenotypic accommodation is discussed in the next section.

A GENERAL MODEL FOR THE ORIGIN OF ADAPTIVE PHENOTYPIC NOVELTIES

The following model is intended to describe the evolutionary origin of all kinds of adaptive traits—morphological, physiological and behavioral, whether induced by a mutation or an environ-

mental factor—at all levels of organization. This is a brief summary of concepts presented in more detail and with more complete supporting evidence elsewhere (West-Eberhard, 2003):

(a) A novel input occurs which affects one (if a mutation) or possibly more (if environmental) individuals. Individuals may experience novel inputs due to evolution in another context (e.g., which moves them into a new environment, or has novel pleiotropic effects on the phenotype via other pathways).
(b) Phenotypic accommodation: Individuals developmentally responsive to the novel input immediately express a novel phenotype, for example, because the new input causes quantitative shifts in one or more continuously variable traits, or due to the switching off or on of one or more input-sensitive traits (causing a reorganization of the phenotype). Adaptive phenotypic adjustments to potentially disruptive effects of the novel input exaggerate and accommodate the phenotypic change without genetic change.
(c) Initial spread: The novel phenotype may increase in frequency rapidly, within a single generation, if it is due to an environmental effect that happens to be common or ubiquitous. Alternatively, if it is due to a positively selected mutation, or is a side effect of a trait under positive selection (Müller, '90), the increase in frequency of the trait may require many generations.
(d) Genetic accommodation (change in gene frequencies under selection): Given genetic variation in the phenotypic response of different individuals, the initial spread produces a population that is variable in its sensitivity to the new input, and in the form of its response. If the phenotypic variation is associated with variation in reproductive success, natural selection results; and to the degree that the variants acted upon by selection are genetically variable, selection will produce genetic accommodation, or adaptive evolutionary adjustment of the regulation and form of the novel trait.

This model requires that at least some individuals in a population are responsive to the new input. As already discussed, the capacity to respond to diverse inputs is likely a property of all living things. The model also depends, for an evolutionary response to selection, on the presence in the population of genetic variation for the developmental change. This also is a realistic assumption for most populations, given the well-documented commonness of genetic variation revealed by electrophoresis, and the evolutionary response of virtually all traits subjected to artificial selection (e.g., see Lewontin, '74; Maynard Smith, '89; West-Eberhard, 2003, 2005).

Standard quantitative genetics models show how threshold selection can lead to the change in frequency of a trait (e.g., see Falconer and MacKay, '96). Previous authors have presented ideas similar to those presented here regarding the origins of novel traits. Wagner and Chiu (2003, p 266), for example, wrote: ''the origin of new characters requires epigenetic opportunity for the new morphological states to occur [i.e., a novel input that provokes a developmental response]. Genetic factors are required for the heritability and subsequent fixation of new morphological states [one possible outcome of genetic accommodation]. This requirement does not imply, however, that the specific nature of a new character is in any sense determined or explained by the mutations that make the character heritable'' [as just discussed, the nature of the new character comes primarily from the reorganization of ancestral developmental pathways].

There is no conflict between this model and the standard view of adaptive evolution as involving variation, selection, and gene-frequency change. But the analysis includes steps of the process that are usually left out, steps with important implications that are sometimes overlooked, for example, the fact that it does not matter, for the initiation of a novelty, whether the original induction is mutational or environmental; and the fact that environmentally induced traits can initially spread without positive selection (all that is needed is recurrence of the inducing factor).

This departs from the view, which may be encouraged by gene-for-trait modeling of evolution by natural selection, that the recurrence or spread of a novel trait is due to the spread of a particular allele, and the associated idea that only genetically induced, mutational or genetically recombinant, novelties have evolutionary potential. Because environmental factors can affect many individuals at once they may be more effective initiators of selectable evolutionary novelties than mutations, which initially affect only single individuals (West-Eberhard, 2003, 2005). In effect, environmental induction jump-starts adaptive evolution by immediately producing a population of phenotypic

variants upon which selection can act. Then, if the phenotypic variance is partly due to genetic variation among individuals, adaptive evolution in response to selection can occur.

In this model, as in Waddington's ('53) concept of genetic assimilation, adaptive evolution depends on evolutionary change in the threshold for a newly induced response, as well as quantitative genetic change in the propensity or ''liability'' to produce the novel trait. But genetic accommodation differs from genetic assimilation in several important respects. For example, it considers genetic change in both the form (e.g., under selection for increased efficiency) and the frequency (e.g., due to change in threshold of expression) of a trait, whereas genetic assimilation treats only the latter. Most importantly, genetic accommodation can lead to a decline in trait frequency and diminished genetic control, or to the establishment of a polyphenism with adaptive conditional expression of alternative forms. Genetic assimilation, by contrast, implies the evolution of increased genetic control and the evolutionary change toward increased frequency or fixation of a trait. For a more detailed comparison of genetic assimilation, genetic accommodation, and the Baldwin effect see West-Eberhard (2003).

Could phenotypic accommodation alone, without gene-frequency change, lead to adaptive evolution? It is sometimes pointed out that developmental plasticity can lead to evolution without gene-frequency change, if the spread of an environmentally induced trait is entirely due to the spread of its environmental inducer, as in the fixation of a conditional alternative phenotype under conditions that induce it, or in the ''behavioral inheritence'' of traits in humans (Avital and Jablonka, 2000). But this would not be Darwinian adaptive evolution, or evolution by natural selection, the subject of this article. Darwin's theory of evolution by natural selection is based on the principle that the spread of a trait in a population is due to the *fitness effect* (advantageousness) of the trait. It is the increased reproductive success, or fitness, of the bearers of the trait that causes the trait to spread. The Darwinian fitness-effect condition is not met if the trait spreads due to the spread of its inducer alone, independent of the fitness effect of the trait. So evolution by increased environmental induction alone may be described as phenotypic evolution—a change in the frequency of a particular phenotype in a population—but not as Darwinian adaptive evolution.

RECIPROCAL CAUSATION IN THE EVOLUTION OF BEHAVIOR AND MORPHOLOGY

There is the potential for circular reinforcement in the evolution of morphology when it is affected by plasticity, especially behavioral plasticity and learning. Diet can affect morphology via phenotypic accommodation during use, and morphology in turn can affect diet, both by phenotypic accommodation due to learning and by making the new diet more profitable. Observations by Greenwood ('65) and others on the African Lake cichlids showed that individuals of a mollusc-feeding species reared on a soft diet in an aquarium develop pharyngeal jaw morphology like that of closely related soft-diet insectivorous species. Then, beginning in the mid 1970s, Liem and Kaufman ('84) demonstrated the reciprocal effect of morphology on diet. When two alternative morphs, one with a mollusc-specialized jaw (the molariform morph) and the other with a soft-diet jaw (the so-called papilliform morph), have an abundant supply of soft food, both prefer the soft diet. But when food is scarce they divergently specialize in accord with their morphological specializations: the fishes with the mollusc-feeder jaw morphology take a greater proportion of molluscs, and those with the soft-food morphology specialize on soft food. So, in conditions of scarcity, morphology affects diet and the resultant divergent behavior would reinforce selection in divergent directions.

A similar phenomenon is well documented in Darwin's finches of the Galapagos islands (Price, '87; Grant and Grant, '89): in times of food scarcity large-beaked finches learn to prefer and efficiently crack large hard seeds, while smaller-beaked individuals learn to concentrate on, and efficiently exploit, small soft seeds. This promotes intermittent diverging selection on the extremes, and generates divergent trends in different populations and species. There is, then, evidence that developmental plasticity in the form of morphology- and diet-associated learning has contributed to the explosive radiations in both the African cichlids and the Galapagos finches (West-Eberhard, 2003).

SIGNIFICANCE FOR RESEARCH ON THE ORIGINS OF MORPHOLOGICAL NOVELTY

A developmental-plasticity approach to the origins of novelty suggests new avenues of

research on the evolution of morphology. To explain the origins of morphological novelty, developmental biology needs to broaden its focus beyond genomic innovation to include behavior and even learning as key mechanisms in the evolution of morphology. These mechanisms need to be included in both microevolutionary and macroevolutionary analyses of change.

Microevolutionary analyses

From observations like those discussed here, I offer the following testable hypothesis: species-specific morphological novelties can result from the evolution, or environmental induction, of species-specific behaviors, and need not involve morphology-specific genetic change (though such change is likely to occur eventually, as genetic accommodation leads to the reorganization of gene expression in traits favored by selection).

This hypothesis suggests a broadened experimental approach to research on the origins of morphological novelty. Suppose you are interested in the origin of the exaggerated sagittal crest in hyenas. In a traditional approach, you might propose breeding experiments to ascertain the heritability of crest height in species that already have a sagittal crest. You might map cranial morphology onto a phylogeny to look for similar structures in related species. Both studies would illuminate the evolution of the crest. But an approach considering developmental plasticity might go further, to examine the correlation between dietary toughness and muscle and bone development, or to examine the possibility of crest induction (e.g., by dietary alteration) in related species that do not normally possess a raised crest.

There are, of course, some taxa in which such plasticity experiments have actually been done. One of the best known is the cichlid fishes, already discussed in the section on reciprocal causation (above). Following the discovery that diet affects feeding morphology in cichlids (Greenwood, '65), various investigators, including Liem and associates (Liem and Osse, '75; Liem and Kaufman, '84), Hoogerhoud ('86), Meyer ('87, '90), Wimberger ('91, '92), and Galis (Galis, '93; Galis et al., '94), experimentally examined the effects of diet on morphology in other cichlid species. These studies confirmed effects of diet on the pharyngeal jaw morphology. The Central American cichlid *Cichlasoma citrinellum* has two trophic morphs: one that feeds primarily on snails and another that has a softer diet. Meyer ('90) found that the alternative pharyngeal jaw morphologies of the two morphs can be reversed in at least some individuals by reversing their diets. He also pointed out that these two body forms parallel the differences between two alternative forms that are very common in fishes, the snail feeder having a body shape like a ''benthic'' or bottom-feeding form, and the soft-food morph resembling a ''limnetic'' form that feeds in the water column. These studies support the hypothesis that recurrent trophic morphologies in cichlids can arise due to phenotypic accommodation under different dietary regimes.

Macroevolution, or major morphological change

Macroevolution may be different in scale to microevolutionary change, but it still requires explanation at a microevolutionary level. That is, it requires explanation in terms of adaptive evolution by natural selection and gene-frequency change within populations. No matter how major the trait, no matter how momentous at the family or phylum or body-plan level, analysis still has to go to the microevolutionary level to find out how the trait began. This suggests another kind of new avenue of research for developmental biologists interested in macroevolutionary aspects of evolution.

To cite just one example, consider the likely role of developmental plasticity in the origin of an undeniably major morphological novelty—a new appendage in a fly. In some genera of sepsid flies (Diptera, Sepsidae), a novel appendage is formed by the fourth sternite of the males. It has evolved independently in several different genera (Eberhard, 2001). In relatively unspecialized species (e.g., *Archisepsis diversiformis*), males have sternal bristles that are rubbed against the female during courtship. In somewhat more elaborate versions (e.g., in an unnamed species of *Pseudo-palaeosepsis*), male sternites have bristled lateral lobes that are semi-articulated and have attached muscles capable of moving them back toward the posterior end of the fly. And in the most highly elaborated examples (e.g. *in Pseudo-palaeosepsis nigricoxa*), the sternal lobes are long, highly articulated, and capable of limb-like movements both toward the posterior and ventrally, forming a novel appendage complete with segments, muscles, and nerves.

Phenotypic flexibility has likely played an important role in the evolution of this hinged,

limb-like structure. First, behavioral movements have evidently taken the lead with abdominal courtship movements preceding the morphological specializations. Then, in somewhat more specialized species, where the lateral lobes are defined, a break in the cuticle allows its pre-existing flexibility or bendability to play a role in the versatility of the males' courtship movements (Eberhard, 2001). The increased modularity of the sternite—now two pieces rather than one—contributes to its flexibility.

The ease with which muscle can be recruited to (or exaggerated at) new attachments, as exemplified in the two-legged goat described earlier, and in these flies, is impressive. But the mechanisms must be different in the flies, where individuals emerge from complete metamorphosis with their adult appendages fully formed and presumably unmodified by exercise. During their development, the walking legs of insects begin as rudiments that grow and then are folded and grooved where they will later become segmented (Chapman, '98, p 343)—a sequence that is not unlike that suggested by the appendage-like lobes of sepsid flies, where the simpler arrangement is a bendable groove or notch, and the more specialized form an articulated structure. It would be of interest to know whether pupal movements play any role in the development of adult insect muscle and cuticular morphology.

Could locomotory appendages like legs or wings have started by a process something like that observed in the diversification of sepsid courtship devices? And if they did, at what point during appendage evolution might the major genes associated with such structures have been co-opted for their development? At what point would you expect to have the newly independent modular parts associated with their own imaginal disks? Such questions cannot be answered, or even asked, in studies of the development of fully formed appendages like those of *Drosophila*. But Julia Bowsher, a graduate student at Duke, is beginning to answer them using sepsid flies. In a study on the developmental genetics of the sternal lobe of *Themira biloba*, a species whose males have an intermediate degree of specialization, possessing a semi-articulated sternal lobe but not a segmented articulated appendage, Bowsher has discovered that at least three genes—engrailed, extradenticle, and notch—which are expressed during the development of the lobes are also expressed during genital appendage development in *Drosophila*. These genes have evidently been co-opted in the development of the novel lobes. In *T. biloba*, the expression of these genes in the lobes occurs at the same time as their expression in the genitalic appendages, and well after sternite patterning, further supporting the interpretation that ancestral appendage genes have been co-opted for expression of a new appendage-like trait. The lobes of this species develop from a cluster of abdominal histoblasts, not from genital imaginal discs, or from any imaginal disc of their own, though the nests of histoblasts are imaginal-disc-like in being set aside during early development, and then proliferating and differentiating to form a specific distinctive structure.

Developmental plasticity and novel morphology under sexual selection

Sexual selection is noted for its ability to produce extreme morphological novelties (Darwin, 1871 [1874]). We often assume that natural selection—survival selection—is responsible for novelty, but we may need to look more closely at how novel structures are used. It is quite possible that limbs, especially appendages like wings in insects and tetrapods, were originally used in displays that evolved under sexual selection, even though they are now associated with survival selection due to their obvious importance in flight.

Developmental plasticity under sexual selection may have affected the diversity of the mouths of African-lake cichlids, contributing to their rapid and extreme radiations in African Lakes Victoria and Malawi (e.g., Greenwood, '64). The cichlid radiations are a story of diversification in teeth, jaws and mouths, so it easy to assume that these aspects of the radiation are entirely explained as trophic innovations. But male cichlids also fight and court using their mouths (Baerends and Baerends-van Roon, '50). They employ behaviors that require extreme development of the muscles that are also used in feeding, and they have been described as trembling like straining acrobats when they opened their mouths wide in nuptial and aggressive threat displays (Baerends and Baerends-van Roon, '50). Such extreme behavior could not help but have affected the form of their flexible and muscles and bones, and would favor the genetic variants best able to respond. Novel social inputs, as well as novel inputs from the non-social environment, could lead to novel or exaggerated behavioral responses and their morphological accommodation.

DISCUSSION AND CONCLUSIONS

One possible objection to the arguments made here is that the traits formed by phenotypic accommodation and novel combinations of ancestral traits are not truly new. Is all of evolution just shifting and accommodating the pieces? If rigidly circumscribed modularity of structures were the rule then the moving-the-pieces objection might hold. But, as shown by the examples described here, when phenotypic accommodation involves the re-use of old pieces in new places, as seen in the co-option of muscles and the remodeling of bone in the two-legged goat, and of ancestral genes in the novel appendages of sepsid flies, the new morphologies are substantially changed in shape and dimensions as well as in the way they are put together. Even mutational genomic change often involves reorganization, duplication and recombination of parts (examples and references in West-Eberhard, 2003), and yet we do not hesitate to call mutations true genetic novelties. As with the concept of homology, the problem is not simple (for discussion of homology relating especially to the nature of innovation, see Müller, 2003; Hall, 2003; West-Eberhard, 2003).

By the broad definition of innovation discussed near the beginning of this article, phenotypic accommodation, including behavioral accommodation and even learning, can be an important source of morphological novelty because it permits immediate reorganization of phenotypes responsive to novel inputs from environment and genome. Although the components of a reorganized phenotype are not themselves new, the combination that makes it distinctive compared to recent ancestors is new, and the components are newly subject to selection in a new context. There is, therefore, some justification for considering novelties due to phenotypic accommodation, once they have been subjected to selection and genetic accommodation, to be true evolutionary innovations.

All novel traits, including macroevolutionary ones, have to be explained in terms of the developmental generation of variation and ultimately in the context of selection within populations, beginning with individuals and species that lack the novel trait. A plausible transition hypothesis, showing how the ancestral phenotype was transformed to produce a novel form, is an important though neglected part of evolutionary biology.

ACKNOWLEDGMENTS

I thank Julia Bowsher, John Skoyles, and Neal Smith for drawing my attention to recent findings relevant to this article, and Julia Bowsher, W.G. Eberhard and two anonymous reviewers for helpful comments.

LITERATURE CITED

Avital E. Jablonka E. 2000. Animal traditions: behavioural inheritance in evolution. Cambridge, UK: Cambridge University Press.

Baerends GP, Baerends-van Roon JM. 1950. An introduction to the study of the ethology of cichlid fishes. Behav Suppl 1:1–243.

Baldwin JM. 1896. A new factor in evolution. Am Nat 30:441–451, 536–553.

Baldwin JM. 1902. Development and evolution. New York: Macmillan.

Bernays EA. 1986. Diet-induced head allometry among foliage—chewing insects and its importance for graminivores. Science 231:495–497.

Bradshaw AD. 1965. Evolutionary significance of phenotypic plasticity in plants. Adv Gen 13:115–155.

Bramble DM, Lieberman DE. 2004. Endurance running and the evolution of *Homo*. Nature 432:345–352.

Chapman RF. 1998. The insects. Cambridge, MA: Cambridge University Press.

Darwin C. 1871 [1874]. The descent of man and selection in relation to sex, 2nd edn. [unabridged but re-numbered text, and figures]. New York: The Modern Library, Random House.

Eberhard WG. 2001. Multiple origins of a major novelty: moveable abdominal lobes in male sepsid flies (Diptera: Sepsidae), and the question of developmental constraints. Evol Dev 3:206–222.

Falconer DS, Mackay TRC. 1996. Introduction to quantitative genetics, 4th edn. Essex: Longman.

Frazzetta TH. 1975. Complex adaptations in evolving populations. Sunderland, MA: Sinauer.

Galis F. 1993. Interactions between the pharyngeal jaw apparatus, feeding behaviour, and ontogeny in the cichlid fish, *Haplochromis piceatus*: a study of morphological constraints in evolutionary ecology. J Exp Zool 267: 137–154.

Galis F, Terlouw A, Osse JWM. 1994. The relation between morphology and behaviour during ontogenetic and evolutionary changes. J Fish Biol 45(Suppl. A):13–26.

Gerhart J, Kirschner M. 1997. Cells, embryos, and evolution: toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Malden, MA: Blackwell.

Goldschmidt R. 1940 [1982]. The material basis of evolution. New Haven: Yale University Press.

Grant BR, Grant PR. 1989. Evolutionary dynamics of a natural population. Chicago: University of Chicago Press.

Greenwood PH. 1964. Explosive speciation in African lakes. Proc R Inst 40:256–269.

Greenwood PH. 1965. Environmental effects on the pharyngeal mill of a cichlid fish, *Astatoreochromis alluaudi*, and their taxonomic implications. Proc Linn Soc Lond 176:1–10.

Hall BK. 2003. Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. Biol Rev 78:409–433.

Hirasaki E, Ogihara N, Hamada Y, Kumakura H, Nakatsukasa M. 2004. Do highly trained monkeys walk like humans? A kinematic study of bipedal locomotion in bipedally trained Japanese Macaques. J Hum Evol 46:739–750.

Holekamp KE, Smale L. 1998. Behavioral development in the spotted hyena. BioScience 48:997–1005.

Hoogerhoud RJC. 1986. Ecological morphology of some cichlid fishes. Thesis, Leiden: University of Leiden.

Kirschner MW. 1992. Evolution of the cell. In: Grant PR, Horn HS, editors. Molds, molecules and metazoa: growing points in evolutionary biology. Princeton: Princeton University Press.

Lewontin RC. 1974. The genetic basis of evolutionary change. New York: Columbia University Press.

Liem KF, Kaufman LS. 1984. Intraspecific macroevolution: functional biology of the polymorphic cichlid species *Cichlasoma minckleyi*. In: Echelle AA, Kornfield I, editors. Evolution of fish species flocks. Orono: University of Maine at Orono Press. p 203–215.

Liem KF, Osse JWM. 1975. Biological versatility, evolution, and food resource exploitation in African cichlid fishes. Am Zool 15:427–454.

Maynard Smith J. 1989. Evolutionary genetics. New York: Oxford University Press.

Mayr E. 1959. The emergence of evolutionary novelties. In: Tax S, editor. Evolution after Darwin, Volume One. Chicago: University of Chicago Press. p 349–380.

Meyer A. 1987. Phenotypic plasticity and heterochrony in *Cichlasoma managuense* (Pisces, Cichlidae) and their implications for speciation in cichlid fishes. Evolution 41:1357–1369.

Meyer A. 1990. Ecological and evolutionary consequences of the trophic polymorphism in *Cichlasoma citrinellum* (Pisces: Cichlidae). Biol J Linn Soc 39:279–299.

Müller GB. 1990. Developmental mechanisms at the origin of morphological novelty: a side-effect hypothesis. In: Nitecki MH, editor. Evolutionary innovations. Chicago: University of Chicago Press. p 99–132.

Müller GB. 2003. Homology: the evolution of morphological organization. In: Müller GB, Newman SA, editors. Origination of organismal form: beyond the gene in developmental and evolutionary biology. Cambridge, MA: MIT Press. p 51–69.

Müller GB, Wagner GP. 1991. Novelty in evolution: restructuring the concept. Ann Rev Ecol Syst 22:229–256.

Price T. 1987. Diet variation in a population of Darwin's finches. Ecology 68:1015–1028.

Schlichting CD. 1986. The evolution of phenotypic plasticity in plants. Ann Rev Ecol Syst 17:667–693.

Schmalhausen II. 1949 [1986]. Factors of evolution. Chicago: University of Chicago Press.

Slijper EJ. 1942a. Biologic-anatomical investigations on the bipedal gait and upright posture in mammals, with special reference to a little goat, born without forelegs. I. Proc Konink Ned Akad Wet 45:288–295.

Slijper EJ. 1942b. Biologic-anatomical investigations on the bipedal gait and upright posture in mammals, with special reference to a little goat, born without forelegs II. Proc Konink Ned Akad Wet 45:407–415.

Strathmann RR, Fenaux L, Strathmann MF. 1992. Heterochronic developmental plasticity in larval sea urchins and its implications for evolution of non-feeding larvae. Evolution 46:972–986.

Sultan S. 1987. Evolutionary implications of phenotypic plasticity in plants. Evol Biol 20:127–178.

Sultan S. 2000. Phenotypic plasticity for plant development, function and life history. Trends Plant Sci 5:537–542.

Waddington CH. 1953. Genetic assimilation of an acquired character. Evolution 7:118–126.

Wagner GP, Chiu C-h. 2003. Genetic and epigenetic factors in the origin of the tetraped limb. In: Müller GB, Newman SA, editors. Origination of organismal form: beyond the gene in developmental and evolutionary biology. Cambridge, MA: MIT Press. p 265–285.

Wake DB, Roth G. 1989. The linkage between ontogeny and phylogeny in the evolution of complex systems. In: Wake DB, Roth G, editors. Organismal functions: integration and evolution in vertebrates. New York: Wiley. p 361–377.

West-Eberhard MJ. 1992. Behavior and evolution. In: Grant PR, Horn H, editors. Molds, molecules and metazoa: growing points in evolutionary biology. Princeton: Princeton University Press. p 57–75.

West-Eberhard MJ. 1998. Evolution in the light of developmental and cell biology, and vice versa. Proc Nat Acad Sci USA 95:8417–8419.

West-Eberhard MJ. 2003. Developmental plasticity and evolution. New York: Oxford University Press.

West-Eberhard MJ. 2005. Developmental plasticity and the origin of species differences. Proc Natl Acad Sci USA 102(Suppl. 1):6543–6549.

Wimberger PH. 1991. Plasticity of jaw and skull morphology in the neotropical cichlids *Geophagus brasiliensis* and *G. steindachneri*. Evolution 45:1545–1563.

Wimberger PH. 1992. Plasticity of fish body shape. The effects of diet, development, family and age in two species of *Geophagus* (Pisces: Cichlidae). Biol J Linn Soc 45: 197–218.

Wimberger PH. 1994. Trophic polymorphisms, plasticity, and speciation in vertebrates. In: Stouder DJ, Fresh KL, Feller RJ, editors. Theory and application of fish feeding ecology. Columbia: University of South Carolina Press. p 19–43.

28 June 1985, Volume 228, Number 4707 SCIENCE

The Sociogenesis of Insect Colonies

Edward O. Wilson

Summary. Studies on the social insects (ants, bees, wasps, and termites) have focused increasingly on sociogenesis, the process by which colony members undergo changes in caste, behavior, and physical location incident to colonial development. Caste is determined in individuals largely by environmental cues that trigger a sequence of progressive physiological restrictions. Individual determination, which is socially mediated, yields an age-size frequency distribution of the worker population that enhances survival and reproduction of the colony as a whole, typically at the expense of individuals. This "adaptive demography" varies in a predictable manner according to the species and size of the colony. The demography is richly augmented by behavioral pacemaking on the part of certain castes and programmed changes in the physical position of colony members according to age and size. Much of what has been observed in these three colony-level traits (adaptive demography, pacemaking, and positional effects) can be interpreted as the product of ritualization of dominance and other forms of selfish behavior that is still found in the more primitive insect societies. Some of the processes can also be usefully compared with morphogenesis at the levels of cells and tissues.

Together with flight and metamorphosis, colonial life was one of the landmark events in the evolution of the insects and evidently served as a source of their ecological success. Preliminary studies indicate that approximately one-third of the entire animal biomass of the Amazonian terra firme rain forest may be composed of ants and termites, with each hectare of soil containing in excess of 8 million ants and 1 million termites (*1, 2*). On the Ivory Coast savanna the density of ants is 20 million per hectare, with one species, *Camponotus acvapimensis*, alone accounting for 2 million (*3*). Such African habitats are often visited by driver ants (*Dorylus* spp.), single colonies of which occasionally contain more than 20 million workers (*4*). And the driver ant case is far from the ultimate. A "supercolony" of the ant *Formica yessensis* on the Ishikari Coast of Hokkaido was reported to be composed of 306 million workers and 1,080,000 queens living in 45,000 interconnected nests across a territory of 2.7 square kilometers (*5*).

The environmental impact of these insects is correspondingly great. In most terrestrial habitats ants are among the leading predators of insects and other small invertebrates (*3, 6, 7*), and leafcutter ants (*Atta* spp.) are species for species the principal herbivores and most destructive insect pests of Central and South America (*8*). *Pogonomyrmex* and other harvester ants compete effectively with mammals for seeds in deserts of the southwestern United States (*9*). Other ants move approximately the same amount of soil as earthworms in the woodlands of New England, and they surpass them in tropical forests. Both are exceeded in turn by termites, which also break down a large part of the vegetable litter and diffuse the products through the humus (*10, 11*).

Edward O. Wilson is Frank B. Baird, Jr., Professor of Science and Curator in Entomology, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138. This article is based on the lecture of the 1984 Tyler Prize for Environmental Achievement, delivered at the University of Southern California on 24 May 1984.

The Reasons for Success

In general, the most abundant social insects are the evolutionarily more advanced groups of ants and termites, in other words, those with the highest percentage of derived traits in anatomy and physiology as well as the more populous and complexly organized societies (*6, 12, 13*). What is the real origin of this competitive advantage in the environment as a whole? At the risk of oversimplification, it can be said that entomologists have come to recognize three qualities as being most important. First, coordinated groups conduct parallel as opposed to serial operations and hence make fewer mistakes, especially when labor is divided among specialists. If different cadres of workers in an ant colony simultaneously forage for food, feed the queen, and remove her eggs to a safe place, they are more likely as a whole to complete the operation than if they perform the steps in repeated sequences in the manner of solitary insects (*13*). Second, groups can concentrate more energy and force at critical points than can single competitors, using sheer numbers to construct nests in otherwise daunting terrain, as well as to defend the young, and to retrieve food more effectively. Finally, there is caste: in ways that vary among species, the food supply is stabilized by the use of larvae and special adult forms to store reserves in the form of fat bodies and nutrient liquids held in the crop, while defense, nest construction, foraging, and other tasks are mostly accomplished by specialists (*14*).

The aim of much of contemporary research on social insects is to identify more fully the mechanisms by which colony members differentiate into castes and divide labor—and to understand why certain combinations of mechanisms have produced more successful products than others. The larger hope is that more general and exact principles of biological organization will be revealed by the meshing of comparable information from developmental biology and sociobiology. The definitive process at the level of the organism is morphogenesis, the set of procedures by which individual cells or cell populations undergo changes in shape or position incident to organismic development (*15*). The definitive

process at the level of the colony is sociogenesis, the procedures by which individuals undergo changes in caste, behavior, and physical location incident to colonial development. The question of interest for general biology is the nature of the similarities between morphogenesis and sociogenesis.

The study of social insects is by necessity both a reductionistic and holistic enterprise. The behavior of the colony can be understood only if the programs and positional effects of the individual members are teased apart, ultimately at the physiological level. But this information makes full sense only when the patterns of colonial behavior of each species are examined as potential idiosyncratic adaptations to the natural environment in which the species lives. At both levels social insects offer great advantages over ordinary organisms for the study of biological organization. Although no higher organism can be readily dissected into its constituent parts for study and then reassembled, this is not the case for the insect colony. The colony can be fragmented into any conceivable combination of sets of its members, manipulated experimentally, and reconstituted at the end of the day, unharmed and ready for replicate treatment at a later time. The technique is used for analysis of optimization in social organization as follows: the colony is modified by changing caste ratios, as though it were a mutant. The performance of this "pseudomutant" is then compared with that of the natural colony and other modified versions. The same colony can be turned repetitively into pseudomutants in random sequences on different days, eliminating the variance that would otherwise be due to between-colony differences (*16*). At a still higher level of explanation, that of the ecosystem, the large numbers of species of various kinds of social insects (more than 1000 each in the ant genera *Camponotus* and *Pheidole* alone) give a panoramic view of the evolution of colonial patterns and make correlative analysis of adaptation more feasible.

Principles of Sociogenesis

In all species of social insects thus far studied, caste differences among colony members have proved to be principally or exclusively phenotypic rather than genetic. The environmental factors in each instance belong to one or more of the following six categories: larval nutrition (which is especially important in ants); inhibition caused by pheromones or other stimuli from particular castes (the key factor in many kinds of termites); egg size and hence quantity of nutrients available to the embryo; winter chilling; temperature during development; and age of the queen (*6, 17, 18*). Phenotypic caste determination is similar to restriction during cell differentiation. That is, the growing individual reaches one or more decision points at which it loses some of its potential, and this diminution continues progressively until it reaches the final decision point, where it is determined to the caste it will occupy as an adult. For example, in the ant genus *Pheidole* the restriction to either the queen line or worker line occurs in the egg; then larvae in the worker line become committed to development as either minor or major workers in the fourth and final instar. The cues affecting these two decisions, which include nutrition, winter chilling of queens, and inhibitory pheromones, are mediated to the developing tissue by juvenile hormone (*19, 20*).

The differentiation of the colony members into physical castes is supplemented in the great majority of social species by a regular progression on the part of most workers through different work roles during aging. In this way the individual belongs not only to one physical caste but to a sequence of temporal castes as it passes through its life-span. By far the most common sequence is for the worker to join in the care of the queen or immature stages shortly after it emerges into the adult stage, then to participate in nest building, and, finally, to forage outside the nest for food. Temporal castes are a derived trait in evolution, having become most clearly demarcated in species with the largest societies. They are typically weak or absent in anatomically primitive species with small colony populations (*6, 21*).

Although individual workers are flexible with respect to caste at the start of their personal development in the egg stage, the colony as a whole is rigidly limited to a single array of castes. Each species also has a particular size-frequency distribution of adult workers (*13, 22, 23*). Workers in the ant genus *Pheidole*, for example, are divided into two subcastes, the minors and the majors, by size and body proportions. Among ten species selected for their taxonomic diversity, the majors were found to range from 3 percent in *Pheidole distorta* to 25 percent in *Pheidole minutula* (*23*). A lesser amount of variation exists among colonies belonging to the same species, and recent work suggests indirectly that some of the variation is genetic. Seven colonies of *Pheidole dentata* raised under uniform laboratory conditions through three brood cycles maintained relatively constant major worker percentages, and these levels varied significantly among the colonies, from approximately 5 to 15 percent (*24*).

The size-frequency distribution can also persist through relatively long periods of geological time. A fragment of a colony of the extinct weaver ant *Oecophylla leakeyi* preserved intact from the African Miocene (the only fossil insect society collected to date) proved to have the same distinctive pattern as the two living species of the genus, *Oecophylla longinoda* and *Oecophylla smaragdina*. In particular, the frequency curve was sharply bimodal, with the major workers somewhat more numerous than the minors and with a small number of medias connecting the two moieties. The allometry, or disproportionate variation in body parts, is also similar between the extinct and living species (*25*).

These several lines of evidence have led to the hypothesis of adaptive demography (*13, 26*), which can be summarized as follows. The vast majority of insect, vertebrate, and other animal populations evolve primarily through selection at the level of the individual organism. As a consequence, survivorship curves and natality schedules are directly adaptive, whereas the age-frequency distribution of the population as a whole emerges as an epiphenomenon. In the advanced social insects, in contrast, selection occurs primarily at the level of the colony, with workers mostly or entirely eliminated from reproduction and colonies competing against one another as compact units. Colonies whose members possess the most effective age-frequency distribution are more likely to survive and to reproduce, regardless of the fate of individual colony members. It is generally believed that the workers will increase the replication of genes identical to their own by promoting the physical well-being of the colony, even if they sacrifice themselves to achieve this end. Hence the age-frequency distribution of the colony members is directly subject to natural selection. Survivorship and natality schedules are indirectly subject to natural selection, in the sense of being shaped according to the effect they have on the age-frequency distribution of the colony as a whole.

The adaptive demography hypothesis has begun to be tested by both correlative analysis and experimentation. For

example, linear programming models predict that as a caste specializes, its members should decrease in proportion within the colony membership (*26*). This relation does hold among the species of *Pheidole* so far studied: the repertory size of the major caste is correlated significantly across species with the percentage of the majors in the worker force. Put another way, as the majors perform fewer tasks and devote more time proportionately to roles for which they are anatomically specialized, they become scarcer in the colony population (*23*).

And yet the major workers of *Pheidole* retain a remarkable flexibility. When the minor-major ratio was experimentally reduced to below 1:1 in three widely different species of the genus, the majors increased the number of kinds of acts they performed by as much as 4.5 times and their rate of activity 15 to 30 times. The change occurred within 1 hour of the ratio change and was reversed in comparably short time when the original ratio was restored. Thus the major workers were found to respond in a manner reminiscent of the genome of a somatic cell. Under normal circumstances most of their brain programs are silent: the active repertory is limited in a fashion appropriate to the tasks for which the majors are anatomically specialized. But when an emergency arises a much larger program is quickly summoned, the majors supply about 75 percent of the activity of the missing minors, and as a result the colony continues to feed and grow (*23*).

A second line of evidence of adaptive demography has been provided by studies of the leafcutter ant *Atta cephalotes*. New colonies of *Atta*, like those of most kinds of ants, are founded by single queens after the nuptial flights. These individuals dig a shaft into the ground, then eject a wad of symbiotic fungus from their mouths onto the ground and fertilize the hyphae with droplets of feces. During the next 6 weeks they rear the first brood of workers with reserves from their own bodies while bringing the small garden to flourishing condition. The queens have only enough ovarian yolk and other storage materials to rear one small group to maturity. In order for the colony to survive thereafter, the workers must range in size from a head width of 0.8 mm, which is small enough to culture the fungus, through 1.6 mm, which is just large enough to cut fresh leaves for the fungal substrate. It turns out that the first brood of workers possess a nearly uniform frequency distribution from 0.8 through 1.6 mm, which comes close to maximizing the number of individuals and at the same time achieves the minimum size range required to grow the fungus on which the colony depends (*27*).

As the leafcutter population expands afterward, the size-frequency distribution of the workers changes in dramatic fashion. The range is increased at both ends and the curve becomes strongly skewed toward the media and major worker classes (Fig. 1). An interesting question then arises: suppose that by some misadventure most of the population of a leafcutter colony were destroyed, reducing it to near the colony-founding state. Would the size-frequency distribution of new workers produced by the colony be characteristic of the beginning stage, or would it remain at the older stage? In other words, which is the more important in the ontogeny of the caste system, the size of the colony or its age? If age were more important, causing much of the available energy to be invested in workers larger than the minimum required to harvest leaves, the colony would be imperiled because of a

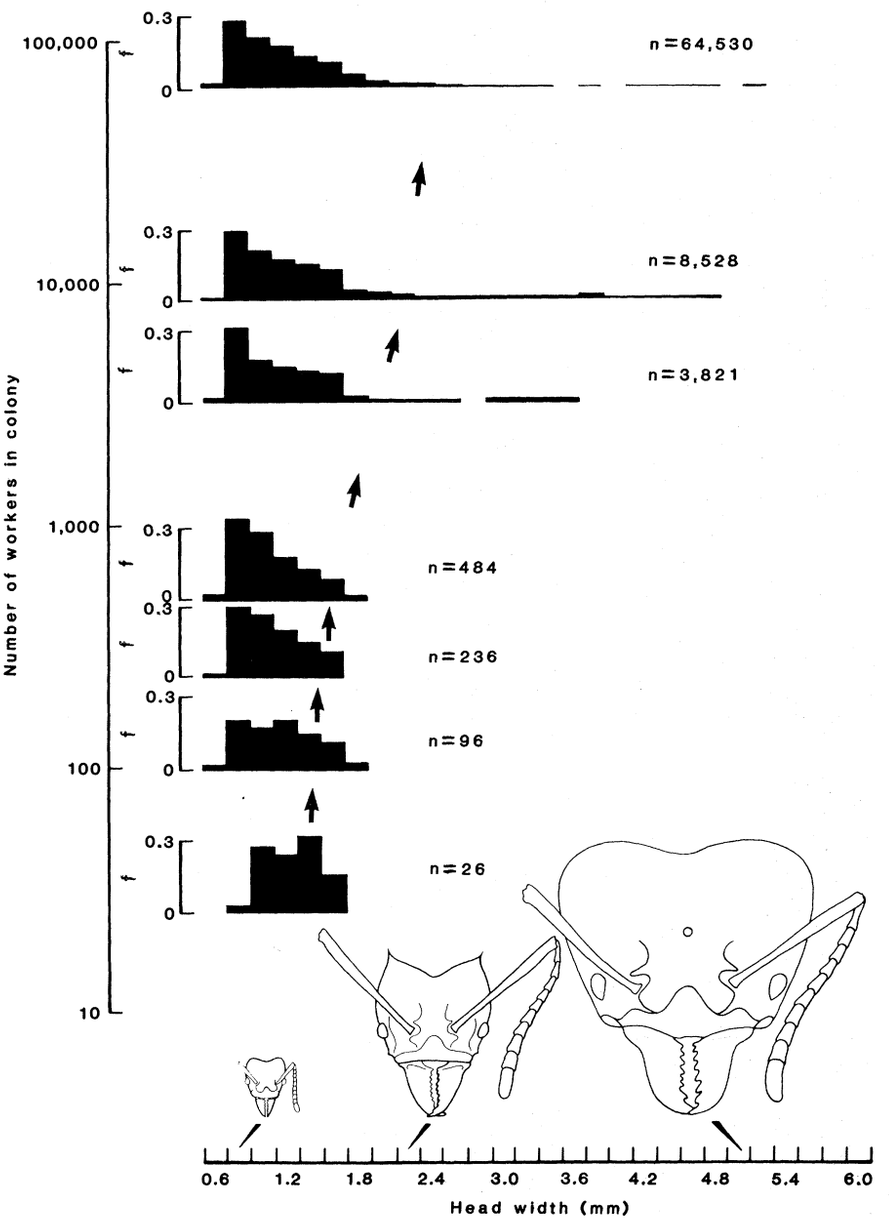


Fig. 1. The ontogeny of the caste system of the leafcutter ant *Atta cephalotes*, illustrated by seven representative colonies collected in the field or reared in the laboratory. The worker caste is differentiated into subcastes by continuous size variation associated with disproportionate growth in various body parts. The number of workers in each colony (*n*) is based on complete censuses; *f* is the frequency of individuals according to size class. The heads of three sizes of workers are shown in order to illustrate the disproportionate growth. Modified from Wilson (*27*).

shortage of the small gardener classes. The creation of just one new major worker, possessing a body weight 300 times that of a gardener worker, would bankrupt the already impoverished colony. In order to provide an answer, I selected four colonies 3 to 4 years old and with about 10,000 workers and reduced the population of each to 236, giving them an artificially imposed juvenile size-frequency distribution. The worker pupae produced at the end of the first brood cycle possessed a size-frequency distribution like that of small, young colonies rather than larger, older ones. Thus colony size is more important than age, and "rejuvenated" colonies are prevented from extinguishing themselves through an incorrect investment of their resources (*27*).

Such programmed resiliency implies the existence of control mechanisms operating at the level of the colony during population growth. An increasing fraction of the research on social insects is now being directed at the discovery of such mechanisms. This work has begun to reveal a fascinating pattern of feedback loops, pacemakers, and positional effects.

An example of negative feedback is provided by the events leading to the fission of honeybee colonies. The queen secretes a "queen substance," *trans*-9-keto-2-decenoic acid, which under most circumstances inhibits the construction of royal cells by the workers and hence the rearing of new queens (*28*). However, in large, freely growing colonies this pheromone must be supplemented by a second substance, the footprint pheromone, which is secreted in relatively large amounts from glands in the fifth tarsal segment of the queen. When bee colonies become overcrowded, the queen is unable to walk along the bottom edges of the comb, where the royal cells are ordinarily built. As a result the inhibition fails in that zone, the cells are built, and the colony reproduces. With the population density now reduced to below threshold density, the queen is able to resume her inhibitory control (*29*).

Most such controls are negative and hence contribute to physiological stability and smooth growth cycles within the colony. What appear to be properties of positive feedback and explosive chain reactions nevertheless do occur during nest evacuation in a few species. When attacking fire ant workers press closely on nests of the ant *Pheidole dentata*, the defending minor workers start laying odor trails back into the brood area. This causes excited movement through the nest and further bouts of recruitment. At the height of this expanding activity the workers and queen suddenly scatter from the nest and seek individual cover. When the fire ants are then experimentally removed, the *Pheidole* adults return to the nest and reoccupy it (*30*).

The coordination of activity is still imperfectly understood. Although the typical insect society is not quite the "feminine monarchie" envisioned by early entomologists (*31*), it is also much more than a republic of specialists. According to the species, certain immature stages and castes function as pacemakers and coordinators of colony activity. Ant larvae are specially effective in initiating foraging and nest construction by the adult workers. In army ants (*Eciton*), the hatching of larvae triggers the monthly nomadic cycle during which the entire colony marches to a new location daily (*32*). But in the great majority of other species thus far studied it is the queen that provides the maximum regulation. In more primitive societies, such as those of bumblebees (*Bombus*) and paper wasps (*Polistes*), she physically dominates her daughters and other females occupying the nests, prevents them from laying eggs, and by these actions forces most into foraging and other nonreproductive tasks. Such influence can transcend simple displacement. For instance, the presence of the queen of *Polistes fuscatus*, probably a typical species at this evolutionary level, increases and synchronizes overall worker activity (*33*). In carpenter ants (*Camponotus*), the mother queen is the principal source of the nest odor (*34*). When she is removed, the workers, now in a more chaotic state, fall back on odor cues emanating from their own bodies (*35*).

Workers of social insects move to different positions with reference to the queen and brood according to their ages. This pattern is usually centrifugal: soon after the worker emerges from the pupa into the adult stage, it attends the queen and immature stages, then drifts toward the outer chambers to assist in nest construction, and finally devotes itself primarily to foraging outside the nest. The progression is accompanied by physiological change. The details vary greatly among species, and even among members of the same colony, but in general the ovaries reach maximum development early in adult life, along with fat bodies and exocrine glands devoted to nutritive exchange (*6, 17, 36–38*). Afterward these tissues regress more than enough to counterbalance the growth of exocrine glands associated with nest construction and foraging, so that the worker declines overall in weight. Mortality due to accidental causes increases sharply among workers when they commence foraging. But this attrition has far less effect on the size-and-age structure of the worker population than if individuals commenced foraging early in life, because the natural life-span is curtailed in any case past the onset of foraging by physiological senescence. In the best documented case, the honeybee worker born in early summer typically begins foraging at 2 to 3 weeks of adult life and dies from senescence by 10 weeks into this period (*39*).

The workers of advanced insect societies are not unlike cells that emigrate to new positions, transform into new types, and aggregate to form tissues and organs. With relatively small adjustments in response thresholds according to size and age, intricate new patterns are created at the level of the colony. In the fungus-growing termite *Macrotermes subhyalinus*, for example, 90 percent of the foragers are large major workers past 30 days of age. Younger major and minor workers accept the grass collected by these foragers, consume it, and pass the partly digested material out into the fungus comb. Workers of various castes older than 30 days eat the fungus comb and produce the final feces (*40*). In the leafcutting ant *Atta sexdens* most of the fresh vegetation is gathered by workers of intermediate size (which, incidentally, achieve the highest net energetic yield of all the size groups). The material is then converted into new fungus substrate within the nest by an assembly-line operation that penetrates ever more deeply into the combs: successively smaller workers cut the leaves into tiny fragments, chew them into pulp, stick the processed lumps onto the growing combs, and transfer strands of fungi onto this newly prepared substrate. Finally, the smallest workers of all care for the proliferating fungus, virtually strand by strand (*16, 41*).

Such patterns are in fact much more intricate than a description of sequences alone indicates. In the ant *Pheidole dentata* and the honeybee *Apis mellifera* the tasks are broken into sets that are linked not by the similarity of the behaviors performed but by the proximity of the objects to which they are directed, thus reducing the travel time and energy expenditure of the individual workers (Figs. 2 and 3). The similarities between the two patterns can only be due to convergent evolution, since ants and bees arose during Mesozoic times from widely different stocks of aculeate wasps (*42*).

The Imperfection of Insect Societies

Although insects as a whole originated at least 350 million years ago, higher social insects did not appear until the Jurassic Period, roughly 200 million years ago, and they began an extensive evolutionary radiation only in the late Cretaceous and early Tertiary Periods, about 100 million years later (*42*). Even then, advanced social organization originated in as few as 13 stocks, 12 within the aculeate Hymenoptera (ants, bees, and wasps) and one in the cockroach-like orthopteroids that produced the termites (*6*).

Two possible explanations for this evolutionary conservatism have emerged from more detailed studies of individual colony members. The first is that the small size of the insect brain and the heavy reliance of social forms on chemical signaling place inherent limits on the amount of information flow through the colonies. This circumstance leads to frequent near-chaotic states and the dependence on colony decision-making by *force majeure*, a statistical preponderance of certain actions over others that lead to a dynamic equilibrium rather than clean binary choices (*6, 13, 43, 44*). Thus when released from threshold concentrations of the queen inhibitory pheromones, some honeybee workers build royal cells while a smaller number of workers set out to dismantle them. The final result is an equilibrial number of cells sufficient for the rearing of new queens (*44*).

On the other hand, a few mechanisms are coming to light that sharpen the precision of mass response and bring it closer to binary action. Markl and Hölldobler (*45*) reported the existence of "modulatory communication" in ants, a form of signaling in one channel that alters the threshold of response in another. For example, when harvester ants of the genus *Novomessor* encounter large food objects they make sounds by scraping together specialized surfaces on the thin postpetiole and adjacent abdominal segment. This stridulation does not cause an overt behavioral change in nestmates but raises the probability that they will release short-range recruitment chemicals. The overall result is a speeding and tightening of the coordination process.

The second force inhibiting social evolution, at least in the case of hymenopterans, is the substantial conflict among individuals for reproductive privileges. Dominance rank orders, once thought to be confined to simply organized societies of halictine bees, bumblebees, and polis-

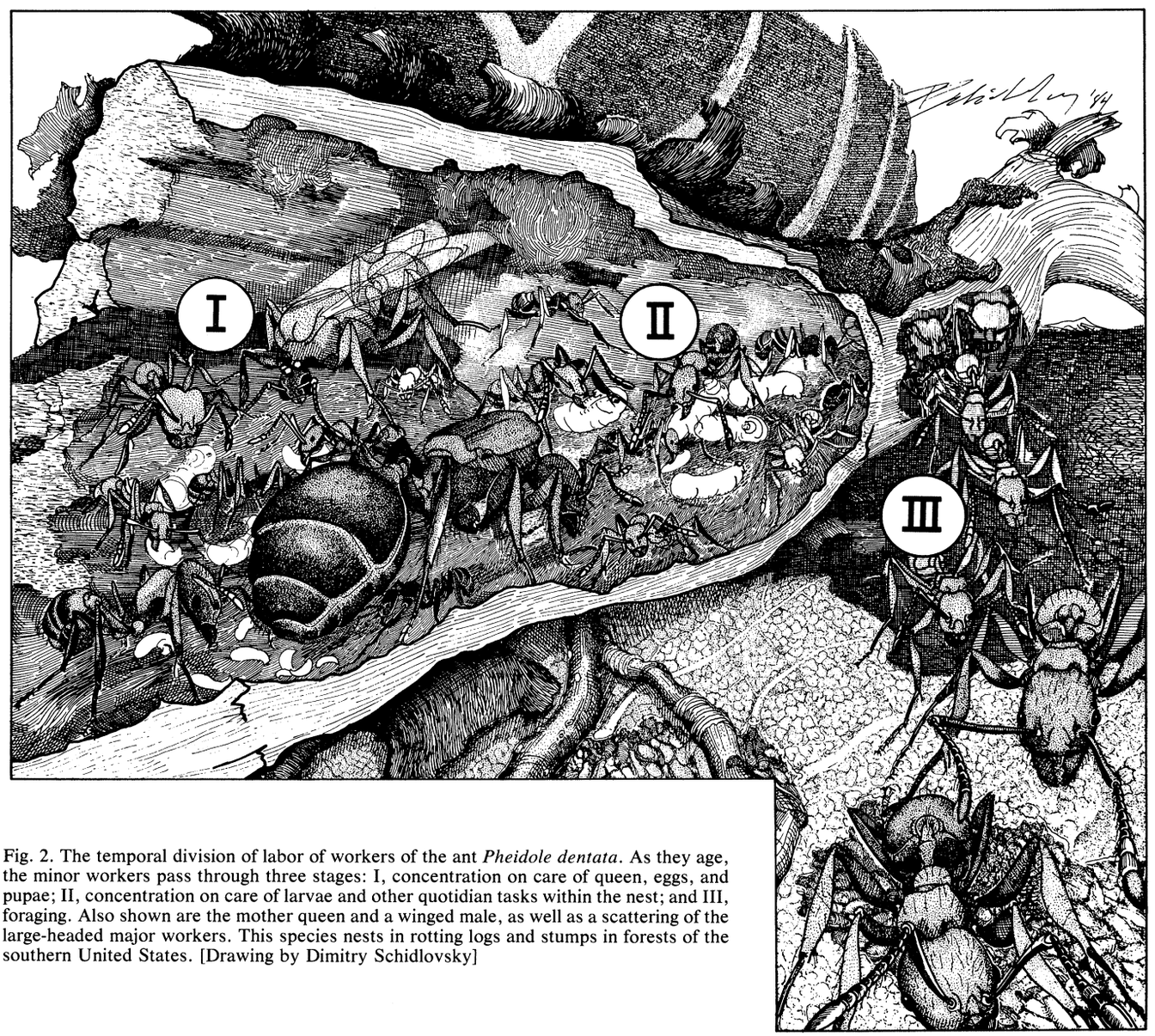


Fig. 2. The temporal division of labor of workers of the ant *Pheidole dentata*. As they age, the minor workers pass through three stages: I, concentration on care of queen, eggs, and pupae; II, concentration on care of larvae and other quotidian tasks within the nest; and III, foraging. Also shown are the mother queen and a winged male, as well as a scattering of the large-headed major workers. This species nests in rotting logs and stumps in forests of the southern United States. [Drawing by Dimitry Schidlovsky]

tine wasps, as well as associations of queens of a few kinds of ants [*Nothomyrmecia, Myrmecocystus*, and *Eurhopalothrix* (*46*)] have also been discovered in the workers of some species of ants as well (*47*). West-Eberhard has argued that competition among workers is more pervasive among advanced societies than has been recognized and that selection at the level of the individual has consequently played a key role in the division of labor (*36, 48*). She explains the centrifugal pattern of temporal castes (Figs. 2 and 3) as the product of such selection. The individual worker, by staying close to the brood chambers while still young and while her personal reproductive value is highest, maximizes her potential to contribute personal offspring. But as death approaches and fertility declines because of senescence, the optimum strategy for contributing genes to the next generation is to enhance colony welfare through more dangerous occupations such as defense and foraging, thus producing more brothers and sisters as opposed to personal offspring. By this criterion, Porter and Jorgensen (*37*) were correct to call foraging harvester ants the "disposable" caste. Hölldobler (*49*) has recently described what may be the ultimate case: aging workers of the Australian tree ant (*Oecophylla smaragdina*) occupy special "barracks nests" around the periphery of the main nest area. They stand idle most of the time and are among the first defenders to enter combat during territorial battles with other tree ant colonies.

Individual selection appears likely to have inhibited the refinement of social behavior, especially in the earliest stages of the evolution. Indeed, there is evidence that species of the bee genus *Exoneurella*, trading production of siblings for the production of offspring, have returned from primitive sociality back to a more nearly solitary state (*50*). Yet there does appear to be a point of no return in the rise of sociality. When colonies become very complex, organized by an intricate caste system and highly coordinated group movements, the advantages of queenlike behavior on the part of workers is diminished and may even disappear. In a few advanced ant genera, such as *Pheidole* and *Solenopsis*, the workers no longer even possess ovaries (*51*).

The pattern emerging from comparative studies suggests that as reproductive competition has declined during the elaboration of sociogenesis, dominance interactions have been ritualized to serve as part of the communicative signals dividing labor. In the more complex societies of bees and wasps, overt aggression is replaced by queen pheromones, but the inhibition of the ovaries of the subordinates and their induction into worker roles remain essentially the same (*6, 14*). Also, traces of aggressive and subordinate interactions persist in ritual form. The workers of stingless bees either hurriedly withdraw from the area when the

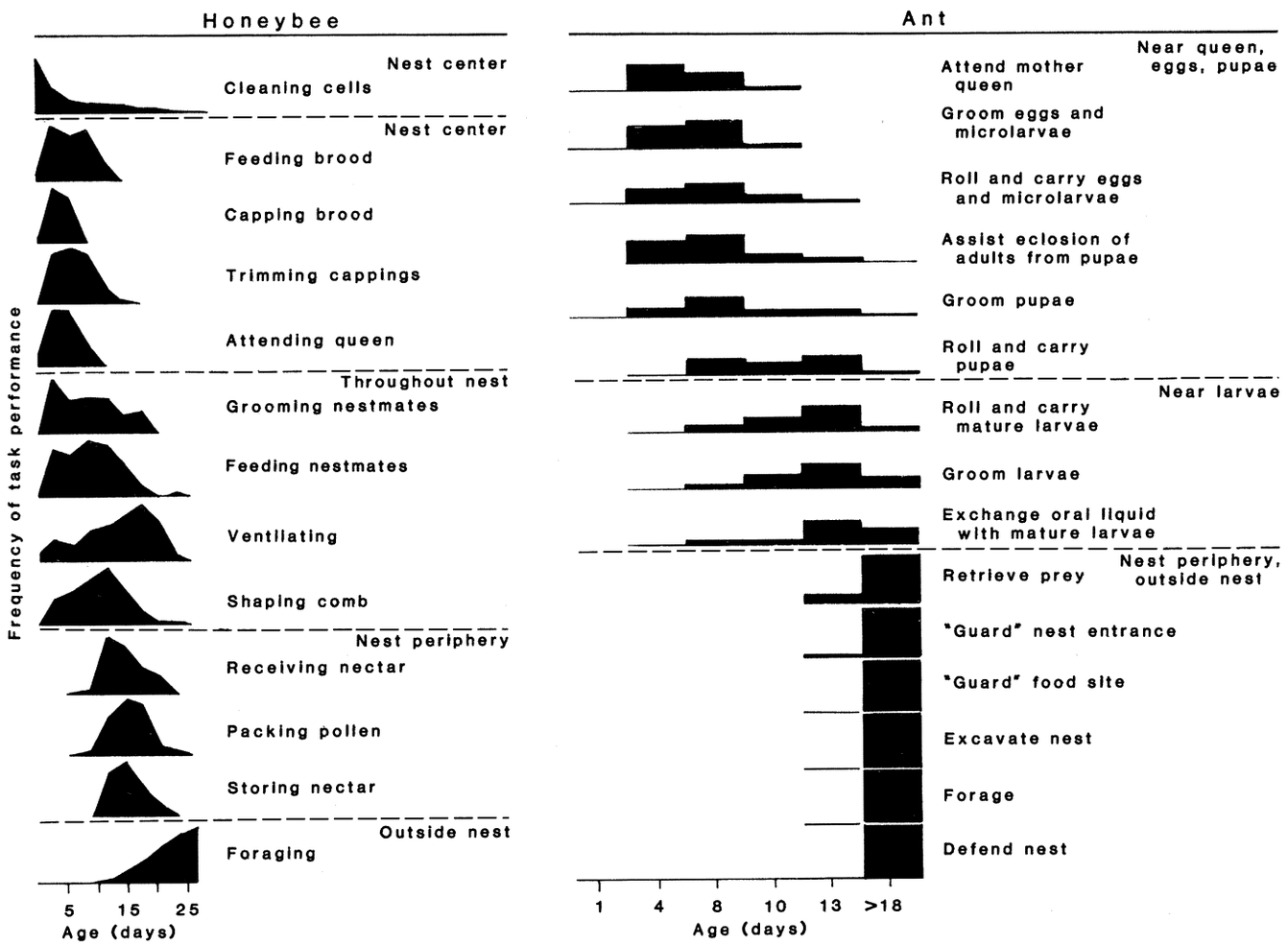


Fig. 3. The temporal division of labor, based on changes of behavior in the adult workers with aging, is shown in the ant *Pheidole dentata* and honeybee *Apis mellifera*: the insects shift from one linked set of tasks to another as they move their activities outward from the nest center (see Fig. 2). The similarities between the two species are convergent and believed to be adaptive. The sum of the frequencies in each histogram is 1.0. Adapted from Wilson (*54*) and Seeley (*55*).

queen approaches, thus clearing a path for her, or else they mock-attack, then bow to her head, and finally swing to her side to become part of the retinue (*52*). Ritualized dominance interactions may also be important between sterile workers. Major workers of the ant *Pheidole pubiventris* turn away from minor workers when they encounter them around the brood, thus yielding most of the care of the immature forms to these smaller nestmates. This aversion neatly divides colony labor into several principal categories (*23*).

Although seldom acknowledged in the literature, regulatory mechanisms are often found lacking even when they are intuitively anticipated by the investigator. For example, the major workers of *Pheidole dentata* are specialized for response against fire ants and other members of the genus *Solenopsis*, but when colonies are stressed continually with these enemies the major-minor ratio remains the same.

In other words, there is no increase in the defense expenditure in the face of a major threat (*24*). Leafcutter workers with head widths from 1.8 through 2.2 mm are responsible for most of the foraging, but when members of this important caste are removed experimentally, the colonies fail to compensate for the loss by increasing representation of the size class in later broods. The result is a reduction in energetic efficiency through two brood cycles (*53*).

On the whole, insect societies display impressive degrees of complexity and integrity on the basis of what appear to be relatively few sociogenetic processes. The mechanisms that do exist, together with their strengths, precision, and phylogenetic distribution, constitute a subject in an early and exciting period of investigation. Of comparable importance are the expected mechanisms that do not exist, so that investigators are likely to pay closer attention to them than has been the case in the past. As the full pattern becomes clearer, it may be possible to compare sociogenesis with morphogenesis in a way that leads to a more satisfying general account of biological organization.

References and Notes

1. L. Beck, *Amazoniana* **3**, 69 (1971).
2. F. J. Fittbau and H. Klinge, *Biotropica* **5**, 2 (1973).
3. J. Levieux, in *The Biology of Social Insects*, M. D. Breed, C. D. Michener, H. E. Evans, Eds. (Westview, Boulder, 1982), pp. 48–51.
4. A. Raignier and J. Van Boven, *Ann. Mus. R. Congo Belg. Tervuren* **2**, 1 (1955).
5. S. Higashi and K. Yamauchi, *Jpn. J. Ecol.* **29**, 257 (1979).
6. E. O. Wilson, *The Insect Societies* (Harvard Univ. Press, Cambridge, Mass., 1971).
7. R. L. Jeanne, *Ecology* **60**, 1211 (1979).
8. N. A. Weber, *Gardening Ants: The Attines* (American Philosophical Society, Philadelphia, 1972); J. M. Cherrett, in *The Biology of Social Insects*, M. D. Breed, C. D. Michener, H. E. Evans, Eds. (Westview, Boulder, 1982), pp. 114–118.
9. D. Davidson, J. H. Brown, R. S. Inouye, *BioScience* **30**, 233 (1980).
10. W. H. Lyford, *Harv. For. Pap.* **7**, 1 (1963).
11. T. Abe, in *The Biology of Social Insects*, M. D. Breed, C. D. Michener, H. E. Evans, Eds. (Westview, Boulder, 1982), pp. 71–75.
12. E. O. Wilson, *Stud. Entomol.* **19**, 187 (1976).
13. G. F. Oster and E. O. Wilson, *Caste and Ecology in the Social Insects* (Princeton Univ. Press, Princeton, N.J., 1978); J. M. Herbers, *J. Theor. Biol.* **89**, 175 (1981).
14. The large literature on the advantages of social life, much of it based on experimental studies, is reviewed by Wilson (*6*) and investigators in (*13*), as well as by C. D. Michener, *The Social Behavior of the Bees: A Comparative Study* (Harvard Univ. Press, Cambridge, Mass., 1974); H. R. Hermann, Ed., *Social Insects* (Academic Press, New York, 1979–1982), vols. 1–4; T. Seeley and B. Heinrich, in *Insect Thermoregulation*, B. Heinrich, Ed. (Wiley, New York, 1981), pp. 159–234.
15. See, for example, N. K. Wessells, *Tissue Interaction and Development* (Benjamin, Menlo Park, Calif., 1977).
16. E. O. Wilson, *Behav. Ecol. Sociobiol.* **7**, 157 (1980).
17. M. V. Brian, in *Social Insects*, H. R. Hermann, Ed. (Academic Press, New York, 1979), vol. 1, pp. 121–222.
18. J. de Wilde and J. Beetsma, *Adv. Insect Physiol.* **16**, 167 (1982).
19. D. E. Wheeler and H. F. Nijhout, *J. Insect Physiol.* **30**, 127 (1984).
20. L. Passera and J.-P. Suzzoni, *Insectes Soc.* **26**, 343 (1979).
21. For example, the very primitive ant *Amblyopone pallipes* appears to lack temporal castes completely [J. F. A. Traniello, *Science* **202**, 770 (1978)].
22. M. I. Haverty, *Sociobiology* **2**, 199 (1977); in *The Biology of Social Insects*, M. D. Breed, C. D. Michener, H. E. Evans, Eds. (Westview, Boulder, 1982), p. 251.
23. E. O. Wilson, *Behav. Ecol. Sociobiol.* **16**, 89 (1984).
24. A. B. Johnston and E. O. Wilson, *Ann. Entomol. Soc. Am.* **78**, 8 (1985).
25. E. O. Wilson and R. W. Taylor, *Psyche* **71**, 93 (1964).
26. E. O. Wilson, *Am. Nat.* **102**, 41 (1968); J. M. Herbers, *Evolution* **34**, 575 (1980).
27. E. O. Wilson, *Behav. Ecol. Sociobiol.* **14**, 55 (1983).
28. C. G. Butler and R. K. Callow, *Proc. R. Entomol. Soc. London* **B43**, 62 (1968).
29. Y. Lenski and Y. Slabezki, *J. Insect Physiol.* **27**, 313 (1981).
30. E. O. Wilson, *Behav. Ecol. Sociobiol.* **1**, 63 (1976).
31. C. Butler, *The Feminine Monarchie* (Barnes, Oxford, 1609).
32. T. R. Schneirla, *Army Ants: A Study in Social Organization*, H. R. Topoff, Ed. (Freeman, San Francisco, 1971).
33. H. K. Reeve and G. J. Gamboa, *Behav. Ecol. Sociobiol.* **13**, 63 (1983).
34. N. F. Carlin and B. Hölldobler, *Science* **222**, 1027 (1983).
35. _____, personal communication.
36. M. J. West-Eberhard, in *Natural Selection and Social Behavior*, R. D. Alexander and D. W. Tinkle, Eds. (Chiron, New York, 1981), pp. 3–17.
37. S. D. Porter and C. D. Jorgensen, *Behav. Ecol. Sociobiol.* **9**, 247 (1981).
38. T. D. Seeley, *ibid.* **11**, 287 (1982).
39. M. Rockstein, *Ann. Entomol. Soc. Am.* **43**, 152 (1950); S. F. Sakagami and H. Fukuda, *Res. Popul. Ecol.* **10**, 127 (1968).
40. S. Badertscher, C. Gerber, and R. H. Leuthold, *Behav. Ecol. Sociobiol.* **12**, 115 (1983).
41. E. O. Wilson, *ibid.* **7**, 143 (1980).
42. F. M. Carpenter and H. R. Hermann, in *Social Insects*, H. R. Hermann, Ed. (Academic Press, New York, 1979), vol. 1, pp. 81–89.
43. P. Hogeweg and B. Hesper, *Behav. Ecol. Sociobiol.* **12**, 271 (1983).
44. D. H. Baird and T. D. Seeley, *ibid.* **13**, 221 (1983).
45. H. Markl and B. Hölldobler, *ibid.* **4**, 183 (1978).
46. S. H. Bartz and B. Hölldobler, *ibid.* **10**, 137 (1982); B. Hölldobler and R. W. Taylor, *Insectes Soc.* **30**, 384 (1983); E. O. Wilson, *ibid.* **30**, 408 (1985).
47. B. J. Cole, *Science* **212**, 83 (1981); N. Franks and E. Scovell, *Nature (London)* **304**, 724 (1983).
48. M. J. West-Eberhard, *Proc. Am. Philos. Soc.* **123**, 222 (1979).
49. B. Hölldobler, *Biotropica* **15**, 241 (1983).
50. C. D. Michener, *Kansas Univ. Sci. Bull.* **46**, 317 (1965).
51. E. O. Wilson, *J. Kansas Entomol. Soc.* **51**, 615 (1978).
52. S. F. Sakagami, in *Social Insects*, H. R. Hermann, Ed. (Academic Press, New York, 1982), vol. 3, pp. 361–423.
53. E. O. Wilson, *Behav. Ecol. Sociobiol.* **15**, 47 (1983).
54. _____, *ibid.* **1**, 141 (1976).
55. T. D. Seeley, *ibid.* **11**, 287 (1982).

56. I am grateful to D. M. Gordon, B. Hölldobler, T. Seeley, and D. Wheeler for critical readings of the manuscript. Supported by a series of grants from the National Science Foundation, the latest of which is BSR 81-19350.

tion processing task such as filtering out spurious input fluctuation (*25*), generating temporal programs of expression (*3, 25*) or accelerating the throughput of the network (*2, 26*). Recently, the same network motifs were also found in the transcription network of yeast (*7*, *27*). It is important to stress that the similarity in circuit structure does not necessarily stem from circuit duplication. Evolution, by constant tinkering, appears to converge again and again on these circuit patterns in different nonhomologous systems (*25*, *27*, *28*), presumably because they carry out key functions (see Perspective (*29*) STKE). Network motifs can be detected by algorithms that compare the patterns found in the biological network to those found in suitably randomized networks (*25*, *27*). This is analogous to detection of sequence motifs as recurring sequences that are very rare in random sequences.

Network motifs are likely to be also found on the level of protein signaling networks (*30*). Once a dictionary of network motifs and their functions is established, one could envision researchers detecting network motifs in new networks just as protein domains are currently detected in the sequences of new genes. Finding a sequence motif (e.g., a kinase domain) in a new protein sheds light on its biochemical function; similarly, finding a network motif in a new network may help explain what systems-level function the network performs, and how it performs it.

Will a complete description of the biological networks of an entire cell ever be available? The task of mapping an unknown network is known as reverse-engineering (*3*, *31–33*). Much of engineering is actually reverse-engineering, because prototypes often do not work and need to be understood in order to correct their design. The program of molecular biology is reverse-engineering on a grand scale. Reverse engineering a nonmodular network of a few thousand components and their nonlinear interactions is impossible (exponentially hard with the number of nodes). However, the special features of biological networks discussed here give hope that biological networks are structures that human beings can understand. Modularity, for example, is at the root of the success of gene functional assignment by expression correlations (*11*, *34*). Robustness to component tolerances limits the range of possible circuits that function on paper to only a few designs that can work in the cell. This can help theorists to home in on the correct design with limited data (*21–23*). Network motifs define the few basic patterns that recur in a network and, in principle, can provide specific experimental guidelines to determine whether they exist in a given system (*25*). These concepts, together with the current technological revolution in biology, may eventually allow characterization and understanding of cell-wide networks, with great benefit to medicine. The similarity between the creations of tinkerer and engineer also raises a fundamental scientific challenge: understanding the laws of nature that unite evolved and designed systems.

References and Notes

1. F. Jacob, *Science* **196**, 1161 (1977).
2. M. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, MA, 1976).
3. M. Ronen, R. Rosenberg, B. I. Shraiman, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 10555 (2002).
4. C. H. Yuh, H. Bolouri, E. H. Davidson, *Science* **279**, 1896 (1998).
5. Y. Setty, A. E. Mayo, M. G. Surette, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7702 (2003).
6. D. Thieffry, A. M. Huerta, E. Perez-Rueda, J. Collado-Vides, *Bioessays* **20**, 433 (1998).
7. T. I. Lee *et al.*, *Science* **298**, 799 (2002).
8. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
9. S. H. Strogatz, *Nature* **410**, 268 (2001).
10. L. H. Hartwell, J. J. Hopfield, S. Leibler, A. W. Murray, *Nature* **402**, C47 (1999).
11. J. Ihmels *et al.*, *Nature Genet.* **31**, 370 (Aug, 2002).
12. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási, *Science* **297**, 1551 (2002).
13. C. R. Myers, arXiv: cond-mat/0305575 (2003).
14. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).
15. D. Bray, *J. Theor. Biol.* **143**, 215 (1990).
16. H. Lipson, J. B. Pollack, N. P. Suh, *Evolution* **56**, 1549 (2002).
17. J. Gerhart, M. W. Kirschner, *Cells, Embryos, and Evolution: Toward a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability* (Blackwell Science Inc, Oxford, 1997).
18. C. H. Waddington, *Nature* **150**, 563 (1942).
19. H. Kacser, J. A. Burns, *Symp. Soc. Exp. Biol.* **32**, 65 (1973).
20. M. Savageau, *Nature* **229**, 542 (1971).
21. N. Barkai, S. Leibler, *Nature* **387**, 913 (1997).
22. U. Alon, M. G. Surette, N. Barkai, S. Leibler, *Nature* **397**, 168 (1999).
23. A. Eldar *et al.*, *Nature* **419**, 304 (2002).
24. D. Fell, *Understanding the Control of Metabolism* (Portland Press, London, 1997).
25. S. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
26. N. Rosenfeld, M. B. Elowitz, U. Alon, *J. Mol. Biol.* **323**, 785 (2002).
27. R. Milo *et al.*, *Science* **298**, 824 (2002).
28. G. C. Conant, A. Wagner, *Nature Genet.* **34**, 264 (2003).
29. A. Wagner, *Sci. STKE* **2003**, pe41 (2003).
30. C. V. Rao, A. P. Arkin, *Annu. Rev. Biomed. Eng.* **3**, 391 (2001).
31. M. E. Csete, J. C. Doyle, *Science* **295**, 1664 (2002).
32. A. Arkin, P. Shen, J. Ross, *Science* **277**, 1275 (1997).
33. T. S. Gardner, D. di Bernardo, D. Lorenz, J. J. Collins, *Science* **301**, 102 (2003).
34. M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14863 (1998).

VIEWPOINT

Social Insect Networks

Jennifer H. Fewell

Social insect colonies have many of the properties of adaptive networks. The simple rules governing how local interactions among individuals translate into group behaviors are found across social groups, giving social insects the potential to have a profound impact on our understanding of the interplay between network dynamics and social evolution.

The formal exploration of social insect colonies as networks is in its infancy. However, social insects such as wasps, ants, and honeybees provide a powerful system for examining how network dynamics contribute to the evolution of complex biological systems. Social insect colonies (and social groups generally) have key network attributes that appear consistently in complex biological systems, from molecules through ecosystems; these include nonrandom systems of connectivity and the self-organization of group-level phenotypes (*1–3*). Colonies exhibit multiple levels of organization, yet it is still possible to track individuals, making these societies more accessible to experimental manipulation than many other complex systems.

How can viewing insect societies as networks shape our understanding of social organization and evolution? First, they have become one of the central model systems for exploring self-organization: the process by which interactions occurring locally between individuals produce group-level attributes. Self-organization in a social insect colony produces emergent properties: social phenotypes that are greater than a simple summation of individual worker behaviors (*2*). The basic rules generating these dynamics are broadly applicable across taxa whose members show social behavior, and they produce ubiquitous patterns of social organization, including mass action responses, division of labor, and social hierarchies (*2*, *4*).

School of Life Sciences, Arizona State University, Tempe, AZ 85287–1501, USA. E-mail: j.fewell@asu.edu

Second, the social insects provide an opportunity to explore how behavior evolves within complex systems. This has led to a shift in focus from variation among individuals to how interactions among individuals and groups shape that variation. Most of the well-studied social insects are eusocial (only one or a few individuals in the colony reproduce), and the colony is considered an adaptive unit made up of related individuals (*5*). Because of this, we are comfortable in relating group dynamics to fitness effects at both the individual and group levels. However, multilevel selection acts on social insect colonies, not just because their members are highly related but also because they are densely connected networks. This emerging view of social groups as networks contributes to a growing awareness of how the fitness of individuals and groups is generated interactively across levels of biological organization (*3*, *6*, *7*).

To explore the relationships between complexity and selection in social systems, we first need to describe the social group as a network. A network is simplistically a system of interacting elements, or nodes, that communicate with each other [see (*8*, *9*) in this issue]. Social insect colonies are dense networks in which individuals have multiple points of contact (*1*, *10*). As dense networks, colonies distribute information rapidly, allowing them to respond flexibly and efficiently to the dynamic environment in which they live. An extreme example is the alarm response of African honeybees, in which an initial release of alarm pheromone by a few guards cascades within a minute to stinging responses by thousands of bees.

Like many biological systems, social insect colonies are also distributed networks (*2*). Although the colony generally has a single queen, she does not centrally control colony function. Instead, workers make decisions based on local information and perform behaviors in parallel (*10*). This is the case, to some degree, even for hierarchical systems such as the wasp network, where the queen controls the reproductive output of the colony but does not individually direct many aspects of day-to-day colony function. We lack sufficient data to accurately characterize the connections that occur between any two individuals within a colony, much less the connections across the society. However, it is clear that connections among nestmates are nonrandomly distributed for many, if not most, colony functions. A few key individuals, or hubs, distribute information (connect) to many more nestmates than do others. The most obvious of these is the queen, who, in honeybees, secrets a pheromone that represses reproduction in workers and maintains colony cohesion. Queen pheromone is transmitted to workers as they groom her, then is rapidly transmitted through the hive via trophallaxis and deposits on nest wax (*11*). Key individuals are also present within worker task groups, where they stimulate performance of a task or provide a central point around which performance is organized (*12*). For example, foraging task groups often include scouts or dancers. They communicate most of the information about resource location and availability and, in ants, often maintain the cohesion of groups of recruits that go out to forage (*10*, *12*, *13*).

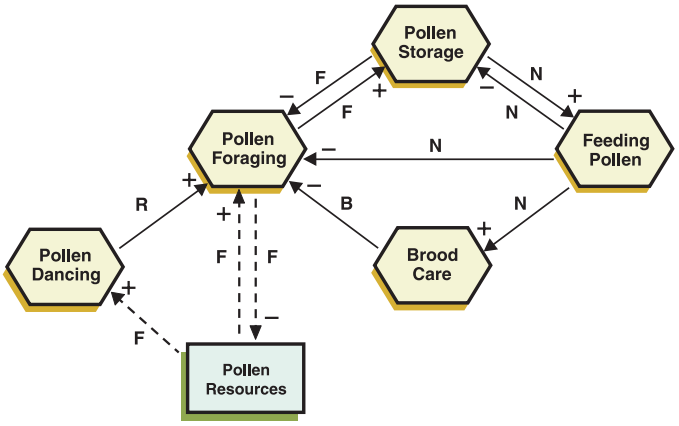


Fig. 1. The network pathways modulating pollen foraging in a honeybee colony (developed with T. Taylor). Nodes are the tasks linked to pollen foraging; vectors are the individuals transmitting information: F, forager; N, nurse; B, brood; R, recruit. Foragers returning with pollen receive information about pollen storage levels as they place pollen loads into cells. The amount of stored pollen is negative feedback for pollen foraging. Pollen is removed from cells by nurse bees, who feed it first to developing brood and give excess to pollen foragers. Receiving pollen from nurses is negative feedback for foraging. Foragers also receive information about pollen stores from brood, who produce a hunger pheromone when they are not fed; brood care reduces hunger levels. Information on pollen availability and location is transmitted by pollen dancers. Dancing elicits recruitment to foraging by workers not actively engaged in foraging (*2*, *20*–*22*).

The importance of these rare individuals makes it likely that for many functions the colony network becomes scale-free, which means that variation in connectivity is best described by a power law rather than a Poisson distribution (*14*). This is important to colony resiliency, because it means that the loss of any of the vast majority of workers would have little effect. In contrast, removing nodes within a randomly distributed network can quickly fragment the system. Although scale-free networks are buffered from the effects of random loss, the removal of key nodes can severely disrupt the system (*1*, *14*). The colony has long-term mechanisms to replace any element, including the queen, but the removal of key individuals does have immediate disruptive effects. Loss of the scout who discovers a foraging trail can completely block the retrieval of a resource, yet removing the recruits who follow the scout has little effect on overall foraging (*15*). Social insect networks are similar in this way to other biological networks, from food webs with keystone species (*16*) to metabolic pathways, in which a few key molecules are involved in most reactions (*9*, *17*).

With these global attributes in place, how does information transfer within a social colony actually occur? Unfortunately, we do not yet have enough empirical data to answer this question well. Models to date have explored networks in the context of task regulation: the amount of effort by individuals or groups that is allocated to different tasks. One approach has been to consider the colony as a regular network (*9*), in which individuals performing the same task form clusters of high connectivity, with weaker links across tasks (*18*, *19*). In a model of recruitment to alternate resource pathways, Bonabeau *et al*. (*19*) showed that colonies can balance efficient utilization of a single resource with flexible allocation across resources by a mixed strategy of within-cluster information transfer coupled with global information transfer across clusters. An important finding of this model [and the Pacala *et al*. model (*18*) on which it was based] is the importance of cross-cluster links in maintaining flexibility for moving individuals from one task or cluster to another.

The assumptions of the Bonabeau *et al*. model (*19*) fit well into the context of trail selection during foraging, where the signals are well defined. However, expanding the model more widely to multiple tasks has been problematic. One reason is that contacts between workers are extremely fluid. Connections between workers in a social insect colony are ephemeral, and signals themselves can outlive connections. Signal systems are also highly diverse in information content and include large-scale signals, such as alarm pheromones, that target the colony globally (*10*).

Social insect networks are traditionally modeled with workers as nodes. However, because worker interactions are so fluid, we can alternatively map the system from the perspective of treating tasks as nodes and individual workers as connectors (symbolic dynamics). Figure 1 describes such a map for the short-term modulation of pollen foraging in honeybees. It is clear from this map that

cross-task connections are the primary pathways for regulating pollen foraging. Pollen foraging is homeostatically regulated around pollen storage levels and is positively regulated around brood, for which it is the main nutrient source (*20*). Pollen foragers collect most of their information about colony pollen need and/or intake either indirectly via changes in pollen stores, from nurse bees who feed foragers when excess pollen is available, or from brood who emit hunger signals (*21*, *22*). The map is not consistent with the assumption of high within-cluster connectivity, but it does support the assertion that connections across tasks are important to allocation (*18*, *19*). They may, in fact, be the primary links modulating task regulation globally. If so, information flow in social insect colonies has an important similarity to that in human social networks, where weak ties across social clusters play an important role in regulating society as a large-scale network (*23*).

Although the complexities of the whole-colony network have not yet been well described, large strides have been made in the analysis of how local interactions within the network affect global colony dynamics. As dense networks, social insect colonies have a high potential for the emergence of large-scale phenomena via self-organization (*1*). Self-organization pervades all aspects of colony function, including foraging, nest defense, resource storage, nest construction, site selection, thermoregulation, and division of labor (*2*).

The growing body of theoretical and empirical work on self-organization is one of the more important contributions of social insect research to understanding biocomplexity (*2*). What is perhaps most important about self-organization in social insects is that it is not based on derived characteristics unique to the taxon. Instead, it is driven by a limited set of nonlinear dynamics that should occur across social systems, from insects to humans (*2*, *4*). As an example, a majority of the emergent components of social behavior can be categorized as "convergent," in which individuals become behaviorally more similar, or "divergent," in which the behavior of one individual reduces the likelihood that the second individual will perform the same behavior.

The minimal components (or minimal rule set) for convergence can be condensed to (i) a positive stimulus for the behavior as a result of its performance; (ii) amplification of the stimulus through successive iterations; and (iii) a decay component, so that signals and cues must be regenerated. A beautiful example of behavioral convergence via these minimal rules is found in the trail marking system of the Argentine ant *Linepithema humile*. Workers traveling to and from resources lay a pheromonal trail. Each time a trail is laid, the local environment at points of choice between alternate trails is changed. Ants reaching these points preferentially choose the trail with more pheromone and add to it, creating a positive feedback loop. Meanwhile, the pheromone marks on the alternate trail decay. As more foragers repeat this process, one trail becomes the primary and often the only route (*2*, *24*). These simple rules underlie trail-making in multiple ant species (*2*). Similar rules describe convergent group behaviors in other social species, such as migrating social spiders who choose a direction of travel based on the accumulation of draglines from others in the group (*25*).

The minimal rule set for divergence can be condensed to two components: (i) performance of a behavior by one individual reduces the probability that others will perform the same behavior, and (ii) stimulus levels for the behavior increase in the absence of performance. Most divergence models also include a positive feedback loop, in which performance of the behavior increases the probability that the individual will perform the behavior again. This self-reinforcement generates divergence even with initially small random differences in behavior and produces a faster and more stable system of divergence (*26*). However, divergence can emerge in the absence of self-reinforcement if individuals initially differ intrinsically in their response thresholds: the stimulus level at which they respond by performing a behavior (*27*, *28*).

This rule set forms the basis for the response threshold models of division of labor (*27*). These models begin with the initial assumption that individuals perform a task when environmental stimuli reach a level that matches the individual's threshold for response. That individual performs the task; in doing so, she reduces the stimulus levels encountered by others and thus reduces their probability of performing the task also. Empirical tests on solitary bees and on ant queens during colony founding have shown that division of labor can emerge even without a history of direct selection (*29*). When normally solitary ant queens are forced into artificial social groups, one individual takes over the task of excavation, whereas the other individual remains in the nest and tends brood. The dynamics of this division of labor fit well with the predictions of the response threshold model.

Similar patterns of divergence occur across other social taxa. Social hierarchies within bumblebees and primates can be modeled by a similar minimal rule set for divergence, coupled with reinforcement (*30*, *31*). Division of labor also appears frequently within social species, including humans. As an example, we can imagine an apartment where housemates share tasks. Used dishes pile up in the sink, producing a continuously increasing stimulus. The dishes go unnoticed until the threshold of the one most sensitive to them is met, and he or she washes them. This removes the dishes as a stimulus, further reducing the likelihood that the other group members will ever wash them. The result is a dishwashing specialist (much to his/her dismay), and a set of nondishwashers. Similar interactions across other chores, from cleaning the bathroom to taking out the garbage, generate a division of labor for the household.

The realization that individuals within a social group are linked as a network is important to our understanding of how selection acts on sociality. The fitness of every individual in the group is produced in part as a result of their interactions with other group members. The emergence of collective behaviors via self-organization also produces phenotypes at the colony level that are themselves subject to selection (*7*). These interactions set the stage for multilevel selection (*32*). Network-level properties, including group size, connectivity, and even variation in individual responsiveness to signals can all shape the adaptive function of the social group (*18*, *28*). As an example, as described above, the emergence of division of labor is based in part on intrinsic variation in worker response thresholds. Honeybee colonies with more diversity in worker thresholds for foraging are able to respond better to changes in the availability and need for resources. This diversity is generated by the extreme polyandry of honeybee queens, who mate with a dozen or more males (*22*).

Network interactions also have a profound influence on individual behavior and fitness. The fitness of each individual in a social group is dependent on the phenotypes of the other group members (*7*); they are each other's social environments. These reciprocal fitness effects are generated by nonlinear interactions within the social network. In some systems, self-organization can actually generate conflicting fitness effects at the individual and group levels. For ant queens, when division of labor spontaneously emerges from small initial differences in behavior (*29*), it produces associated fitness disparities, because the queen who takes over the task of nest excavation is more likely to die. Whether an individual becomes the excavator, and suffers the associated fitness consequences, depends on which group they land in and the thresholds of everyone in that group.

What should be done next in the exploration of social groups as networks? We need to expand our models from elegant descriptions of single behaviors to incorporate the more complex dynamics of the group as a whole. We also need to test those models empirically

on a wider range of social systems. Finally, to understand the evolutionary significance of network dynamics, we must explicitly measure their fitness effects on the social group (*7*). This interplay between network dynamics and selection is just beginning to be explored, and social insects have the potential to be on the leading edge.

References

1. A.-L. Barabási, *Linked: The New Science of Networks* (Perseus, Cambridge, MA, 2002).
2. S. Camazine *et al.*, *Self-Organization in Biological Systems* (Princeton Univ. Press, Princeton, NJ, 2001).
3. R. Solé, B. C. Goodwin, *Signs of Life: How Complexity Pervades Biology* (Basic Books, New York, 2000).
4. C. K. Hemelrijk, *Ethology* **108**, 655 (2002).
5. T. D. Seeley, *Am. Nat.* **150**, S22 (1997).
6. D. S. Wilson, L. A. Dugatkin, *Am. Nat.* **149**, 336 (1997).
7. A. J. Moore, E. D. Brodie, J. B. Wolf, *Evolution* **51**, 1352 (1997).
8. U. Alon, *Science* **301**, 1866 (2003).
9. D. Bray, *Science* **301**, 1864 (2003).
10. B. Hölldobler, E. O. Wilson, *The Ants* (Belknap Press of Harvard Univ. Press, Cambridge, MA, 1990).
11. K. Naumann, M. Winston, K. Slessor, G. Prestwich, F. Webster, *Behav. Ecol. Sociobiol.* 29, 321 (1991).
12. S. K. Robson, J. F. A. Traniello, in *Information Processing in Social Insects*, C. Detrain, J. L. Deneubourg, J. M. Pasteels, Eds. (Birkhauser, Basel, Switzerland, 1999), pp. 239–259.
13. T. D. Seeley, S. Camazine, J. Sneyd, *Behav. Ecol. Sociobiol.* **28**, 277 (1991).
14. R. Albert, A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
15. D. M. Gordon, *Am. Nat.* **159**, 509 (2002).
16. R. Sole, J. Montoya, *Santa Fe Working Paper* 00-11-060 (2000).
17. H. Jeong, B. Tombor, R. Albert, Z. Oltvai, A.-L. Barabási, *Nature* **407**, 651 (2000).
18. S. W. Pacala, D. M. Gordon, H. C. J. Godfray, *Evol. Ecol.* **10**, 127 (1996).
19. E. Bonabeau, G. Theraulaz, J.-L. Deneubourg, *J. Theor. Biol.* **195**, 157 (1998).
20. J. H. Fewell, M. Winston, *Behav. Ecol. Sociobiol.* **30**, 387 (1992).
21. S. Camazine, *Behav. Ecol. Sociobiol.* **32**, 265 (1993).
22. R. E. Page, J. Erber, *Naturwissenschaften* **89**, 91 (2002).
23. M. Granovetter, *Am. J. Sociol.* **78**, 1360 (1973).
24. S. Goss, S. Aron, J. L. Deneubourg, J. M. Pasteels, *Naturwissenschaften* **76**, 579 (1989).
25. F. Saffre, R. Furey, B. Krafft, J. L. Deneubourg, *J. Theor. Biol.* **198**, 507 (1999).
26. G. Theraulaz, E. Bonabeau, J. L. Deneubourg, *Proc. R. Soc. London Ser. B* **265**, 327 (1998).
27. S. N. Beshers, J. H. Fewell, *Annu. Rev. Entomol.* **46**, 413 (2001).
28. R. E. Page, S. D. Mitchell, *Apidologie* **29**, 171 (1998).
29. J. H. Fewell, R. E. Page, *Evol. Ecol. Res.* **1**, 537 (1999).
30. P. Hogeweg, B. Hesper, *Behav. Ecol. Sociobiol.* **12**, 271 (1983).
31. C. K. Hemelrijk, *Biol. Bull.* **202**, 283 (2002).
32. D. S. Wilson, *Am. Nat.* **150**, S1 (1997).

REVIEW

Communication in Neuronal Networks

Simon B. Laughlin[1] and Terrence J. Sejnowski[2,3]*

Brains perform with remarkable efficiency, are capable of prodigious computation, and are marvels of communication. We are beginning to understand some of the geometric, biophysical, and energy constraints that have governed the evolution of cortical networks. To operate efficiently within these constraints, nature has optimized the structure and function of cortical networks with design principles similar to those used in electronic networks. The brain also exploits the adaptability of biological systems to reconfigure in response to changing needs.

Neuronal networks have been extensively studied as computational systems, but they also serve as communications networks in transferring large amounts of information between brain areas. Recent work suggests that their structure and function are governed by basic principles of resource allocation and constraint minimization, and that some of these principles are shared with human-made electronic devices and communications networks. The discovery that neuronal networks follow simple design rules resembling those found in other networks is striking because nervous systems have many unique properties.

To generate complicated patterns of behavior, nervous systems have evolved prodigious abilities to process information. Evolution has made use of the rich molecular repertoire, versatility, and adaptability of cells. Neurons can receive and deliver signals at up to 10^5 synapses and can combine and process synaptic inputs, both linearly and nonlinearly, to implement a rich repertoire of operations that process information (*1*). Neurons can also establish and change their connections and vary their signaling properties according to a variety of rules. Because many of these changes are driven by spatial and temporal patterns of neural signals, neuronal networks can adapt to circumstances, self-assemble, autocalibrate, and store information by changing their properties according to experience.

The simple design rules improve efficiency by reducing (and in some cases minimizing) the resources required to implement a given task. It should come as no surprise that brains have evolved to operate efficiently. Economy and efficiency are guiding principles in physiology that explain, for example, the way in which the lungs, the circulation, and the mitochondria are matched and coregulated to supply energy to muscles (*2*). To identify and explain efficient design, it is necessary to derive and apply the structural and physicochemical relationships that connect resource use to performance. We consider first a number of studies of the geometrical constraints on packing and wiring that show that the brain is organized to reduce wiring costs. We then examine a constraint that impinges on all aspects of neural function but has only recently become apparent—energy consumption. Next we look at energy-efficient neural codes that reduce signal traffic by exploiting the relationships that govern the representational capacity of neurons. We end with a brief discussion on how synaptic plasticity may reconfigure the cortical network on a wide range of time scales.

Geometrical and Biophysical Constraints on Wiring

Reducing the size of an organ, such as the brain, while maintaining adequate function is usually beneficial. A smaller brain requires fewer materials and less energy for construction and maintenance, lighter skeletal elements and muscles for support, and less energy for carriage. The size of a nervous system can be reduced by reducing the number of neurons required for adequate function, by reducing the average size of neurons, or by laying out neurons so as to reduce the lengths of their connections. The design principles governing economical layout have received the most attention.

Just like the wires connecting components in electronic chips, the connections between neurons occupy a substantial fraction of the total volume, and the wires (axons and dendrites) are expensive to operate because they dissipate energy during signaling. Nature has an important advantage over electronic circuits because components are connected by wires in three-dimensional (3D) space, whereas even the most advanced VLSI (very large scale integra-

[1]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [2]Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [3]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

*To whom correspondence should be addressed. E-mail: terry@salk.edu

Jeffry L. Ramsey

Statement

and

Readings

THAT, AND WHY, CHEMISTRY IS REDUCIBLE TO PHYSICS, BUT NOT FULLY SO

Jeff Ramsey
Smith College

If reductionism is "the thesis that the results of inquiry in one domain . . . can be understood or are explained by the conceptual resources of another, more fundamental domain" (Wimsatt and Sarkar 2006, 696), then chemistry is reducible to physics. However, it is not fully reducible if the notion of a full reduction is taken to imply a loss of explanatory power or ontological robustness for the higher level of phenomena.

The existing philosophical and scientific literature supports the claim that chemistry is reducible, but not fully so, to physics. Examples that illustrate this include the periodic table, the use of orbitals to explain bonding and concept of molecular shape. In each case, the conceptual resources of physics have helped us understand or explain the chemical phenomena. But in each case the relation has not led (and is not likely to lead) to elimination or loss of explanatory power for the chemical concept or phenomenon.

Why should the situation be thus? I speculate that, from the perspective of physics, chemistry is a compositional science. The above examples of successful reductions rely on referential identities or localizations rather than relations among theories. Thus, the relevant conception of reduction here is the explanation of one level through the operations of often qualitatively different mechanisms at a lower level. This makes defensible a claim of ontological and explanatory autonomy of the higher level.

CAMBRIDGE STUDIES IN PHILOSOPHY AND BIOLOGY

This major series publishes the very best work in the philosophy of biology. Nonsectarian in character, the series extends across the broadest range of topics: evolutionary theory, population genetics, molecular biology, ecology, human biology, systematics, and more. A special welcome is given to contributions treating significant advances in biological theory and practice, such as those emerging from the Human Genome Project. At the same time, due emphasis is given to the historical context of the subject, and there is an important place for projects that support philosophical claims with historical case studies.

Books in the series are genuinely interdisciplinary, aimed at a broad cross-section of philosophers and biologists, as well as interested scholars in related disciplines. They include specialist monographs, collaborative volumes, and – in a few instances – selected papers by a single author.

Alfred I. Tauber: *The Immune Self: Theory or Metaphor?*
Elliott Sober: *From a Biological Point of View*
Robert Brandon: *Concepts and Methods in Evolutionary Biology*
Peter Godfrey-Smith: *Complexity and the Function of Mind in Nature*
William A. Rottschaefer: *The Biology and Psychology of Moral Agency*

Genetics and reductionism

SAHOTRA SARKAR

CAMBRIDGE UNIVERSITY PRESS

3

Types of reduction: Substantive issues

This chapter will try to show that the formal issues that were discussed in the last chapter pale into insignificance – or, at least, into scientific and philosophical disinterest – in comparison with the substantive issues about reduction that arise once scientific explanations are considered in their full complexity. These issues are clustered around two questions:

(i) how is the system that is being studied [and the behavior of which is potentially being explained (or reduced)] *represented*?; and
(ii) what, exactly, has to be assumed about objects and their interactions for the *explanation* to work?

These questions have been posed generally enough to be applicable to all (natural) scientific contexts, including the physical sciences. Similarly, the analysis of reduction that is developed here is intended to be applicable to these other contexts. Nevertheless, given the limited scope of this book, the examples analyzed here will all be from genetics and molecular biology. Similarly, the general philosophical implications of this analysis that are drawn at the end of this chapter (§ 3.7–§ 3.12) will also be geared towards molecular biology and genetics though they are intended to be more generally applicable.

The basic strategy of this analysis will be to develop and use three substantive criteria to distinguish five different types of reduction. Three of these types of reduction are more important than others, and the rest of the book will proceed to use them to analyze the various types of explanation that are encountered in genetics. Two broad intuitions about reduction have guided the choice of the three criteria.

(i) Reduction involves the explanation of laws or phenomena in one realm by those in another. In this sense, reductions raise the potential for unification of knowledge, though this issue will turn out to be subtler –

and more controversial – than it may initially appear (see § 3.12). This intuition, that is, the requirement of two different realms, arises out of a desire to ensure that there remains some distinction between reductions and other kinds of explanation.

(ii) Many of the reductions will be attempts to explain properties of complex wholes in terms of their constituent parts. This was the kind of reduction that was expected of the mechanical philosophy of the seventeenth century, whether in the realm of physics or of the biological sciences. It was the motivation behind the formulation of the kinetic theory of matter, and the well-known attempts to reduce thermodynamics to the kinetic theory during the latter half of the last century. This is the type of reduction that is supposed to occur in molecular biology, when biological phenomena originally studied "classically" are explained on the basis of molecular mechanisms.

In the spirit of what was said in the last chapter, no assumption about the form of explanation will be made in the discussions of this chapter (or at any later point in this book). This is part of the general attempt to shift attention away from formal issues to substantive ones. However, a few general assumptions about what any explanation must presume will be necessary for the discussions and will be explicitly stated in § 3.1. These will be kept as minimal and uncontroversial as possible and will be self-consciously formulated as substantive assumptions. Attention will then be focused on the additional criteria that must be satisfied by an explanation for it to be a reduction and, if it is a reduction, to be the type of reduction that it is. These additional criteria are the substantive criteria mentioned above. These criteria will also be formulated in a fashion general enough to be consistent at least with any of the models of explanation that are currently in vogue. In particular, they will be consistent with both deductive–nomological explanation and with the types of explanation that are based on Salmon's (1971) "statistical relevance" model. Consistency with the former is important, because its miscellaneous variants, together, continue to provide the most popular candidates for deterministic explanations. The types of explanation that emerge from modifications and extensions of the statistical relevance model, meanwhile, have much in their favor purely as a models of statistical explanation, even though their scope might not be quite as general as Salmon (1971) had initially claimed.

Thus, this analysis divorces the criteria for reduction from those for explanation in contrast to what occurs in the older analyses of reduction that were discussed in the last chapter. This strategy has two benefits: (i) it makes this analysis of reduction immune to specific criticisms of various models of

explanation; and (ii) the separation has the additional virtue of encouraging a more concentrated focus on the precise nature of reduction than has been customary. For instance, a clear separation of issues connected to explanation in general from those that arise specifically in the context of reduction helps avoid problems about some models of reduction that are, ultimately, problems about explanation (in general). For example, Sarkar (1989) criticized Wimsatt's (1976b) model of reduction (discussed in the previous chapter, § 2.4) on the ground that it is counterintuitive because laws are treated as groups of individual facts. However, this is actually a general problem with Salmon's (1971) account of explanation, which Wimsatt adopts (with some modifcations, as noted in Chapter 2, § 2.5), rather than a problem that is specific to the ideas about reduction that Wimsatt introduces. Separating the issues clarifies this point. Moreover, Wimsatt's cryptic claim that some reductions involve "compositional redescription," which is a specific claim about reduction, is exactly the sort of issue that receives the further attention that it deserves, once the issues connected only with reduction are separated from those connected with explanation in general. The danger of this strategy adopted here is that even the minimal assumptions about explanation that are made may prove to be incompatible with some model of explanation. This cannot be ruled out a priori but, in the absence of a potential candidate model of explanation that raises such a problem, this possibility will be ignored.

3.1. EXPLANATION

To talk of reduction at all, some basic assumptions about "explanation" will be necessary. For the reasons indicated above, these will be kept as general and as mild as possible. There are four such assumptions.

(i) It will be assumed that an explanation begins with a *representation* of the system. The distinction between a system and a representation is important. What, in everyday language, would be called the "same system" can have more than one representation, depending on the context of investigation. A chromosome can be represented as a group of loci (as, for instance, in linkage analysis) or as a physical object (as in cell biology), depending on what the context (of investigation and explanation) is. A cell may be represented as a chemical system of a particular sort, or as a cybernetic system.[1] Obviously, when the same system has different representations, interesting questions about their mutual consistency may arise. These questions can be nontrivial. For instance, it was often suggested that the linear order of loci on chromosomes,

as used in classical genetics, may not be consistent with the order of the physical parts of the chromosome that correspond to these loci; this point will be discussed in detail later in this book (see Chapter 5, § 5.5). Similarly, the choice of a representation is nontrivial; an explanation can fail because of a poor choice of representation. Representations are often indicated diagramatically. These pictorial representations will routinely be used in this book. Perhaps the most sophisticated type of these pictorial representations are the three-dimensional models – such as the double helix model of DNA – that have played a major role in the development of molecular biology. Finally, to emphasize what may be an obvious point, a representation need not be a description of a system in physical space. If **A, a** and **B, b** are each a pair of alleles at two different loci of a diploid organism in some population, **AABb**, **Aabb**, and so on are all adequate representations of genotypes of different individuals for the type of explanation that is attempted in classical genetics (see Chapter 5, § 5.4).

(ii) It will be assumed that what is being explained is some feature of a system as represented – a law it (fully accurately, the representation) obeys, an event in which it participates, and so on. Thus, the account here will be neutral about the role of laws, theories, or individual events as the explananda of a reduction.

(iii) It will be assumed that, given a representation, an explanation involves a process of scientific reasoning or argumentation that will generally be called a *derivation*. "Derivation" as used here must not be confused with the logician's notion of derivation – as, for instance, used in the Nagel-Woodger and Schaffner models discussed in the previous chapter. In Chapter 2, § 2.3, those were called deductions.[2] Derivations will have varying degrees of precision and mathematical rigor. In general, the degree and type of rigor that is appropriate depends on the scientific context. Mathematical rigor, for instance, is no virtue if to achieve that rigor, questionable assumptions must be introduced into the representation of the system. This point will be taken up in § 3.4. Some of these derivations (for instance, those that typically occur in molecular biology) will be relatively trivial. This will generally be the case in contexts where mathematical explanations are customarily not used.[3] In many of these instances, most of the explanatory work goes into the construction of the representation (or model). In many of these cases, the ultimate derivation may be no more than a verbal argument.

(iv) It will be assumed that any explanation uses a set of explanatory "factors" that are presumed to be the relevant ones.[4] These factors bear the "weight" of an explanation in the sense that they provide it with

whatever force or insight that it has. If the explanation is to be put into deductive–nomological form, these are the factors that will be referred to in the general law that forms the basis for the explanation, that is, they are what makes the law *nomological*. If the explanation has the form of Salmon's list, then these are the factors used to partition the reference class into homogeneous subclasses. Using "factors," therefore, provides a neutral and unified term to refer to the relevant entities in radically different models of reduction.[5] The list of factors, as the explanation itself, is context-dependent. The context will determine what factors are relevant, that is, when explanations may stop and when they are incomplete.

These assumptions can, no doubt, be formalized further, though it is open to question whether such formalization would add any insight or would rather bias the discussion toward some particular class of formal models of explanation; this point will not be pursued here. In passing, it should be noted that assumptions (i) and (iii) will generally play the same role here as Nagel's conditions of connectability and derivability, respectively. That assumption (iii) will play such a role should be obvious. However, assumption (i) appears rather different, at least in form, from the condition of connectability. The reason that it generally plays the same role as that condition is that the representation of a system indicates how the system fits into the two realms that would potentially be connected through a reduction.

3.2. SUBSTANTIVE CRITERIA AND TYPES OF REDUCTION

With these assumptions about explanation, three criteria will be used to analyze and characterize different types of reduction. These criteria are substantive: they are about what assumptions are made during a (putative) reductionist explanation, rather than about the form that such an explanation may take. Briefly, the criteria are as follows.

(i) *Fundamentalism*: the explanation of a feature of a system invokes factors from a different realm (from that of the system, as represented) and the feature to be explained is a result only of the rules operative in that realm.

(ii) *Abstract hierarchy*:[6] the representation of the system has an explicit hierarchical organization, with the hierarchy constructed according to some independent criterion (that is, independent of the particular putative explanation), and the explanatory factors refer only to properties of entities at lower levels of the hierarchy.

(iii) *Spatial hierachy*:[7] the hierarchical structure referred to in (ii) is a hierarchy in physical space; that is, entities at lower levels of the hierarchy are spatial parts of entities at higher levels of the hierarchy. The independent criterion invoked in (ii) now becomes spatial containment.

These criteria will be discussed in detail in § 3.3, and § 3.5–§ 3.6. On the basis of these criteria, five different types of reduction can be distinguished.

(a) *Weak reduction*: only substantive criterion (i) is satisfied. A genetical example, which will be discussed in detail in Chapter 4, is the attempt to explain phenotypic features of an organism from a genetic basis using the properties of heritability.

(b) *Approximate abstract hierarchical reduction*: only substantive criterion (ii) is (fully) satisfied, whereas (i) is approximately satisfied.[8] This type of reduction arises from type (c) below, when the assumptions or approximations used in the derivation (of what is being reduced) cannot be fully justified from the rules operative in the more fundamental realm. (As will be discussed in § 3.3, the satisfaction of criterion (i) should be seen as a matter of degree.) This type of reduction is perhaps best seen as an intermediate step in the path toward a reduction of type (c). A genetical example is an explanation on the basis of linkage analysis (see Chapter 5, § 5.4), when not all the properties that have been assumed for the various loci and alleles involved can be fully justified. However, reductions of this type are rare in genetics and will not be considered any further in this book.

(c) *Abstract hierarchical reduction*: only substantive criteria (i) and (ii) are satisfied. Reduction in classical genetics is of this type.[9] The set of alleles and loci form a hierarchically structured genotype. The rules of genetics are assumed to be more fundamental than those governing the phenotype. However, this hierarchy is not necessarily embedded in physical space. This will be discussed in detail in Chapter 5, § 5.5. In the genetic context, that is, in the context of this book, this type of reduction will be called "genetic reduction."

(d) *Approximate strong reduction*: only substantive criteria (ii) and (iii) are (fully) satisfied, whereas (i) is approximately satisfied. This type of reduction arises from type (e) when, as in the case of type (b), the explanations involve approximations that cannot be fully justified. A genetical example (see § 3.6) is the use of "information"-based explanation in molecular genetics, even after it becomes clear that there is no fundamental theory of information transfer that can provide a basis for such explanations (Sarkar 1996). An even more important example is the use of the lock-and-key model of macromolecular interaction in

explanations, if the reduction is being assumed to be a reduction to physics (or chemistry); see § 3.6 and Chapter 6, § 6.4.

(e) *Strong reduction*: all substantive criteria (i), (ii), and (iii) are satisfied. This is the type of reduction where the properties of wholes are explained from those of the parts. Note that when a move is made from types (b) and (d) of reduction to types (c) and (e), respectively, what is assumed as the fundamental realm may change. There is no constraint on what the interactions between the lower elements of an (abstract) hierarchy may be, as long as they are assumed to be more fundamental than those at higher levels. However, once the entities in the hierarchy become spatial parts, their interactions are defined by the known interactions of these spatial parts. In natural scientific contexts, these interactions will be physical interactions, where "physical" is to be construed broadly to include all chemical, macromolecular, and other such interactions from any of the physical sciences. This is the type of reduction that was involved in the mechanical philosophy, the kinetic theory of gases, and, as will be shown in Chapter 6, in many explanations in molecular biology. "Strong reduction," in natural scientific contexts (including that of this book) will also be called "physical reduction."[10]

Types (a), (c), and (e) are the most interesting types of reduction in the context of genetics. In general, it is open to question whether types (b) and (d), neither of which requires the full satisfaction of criterion (i) (that is, fundamentalism), should be regarded as types of reduction at all. Moreover, when the fundamentalist assumption fails, one could wonder whether the explanation at hand is, in fact, an explanation at all. These points will be taken up in detail in § 3.4.

Nickles (1973) was the first to make a distinction between what he called "domain-preserving" and "domain-combining" reductions which, in the classification given here, is basically a distinction between weak reduction [type (a)] and all the others. For domain-combining reductions, Nickles assumed that some variant of Schaffner's (1967b) model is appropriate. Domain-preserving reductions, according to Nickles, occur between a theory and its successor. For him, the succeeding theories get reduced to the preceding ones: special relativity, for instance, gets reduced to Newtonian mechanics when an appropriate limit is taken, such as the speed of light, $c \to \infty$. In such cases there is clearly no explanation of the more general reduced (special relativity in the example) theory by the less general reducing theory (Newtonian mechanics).

Such "reductions" obviously cannot be explanations – a preceding theory cannot explain its successor in any reasonable sense of explanation.

Moreover, this use of "reduction," which seems to have been borrowed from mathematics, is unusual in a scientific context.[11] Wimsatt (1976b) accepted Nickles' distinction and clarified it as a distinction between "intra-level" or "successional" reduction (Nickles' "domain-preserving" or "weak reduction" here) and "inter-level" reduction, by which Wimsatt meant what is being called "strong reduction" here. For interlevel reduction, Wimsatt offered the model that was discussed in Chapter 2, § 2.4. Wimsatt (1995) has something at least akin to the spatial hierarchy criterion [criterion (iii)] in mind when he refers to "material compositional" levels of organization. Since he recognizes other kinds of hierarchies, at least implicitly, he assumes the distinction made here between abstract hierarchical reductions of type (b) or (d) and spatial hierarchical reductions of type (c) or (e).[12] The classification developed in the preceding paragraphs captures these distinctions and offers a finer resolution of the types of reduction that may be separated on the basis of substantive criteria.

3.3. FUNDAMENTALISM

Reduction is pursued because of a belief that some other realm is more fundamental – that is, it can provide deeper understanding, can correlate disparate insights, and so forth – than the one that has been studied. It is necessarily a fundamentalist enterprise at least in this mild sense. This, rather than any sort of more ideological or ontological fundamentalism, is what the substantive criterion (i) tries to capture. It incorporates three distinct requirements:

(a) that a potential reduction draws its explanatory factors from a different realm;
(b) that the rules from that realm are, for some reason or other, considered to be more fundamental than those of the original realm; and
(c) what is to be explained can be derived from these rules using only fully justifiable logical, mathematical, or computational procedures.

In this analysis, these three requirements will not all be accorded the same status. The first two have to be met in order for criterion (i) to be approximately satisfied. Satisfaction of the third requirement is a matter of degree. If the first two requirements are met, but the third is not or is only met to a very limited extent, it will be said that the substantive criterion (i) is only approximately satisfied.

The asymmetry of status of the three requirements listed here needs some justification. The first requirement is necessary to distinguish reductions from any explanation: unless what is explained and what does the explaining

come from different realms, all explanations will turn out to be at least weak reductions. However, in this type of reduction, there is some ambiguity about what "realm" can be taken to mean. [Once criterion (ii) is introduced, "realm" can be unambiguously specified as referring to a level in the hierarchy; see § 3.5.] Roughly, realms will be considered to be different if they are the domains explored by different research traditions. (Even special relativity will have a different realm, in this sense, than classical mechanics. Thus, it will make sense to talk of the reduction of classical mechanics to special relativity.) The rules operative in the more fundamental realm play, here, the role that the reducing theory (or branch) played in accounts of reduction such as those of Nagel or Schaffner.[13] In the discussion that follows, this realm will be called the "**F**-realm"; its rules will be collectively referred to as "**F**-rules"; and "**F**-justified" will mean fully justified on the basis of the **F**-rules.[14]

The second requirement is almost trivial in the sense that unless some such fundamentalist assumption is made, it is hard to see why a putative explanation would be an explanation at all. However, such an assumption was not introduced generally in explanations in § 3.1 in order to assume as little there as possible. Thus, that discussion permits the sort of explanation that involves telling a plausible story – for instance, the kind of story that is rampant in descriptions of evolutionary history or in other "historical explanations." However, reduction, as construed here, at least requires explanation of a somewhat stronger nature. That is why the second requirement has been made explicit, and also why that requirement, along with the first, must be met by every reduction.

The situation with the third requirement is rather different. If scientific explanations were always logical deductions of the sort that were required by the logical empiricists, or if they were even fully rigorous mathematical arguments with no implicit problematic assumptions, then the third requirement would be gratuitous. All explanations would then be naturally **F**-justified. However, especially in contexts where explanations bridge two different realms, approximations are endemic. Moreover, as will be seen in the next section, some types of approximations raise serious epistemological and interpretive problems. This has often been recognized, even in the context of physics, where, since arguments are usually in mathematical form, it is easier to trace and analyze approximations.[15] In particular, approximations may introduce factual assumptions about the system and thereby become, as Leggett (1987, p. 116) has put it, "more or less intelligent guesses, guided perhaps by experience with related systems." In circumstances such as these, **F**-justification will clearly be lost. To get a better grasp on this problem, a more systematic treatment of approximations will be attempted in the next section.

3.4. APPROXIMATIONS

On those rare occasions when logical empiricists and their analytic descendants have addressed the fact that approximations are endemic to scientific explanations, they have generally attempted to incorporate them into the deductive–nomological model to have "approximative D-N [deductive-nomological] explanation."[16] Usually the strategy has been similar to that in Schaffner's (1967) model of reduction, where what is reduced is a theory that has a strong resemblance to the original target of reduction. Discussions of this sort only avoid the problematic substantive issues raised by approximations which, in turn, require a careful treatment of the different types of approximations that are routinely made.[17] The following six sets of distinctions are designed to help such a classification. They almost certainly do not exhaust all the interesting distinctions about approximations that may be made, as they are designed specifically to help address questions that are pertinent to reduction.

(i) Approximations may be *explicit* or *implicit*. There are at least two standard strategies of implicit approximation.
   (a) The invocation of a customary procedure that implicitly makes an approximation. In classical genetics (for instance, in linkage analysis) it is standard to assume implicitly that crossing-over (that would lead to recombination) will not occur within a gene-specifying segment of DNA (see Chapter 5, § 5.4). This is innocuous in almost all contexts. However, there often are more problematic implicit assumptions in linkage analysis, for instance, that the penetrance of an allele – that is, the probability that it will have a recognizable effect – is equal to 1.0.
   (b) The invocation of a model or formula that makes such an approximation. The use of atomic models with spherical atoms and definite surfaces is perhaps the most routine example of such an approximation in molecular biology (see Chapter 6, § 6.2).

   There is much to be said for keeping approximations explicit: it makes it easier to gauge the effects of the approximation. But a stricture that all approximations should always be explicit would probably prove cumbersome in many explanatory contexts: the socialization of scientists guards against errors from most common implicit assumptions. In general, implicit approximations – like other implicit assumptions – are more likely to be made explicit when an explanation involving them runs into trouble. Implicit approximations are not necessarily problematic

for reduction. However, when an approximation is not **F**-justified, and this is not recognized because the approximation is implicit, there is a potential for mistakenly judging a reduction to be successful as well as for incorrectly classifying it by type.

(ii) Approximations may be *corrigible, incorrigible in practice*, or *incorrigible in principle*. Corrigibility is not to be construed absolutely. Rather, all that is required is the knowledge of some procedure for decreasing the effects of an approximation. Usually, this involves a procedure for introducing corrections. In classical genetics, the assumption of an infinite population is corrigible in principle and also in practice in many circumstances. The crucial steps were initiated by Fisher (for instance, 1930) and Wright (for instance, 1931) who investigated stochastic models of gene transmission, that is, models with a finite number of individuals in a population. Since reduction in classical genetics (see Chapter 5) uses rules from population genetics among its **F**-rules, the diverse strategies for the incorporation of population size into population genetics provide methods for correcting the approximation that the population size is infinite. However, these methods are cumbersome in all except the simplest (for instance, one locus) models of genetic influence.

   In molecular genetics, assumptions about the size and shape of atoms within a macromolecule are incorrigible in principle if the **F**-rules that are used are those from quantum mechanics, which are necessary for a general account of chemical bonding, but which do not allow such atoms to exist (see Chapter 6, § 6.3). There is no general procedure for determining when an incorrigibility in practice reflects an incorrigibility in principle. In a particular context, however, it is often possible to make this judgment.

(iii) The maximal effects of an approximation may be *estimable, not estimable in practice*, or *not estimable in principle*. The question of estimability is different from that of corrigibility because even if the effect of an approximation can be estimated, for instance, up to an upper bound, it need not be removable. Conversely, an approximation may be partly corrigible without its (full) effect being estimable. The inability to estimate the effect of an approximation may be due to theoretical or experimental reasons. For instance, as will be seen in Chapter 4 (§ 4.6), the effect of assuming that the variance of a phenotypic character can be approximately represented as the sum of a genotypic and an environmental variance cannot be estimated in many experimental situations. The problem is particularly severe in the case of human populations

where, for obvious ethical reasons, systematic breeding experiments cannot be carried out. While the latter problem may be regarded only as one of estimation in practice, the fact that not all ranges of possible genotypes and environments can ever be explored is a problem of estimation in principle.

(iv) Approximations may involve: (a) only *mathematically justified* procedures (such as taking limits); (b) only **F**-justified procedures; (c) both of these; or (d) neither.[18] **F**-justification is as strong a condition as mathematical consistency. Justification must come from prior factual commitments made on the basis of the **F**-rules, that is, not through the implicit introduction of new assumptions into a derivation. In population genetics the limit of an infinite population is both **F**- and mathematically justifiable. In molecular biology, if the **F**-realm is taken to be either quantum mechanics or quantum chemistry, the shapes attributed to the atoms are not **F**-justifiable.

(v) Approximations may be *context-dependent* or *context-independent*. Once again, this is a question of degree. In classical genetics, including segregation and linkage analysis, the number of alleles at any locus that are assumed to be relevant to a particular problem is a context-dependent approximation (and one that, though corrigible in principle, is usually not corrigible in practice). The usual procedure is to consider all easily distinguishable alleles as distinct and to lump all the others together as a single allele (the "normal" one). In molecular biology the assumptions about the behavior of water molecules (especially how they tend to form ordered structures) that are invoked to explain the hydrophobicity [see Tanford (1980)] are context-dependent approximations in the sense that water is not assumed to have exactly the same properties in other situations, for instance, in the discussion of ionic reactions (in solution).[19] Context-dependence is a useful heuristic for suspecting the lack of **F**-justification of an approximation. However, caution should be exercised in the use of this heuristic; for instance, there is no such justificatory problem for the context-dependent approximation in classical genetics when the number of alleles at a locus is set to one more than those that are easily distinguishable.

(vi) Approximations may involve *counterfactual* assumptions, or not. Throughout this book, "counterfactual" is to be construed simply as referring to assumptions not permitted by the **F**-rules that are assumed.[20] Counterfactual assumptions are endemic, though the extent to which they involve serious violations of **F**-rules is often hard to gauge. The

ubiquity of counterfactual assumptions raises the obvious ontological question of the status of results (such as the existence of certain processes) obtained using them. One response would be to assume that all theories are approximations, and that the "underlying world" poses no problem – in effect, use instrumentalism to rescue realism. Another is to admit that these counterfactual assumptions should be regarded as new factual hypotheses of the special sort indicated in the text. This would already raise problems about whether the fundamentalism criterion for reductions can even be approximately satisfied but, perhaps worse, this raises even more serious conceptual difficulties. What kind of factual hypothesis is it that allows the number of loci or alleles to be infinite, especially in a finite population? At the very least, further distinctions between types of counterfactual approximations will be necessary. This point will not be pursued here. Rather, the position that will be adopted is that counterfactual approximations, however they are interpreted, pose problems for judging the success and classificatory status of reductions.[21]

If an approximation, preferably explicit, is corrigible, its effects estimable, both **F**- and mathematically justified, context-independent, and involves no counterfactual assumption, it is presumably philosophically unproblematic, whether one is interested in only epistemological questions or also in ontological ones. As has previously been noted, too many philosophers who have analyzed reduction have assumed that derivations involve no approximation, or only approximations of this sort, in which case they can potentially be removed to recover, at least partly, the logical cleanliness that these philosophers value. Note that even in this situation, the use of approximations does not allow at least one conclusion that the traditional accounts of reduction assume to be true, namely, that reductions are transitive.[22] A sequence of approximations, however justifiable each may be by itself, need not be so justifiable.

Scientific developments rarely proceed according to the strictures on approximations imposed at the beginning of the last paragraph. Most approximations violate many of these strictures. As has been noted before, that an approximation is implicit is not necessarily problematic, and not much that is sensible can be said about the problem of counterfactual assumptions in general. There is also no general procedure for systematically judging the extent of the problems posed by the other violations of those strictures, that is, by violations of corrigibility, effect estimatibility, mathematical justification, and **F**-justification, either individually or jointly. Suffice it, here, to

note only five general points about these approximations, the first four of which are generally relevant to explanations and emphasize the fact that approximations are not always epistemologically devastating. The last, which concerns what is perhaps the most problematic type of approximation, is particularly troublesome in the context of reduction.

(i) Even if an approximation is incorrigible, an explanation may have force, especially if the data to be explained have more uncertainty than that induced by the approximation. In the genetics of natural populations, where accurate measurements are often impossible in practice, this is often the case.

(ii) If the effect of an incorrigible approximation can be fully estimated, the point made in (i) can be made even stronger to test whether the approximation leaves an explanation within the smear of the data. In classical genetics one could, for instance, show how much difference full or no dominance would make for a model that assumes nothing in particular about dominance (for an example, see Chapter 4, § 4.5).[23]

(iii) Even if the effects of an approximation cannot be estimated, an approximation may not be particularly problematic. It might, for example, be corrigible to a significant extent. Or, it might be both **F**- and mathematically justified and involve no counterfactual assumption, which would be reason enough to tolerate it.

(iv) Moreover, experience has shown that mathematically suspect procedures do not doom a scientific research program. More often, as with Galton's regression procedure for the study of continuous traits in populations, new mathematics can be constructed to rationalize what should be regarded as earlier heuristic procedures.

(v) However, if an approximation is not **F**-justifiable, there is necessarily a problem with the satisfaction of the fundamentalism criterion for reductions. In the discussions of this book, the degree to which that criterion will be judged to have been satisfied will largely depend on the extent to which the approximations (if any) that are used during a derivation (in a reduction) are **F**-justifiable. (These arguments, in the present context, will have to be qualitative.) In passing, it should be noted that while the lack of **F**-justification is philosophically troublesome, it may also open the way for scientific developments, as the **F**-rules may get refined or changed or new realms that can serve as **F**-realms get invented.[24] In fact, it is at least arguable that the ability to choose which type of violation of existing **F**-rules is likely to be scientifically fruitful, and which would not, is an important measure of scientific insight.[25]

3.5. HIERARCHICAL ORGANIZATION

It was explicitly assumed in § 3.1 that an explanation begins with a representation of a system. This representation may or may not be hierarchical. The abstract hierarchy substantive criterion (§ 3.2) of reductions requires that this representation be hierarchical with the system at the top of the hierarchy and other entities at lower levels; that the hierarchy be constructed in accordance with some "independent" criterion; and that the explanation of some feature of the system refer only to factors at lower levels. Using the terminology that has been developed since, the **F**-realm can only include entities at these lower levels of the hierarchy, and the **F**-rules are those that they follow.[26] That the hierarchy be constructed according to an "independent" criterion means that it should not have been posited only for the sake of the explanation at hand, that is, there must be some independent reason for constructing it. There are at least three ways in which this independence condition could be satisfied (these are not independent): (i) there could have been other explanations (perhaps even of similar phenomena) that used such a construction, as was the situation with seeking Mendelian explanations around 1900, when it was not clear how the Mendelian factors would correspond to any physical entities of organisms; (ii) the hierarchy is constructed according to the dictates of a general research program such as the search for physical explanations in biology; (iii) the same hierarchy is used for some entirely different type of explanation. For instance, the genetical hierarchy can be used to explain the details of the origin of a complex trait (gene expression), while the independent reason for constructing it could be provided by the transmission of the trait. In the examples of genetic reduction discussed in Chapter 5, this is precisely the explanatory strategy that will be followed.

A standard way to represent such a hierarchical explanation is by a directed graph with at least one sink, no cycles, and with the edges being those admitted by the hierarchy, and their direction being determined by the direction of explanation.[27] Since not all entities of a hierarchically organized system will have to be invoked in every explanation, this graph will only select those entities whose interactions are relevant. (It can, therefore, be embedded in a larger graph representing the entire hierarchy.) One of the sinks in the graph represents the system whose features are being potentially explained. The "level" of an entity is the number of edges from that sink to the vertex representing the entity. The higher that this number is, the lower the level (see Figure 3.5.1).

Particularly simple hierarchies can be represented as trees, where there is only one sink, and every other node has a unique ancestor and any number of descendants (including 0). A typical hierarchical structure encountered

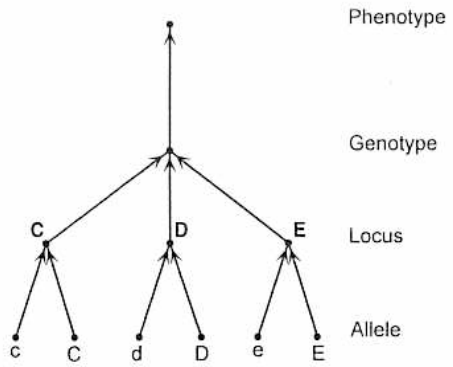


Figure 3.5.1. Graph Structure of RH Blood Group (in humans). Three (closely linked) loci, **C**, **D**, and **E**, determine the RH phenotype of an individual. It is usually assumed that there are two alleles at each locus: these are symbolized *c* or *C*, *d* or *D*, and *e* or *E*. The graph shown is for an individual with a *cde/cDe* genotype. The individual is only heterozygous for the **D**-locus. The arrows are in the direction of explanation.

in genetics is shown in Figure 3.5.1. A phenotypic feature (of humans) is the *Rhesus* (RH) blood group (Vogel and Motulsky 1986). The particular group that an individual has is explained on the basis of three loci, and the two (of many possible) alleles at each locus. This graph has a simple tree structure. However, should a locus be pleiotropic, and several features be studied simultaneously (say, as a complex trait), then the tree structure may be lost even in a genetic context. An example, represented in Figure 3.5.2, is a morphological abnormality (called the podoptera effect) in *Drosophila melanogaster* that is explained by abnormalities in both the wings and the legs (Goldschmidt 1955). These both seem to arise from a mutation at a single locus (with two alleles). The explanatory graph is no longer a tree.

Can the hierarchy criterion fail in biology? Sometimes, though not in the usual circumstances encountered in genetics. It fails in those evolutionary explanations that invoke fitness and, rather than postulating fitnesses as primitive properties of entities, attempts to explain fitnesses in terms of higher-level entities such as the (ecological) environment of the entities.

Note that directed graphs of this sort can represent any (abstract) hierarchy, and this representation makes no commitment to any entities or processes in physical space. Explanations represented in this way reflect intuitions about reduction because there is a definite hierarchy, allowing the assignment of "levels" to which different factors belong, and the requirement

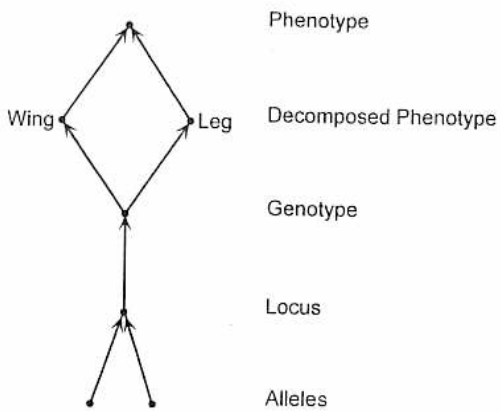


Figure 3.5.2. Graph Structure of Podoptera Effect (in *Drosophila melanogaster*). The phenotype is subdivided into two separate phenotypic characters, a wing abnormality and a leg abnormality. They are both caused by the genotypic character of a single locus with two alleles. The arrows are in the direction of explanation.

that there be no cycles ensures that explanations proceed in a definite direction, that is, intuitively from lower through higher levels toward a sink. If an explanation only fully satisfies the hierarchy criterion – that is, it is representable in this way, while barely satisfying the fundamentalism criterion – then it is a reduction of type (b), which is an approximate abstract hierarchical reduction. If the fundamentalism criterion is also fully (or, at least, to a great extent) satisfied, then the explanation is a reduction of type (c), an abstract hierarchical reduction. As noted before, distinguishing between type (b) and type (c) reductions does not add much insight to discussions of explanations in genetics. This distinction will, therefore, be ignored here and attention will be restricted to type (c).

3.6. WHOLES AND PARTS

Now suppose that the directed edges along one of these graphs not only represent the direction in which explanation must proceed, but also the relation "is a spatial part of." The hierarchy will then be said to be a "spatial heirarchy." The levels of the hierarchy are then usually called "levels of organization."[28] What this means is that the independent criterion by which the hierarchy is constructed is that of spatial containment. As in the general case of using graphs to represent the parts of the hierarchy that are

relevant to a given explanation, not all spatial parts of entities at each such level must occur in the directed graph representing the explanation. The only ones that will be vertices of the graph are the ones that are relevant to an explanation.

Figure 3.6.1(a → f) is an abstract representation of the *lac* operon in *Escherichia coli*. It is intended, at a gross topological level, to reflect the actual three-dimensional structure of the system. A regulator locus or site (i) is responsible for the synthesis of a repressor molecule that binds to an operator locus or site (o) in the absence of the inducer molecule (e.g., lactose). Presumably because of steric hindrance, when the repressor molecule is bound to the operator locus, synthesis of lactose does not take place. In the presence of lactose, because of some interaction between the lactose and the repressor molecule, the latter can no longer bind to the operator locus. In such a circumstance, β-galactosidase, which digests lactose, can be produced through the usual cellular transcription and translation processes. Another protein that is also produced is a permease molecule that aids the transport of lactose across the cellular membrane. The function in lactose digestion of a third protein, acetylase, if any such function is present, has not yet been deciphered. Figure 3.6.2(a,b) is the graph-theoretic representation of the explanation of gene regulation by the operon model.

This full significance of this example will become clearer in Chapter 6, § 6.2. Suffice it here to note six points that are particularly relevant to the present context.

(i) Once spatial representations such as Figure 3.6.1 are available, the graph-theoretic representations may not be particularly informative, at least in relatively simple cases. Graph-theoretic representations will be dropped in any further consideration of strong reductions in this book. Their utility is generally limited to situations when one needs a neutral way to represent what are abstract (nonspatial) hierarchies.

(ii) The abstract representation in Figure 3.6.1 is supposed to reflect actual steric properties. Most importantly, it reflects the belief that steric lock-and-key fits are the critical mechanism by which the various biological interactions are mediated. This point will be taken up in Chapter 6, § 6.2 and § 6.3. Ultimately, the abstract representation would be replaced by an actual model with (models of) the relevant atoms in place. This model is usually constructed by crystallizing the molecular complexes and solving the crystal structure, which is a laborious task.

(iii) The graph-theoretic hierarchical representation of the system, though obviously dependent on the spatial relations in the system, is not visually isomorphic to the three-dimensional structure of the system.

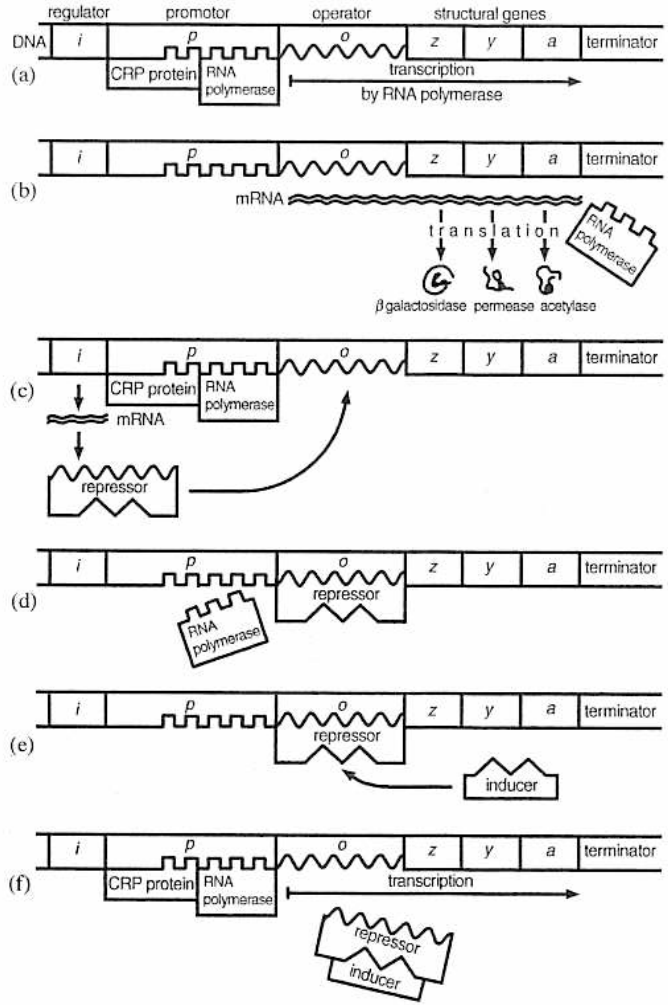


Figure 3.6.1. Gene Regulation at Lac Operon [after Strickberger (1976) p. 680; reprinted by permission of Prentice-Hall, Inc., Upper Saddle River, NJ]. (a) In the absence of a repressor protein [see (c)], and in the presence of CRP-protein attached to part of the promoter site (p), RNA polymerase also attaches to p, and transcription begins; (b) subsequently, translation takes place, and the proteins β-galactosidase, permease, and acetylase are produced; (c) transcription from the regulator site i and subsequent translation produces a repressor protein molecule that attaches to the operator site (o); (d) attachment of the repressor protein to o prevents transcription through (steric) hindrance. It is possible that it also prevents the attachment of RNA polymerase to p through the same mechanism. (e) An inducer attaches to the repressor and the complex detaches from o; see (f). This takes the system to the state described in (a).

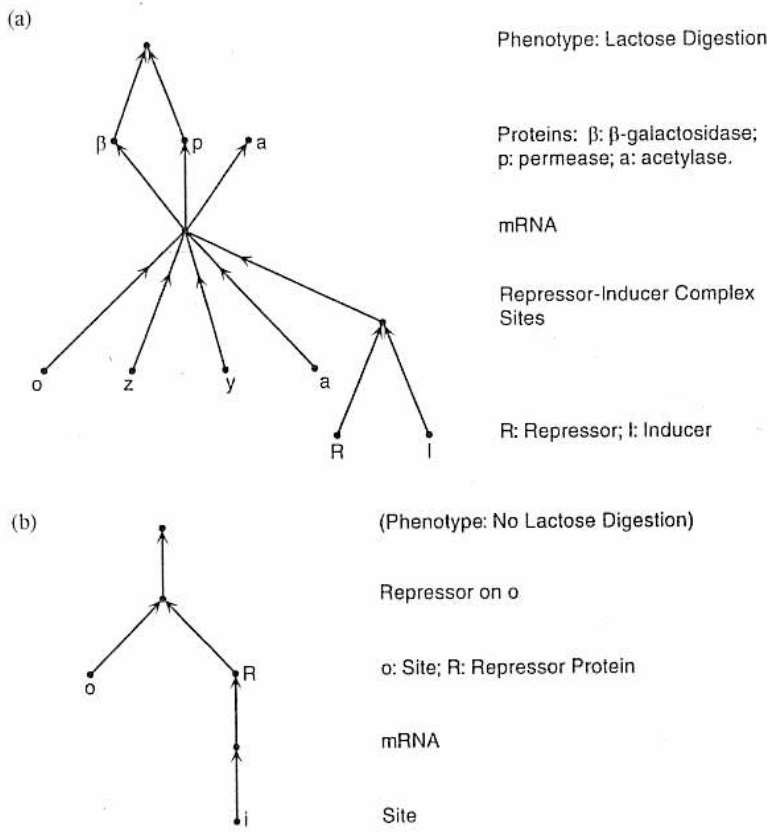


Figure 3.6.2. Graph for Gene Regulation at Lac Operon. (a) Lactose digestion is taking place. This corresponds to Figure 3.6.1 (a), (b), and (f). Note that not all details, such as the involvement of RNA polymerase, are shown. It is assumed that acetylase does not play an identifiable role in the digestion of lactose. (b) No lactose digestion is taking place. This corresponds to Figure 3.6.1 (d).

(iv) There is another kind of failure of isomorphism that is perhaps even more interesting. The "natural levels" determined by the spatial hierarchy of organisms (organism → organ → tissue → cell → organelle → macromolecule → molecular moiety, and so forth) are not usually the same as the levels in the graph representing the reduction (which, by definition, is what is being called "level" here). The explanation of the behavior of the cell may involve organelles, macromolecules, and solvent particles. This point, which emerges naturally from the discussion of reduction given here, has sometimes been systematically elaborated as the claim that reductions can require interfield theories.[29] However, this point is nontrivial if (and probably only if) "reduction" is construed according to the Nagel- or Schaffner-type models.

(v) Note that the type of strong reduction this explanation provides is incomplete in the sense that some parts of the model, such as the appeal to steric hindrance, are no more than conjectures at this stage. However, they are valuable as heuristics in the sense that they provide specific targets for future inquiry.

(vi) The explanation is incomplete in yet another way insofar as one can ask why steric hindrance occurs. After all, this seems to have been a principle gathered from mechanical models built with everyday objects rather than from fundamental physical principles. It is also possible that this incompleteness will generate further inquiry. However, historically, this has not been the case in molecular biology. This point will be briefly addressed below and in more detail in Chapter 6.

Note that the satisfaction of the physical hierarchy regulatory criterion requires the satisfaction of the abstract hierarchy criterion. If an explanation fully satisfies only these two criteria and approximately satisfies the fundamentalism criterion, it is a reduction of type (d), that is, an approximate strong reduction. If it fully (or to a very large extent) satisfies the fundamentalism criterion, it is a reduction of type (e), that is, a strong reduction. Strong reductions have a long and illustrious history, from the mechanical philosophy through the kinetic theory of gases at the end of the nineteenth century to contemporary molecular biology.

However, whether reductionist explanations in molecular biology are actually strong reductions or only approximate strong reductions [of type (d)] is a rather subtle question.[30] The answer depends critically on what is taken to be the **F**-realm. If the **F**-realm is taken to be the physics and chemistry of macromolecules, where the **F**-rules have been determined empirically, then it is plausible to argue that strong reductions are taking place. This is what was implicitly assumed in the discussion of the operon model that was given

above. However, there is something epistemologically unsatisfactory about such a move. The physics and chemistry of macromolecules, at least at this point in the development of science, is not a particularly well-developed or organized domain of inquiry. At best, one can say that there are some experimentally gleaned rules (such as the importance of steric hindrance) that define it. Moreover, at present these rules cannot be reduced to any better-defined (and more fundamental, in this context) realm such as physics and chemistry. Not only have systematic attempts in this direction been very rare, but also the types of approximation involved suffer from all the problems noted in § 3.4; this point will be developed in Chapter 6, § 6.3. Moreover, it is far from clear that the approximations that seem to be necessary in one context would be consistent with those needed in another. In most situations the posited macromolecular interactions assume a spherical water molecule. But for water to have the internal structure to account for the hydrophobic interaction, it cannot be spherical. There is no glaring inconsistency here, but sufficient inconsistency at the level of fine detail to have some reason for worry, if not for serious skepticism.

There is a temptation, therefore, to ask for a reduction where the **F**-realm is fundamental physics or chemistry. The motivation is that most of the interactions in the molecular realm are known to be mediated by quantum mechanics, as in the case of usual (covalent) chemical bonding. The trouble is that the fact noted at the end of the previous paragraph, that there is no epistemologically respectable way to carry out derivations from this realm to the experimental rules discovered for macromolecules, precludes any easy direct reduction of biological phenomena to such a quantum **F**-realm. (Of course, even if the former reduction was possible, it would not guarantee that the total reduction could be accomplished. The "goodness" of approximation is not always a transitive relation, as was noted in § 3.4.) Why has this situation not generated further work? The answer simply seems to be that the usual models of molecular biology, with all the approximations and possible inconsistencies built into them, work very well at the level of experimental detail that is currently available.[31] This is the beauty – and puzzle – of molecular biology from a philosophical point of view. Further discussion of these points will be found in Chapter 6.

3.7. EPISTEMOLOGICAL ELIMINATIVISM

The last six sections of this chapter will very briefly discuss some general philosophical issues that are affected by what is decided about reduction. These discussions are less than comprehensive in two ways: (i) given the

restricted scope of this book, only those issues that are interesting in the context of genetics will be treated; (ii) even when an issue is discussed, those aspects of it that are not particularly interesting in a genetic context will generally be ignored. This is particularly true of ontological questions that are far less pressing in the case of genetics than, for instance, in the case of neurobiology.

This section will reiterate a point about epistemological eliminativism that was already made in the previous chapter (§ 2.4). It has sometimes been thought (for examples, see § 3.12) that reductions or, at least, reductions of type (a), (c), or (e) would lead to the disposal of the reduced rules, theories, and so forth, in favor of those that provide the reductions. These would be appropriately called "reductions with replacements." This is a strong thesis of epistemological eliminativism: the reduced entities would be entirely dispensed with because of reduction. Thus, special relativity would replace Newtonian mechanics, the kinetic theory of gases would replace the thermodynamics (of gases) and, rules about genes would replace those about phenotypic features. There is little to be said in defense of this strong thesis. In fact, from Hull (1972, 1974) on genetics, to P. Churchland (1979, 1981, 1984) and P. S. Churchland (1986) on folk psychology, replacement is usually presented as the type of scientific change that occurs when reduction does not take place. In fact, as modified by Sarkar (1989), Wimsatt's (1976b) model of reduction provides an explicit rationale for not abandoning the laws or theories of the realm being reduced (see Chapter 2, § 2.4).

There are at least two reasons why reduction with replacement is an unlikely possibility.

(i) All known cases of reduction show that using only the rules from the **F**-realm would lead to much more complex explanations than tolerating the rules from the reduced realm.[32] This was the important point that Wimsatt (1976b) made. It is because of this fact that in biology, cell biology and organismic biology continue to be used with much of their traditional conceptual apparatus, in spite of many partial reductions to molecular biology (see Chapter 6).

(ii) In fact, in many cases reductions generate further confidence in the use of rules from the reduced realm because they explicate the exact range of applicability of those rules.[33] The success of special relativity, and the reduction of Newtonian mechanics to it, shows that Newtonian mechanics can be used without worry at low velocities. Similarly, Newton's law of gravitation remains adequate for the strength of gravitational interactions experienced on earth even though it can be reduced to general relativity. Even dubious reductions involving many problematic

approximations can serve this role. A genetic example, Fisher's reduction of biometry to Mendelian genetics, will be discussed in detail in Chapter 5, § 5.2.

There is a weaker thesis of epistemological eliminativism that is more interesting. This would say that, although rules from the reduced realm continue to be used for pragmatic reasons, they receive their epistemological force from the **F**-rules or, in other words, the weight of the explanation is still borne by the **F**-realm. Many different versions of this thesis, flushing out different senses of explanatory weight, can easily be generated. However, there is a compelling general argument against them. If reductions never involved problematic approximations, this thesis could probably be maintained. The approximations that are obviously most problematic are those that are context-dependent approximations and those that cannot be either **F**- or mathematically justified. But even other ones, including those that involve counterfactual assumptions, are reason for worry. These worries are particularly troublesome if the reduced rules have strong empirical support. If, indeed, their support is significantly stronger than that of the **F**-rules, then, along with Cartwright (1983), one could even worry about the correctness of the **F**-rules in the presence of such approximations. Moreover, once all these approximations are banished, very few reductions remain.[34] Therefore, at best, even this weak thesis can only be very occasionally maintained.

3.8. ONTOLOGICAL ELIMINATIVISM

Like epistemological eliminativism, ontological eliminativism comes in two versions, one strong and rather absurd, and the other weak and occasionally defensible. The strong version would replace all entities and processes from the reduced realm with those from the **F**-realm following a reduction. It is hard to find anyone who explicitly advocates such a position.[35] Nevertheless, unless this result is the goal, there is no rationale for the obsession with synthetic identities that was noted in the previous chapter (§ 2.5). All other possible ends of reduction, including the two types of epistemological eliminativism and the weaker form of ontological eliminativism, do not even require biconditionals, let alone identities, as the relation between the entities and processes of the **F**- and other realms. Conditional statements would suffice for these purposes.

The strong form of ontological eliminativism is impossible to sustain. The problem, once again, is a consequence of the ubiquitous approximations that

lead to incompatible properties routinely being ascribed in different realms to what is supposed to be the same entity (or process). It is hard to imagine, unless one suspends all concern for conceptual rigor, how the atom admitted by quantum mechanics can replace all the atoms of chemistry, let alone molecular biology.[36] The attempt to do so, like many other forms of ontological fundamentalism, could generate interesting scientific projects: a devoted quantum mechanical atom-fundamentalist could do good science trying to carry out such an elimination. Strong ontological convictions, however unjustifiable they always seem in retrospect, have the obvious virtue of quite often being able to generate worthwhile research programs. But that is a different question from that of their actual viability in practice.

A weaker form of ontological eliminativism accepts all the problems raised by approximations, and so on, but nevertheless asserts that all entities and processes occurring in some other realm are nothing but those occurring in the **F**-realm. In biology, with physics and chemistry providing the **F**-realm, this sort of eliminativism is perhaps most famously associated with Loeb's (1912) "mechanistic conception" of life. Positions that invoke this weak form of ontological eliminativism usually come with an associated epistemological program of attempting to carry out reductions with as few dubious approximations as possible. Unless such a program is associated with it, weak ontological eliminativism is a rather innocuous and vacuous thesis since there is no other way to spell out the meaning of the "nothing but" in its conception.

Nevertheless, even when there is an associated epistemological program, there are (at least) two ways to deny even the weak version of ontological eliminativism.

(i) One could deny it on ontological grounds and argue that new – or, at least, slightly different – entities and processes exist in different realms. (This would, among other things, explain why epistemologically questionable assumptions are necessary when traversing realms.) Traditional Marxism – or, at least, its dubious official metaphysics, dialectical materialism – required a move of this sort: each level or organization of matter had its own (obviously "emergent") laws and so on.[37] So did vitalism and other such moribund doctrines that probably reflect little more than emotional discomfort with the anemic ontology of the eliminativists.[38]

(ii) Or, one could deny it on conceptual and methodological grounds. The point here is that in the absence of any reasonably consistent scheme that encompasses the different realms of inquiry, it is hard to see how one can make firm (let alone final) ontological commitments about what an

electron, atom, or fitness is. This is the conceptual ground for denying even weak ontological eliminativism. The methodological ground is that it is equally hard to see – unless one is obviously and deeply influenced by some extreme (e.g., Quinean) form of philosophical fundamentalism – why one would want to suggest that entities and processes in the reduced realm are "nothing but" those in the **F**-realm. Commitments of this sort are not necessary to pursue science – to construct useful or intellectually illuminating models of the world. By this point, no reader should be surprised that this is the option that will be endorsed in this book. (There is, no doubt, something reminiscent of logical empiricism in this move; to some extent it captures part of what was valuable in that school's resolute refusal to concern itself with ontology.)

To reiterate the position being advocated here (in this section and the previous), to say, for instance, that a biological system is "nothing but" a physical system has only one useful consequence: It might lead to an epistemological program of the physical investigation of that system. But one can proceed with such an investigation without any such eliminativist move. In fact, as noted in Chapter 2, § 2.2, and as discussed again in Chapter 6, § 6.1, Delbrück, following Bohr, helped initiate molecular biology by pursuing physical explanations hoping that they would ultimately fail.[39] Thus, not only was ontological eliminativism not endorsed, but even epistemological eliminativism was rejected.

3.9. REDUCTION VERSUS CONSTRUCTION

Those who doubt the possibility of reduction sometimes raise the point that, whereas one might find a reductionist explanation of some phenomenon after it has been described in its own realm, one would not have been able to suggest (or in a weaker version would not, in practice, have suggested) such a phenomenon if one only had access to what is known about the **F**-realm. This observation is clearly correct for many cases but it should not be interpreted to be more important than it is. Rather, three points should be noted about it.

(i) The issue here is partly one about the difference between prediction and explanation. At best what this criticism points out is that one would not have predicted all the consequences of some set of assumptions (even if all the rules were fully deterministic). Explanation is usually a weaker category than prediction.[40] Reduction, being a type of explanation, may

not always permit prediction. However, this is a general problem about explanation; it does not provide any compelling reason to doubt the value of reductions.

(ii) If either physics or biology is a guide, what happens in any realm depends on both the entities and processes and the (boundary or initial) conditions of the entities on which those processes act. In biology, those conditions are often critical in the explanation of phenomena – this is one version of what is sometimes called the principle of historicity in biology. The entities and processes are clearly not sufficient to construct the outcomes. Moreover, since the set of possible conditions in which the entities may find themselves is large, it should come as no surprise that biological outcomes – the result of a particular evolutionary and a particular developmental history – would not be predicted or constructed (from some other **F**-realm) in practice. But this, once again, is no argument against the value of reductions.

(iii) Finally, and this is probably the most important point to be made in this context, approximations are not only generally intransitive (as has already been noted in § 3.4), but they are also often degenerate in the sense that there is no unique approximation that is the only correct one in a particular context. When approximations are even to a slight degree context-dependent, there is little chance that the correct approximation can be chosen without reference to what is being explained.

3.10. REDUCTION AND SCIENTIFIC METHOD

The discussion in this book has so far been focused on the questions of whether certain explanations are reductions and, if they are, what type of reduction subsumes them. Yet, there is an entirely different question about the role of reduction in science, namely, the part that it plays in methodology. One aspect of this, reduction as a research strategy, has often been studied.[41] The basic idea here is that research strategies could be designed to search for reductions. How reduction is construed is critically important in this context. If it is construed according to Schaffner's (1967b, 1993a) model, then as Schaffner (1974, 1993a) has noted, reduction is "peripheral" to the practice of science (at least in the case of molecular biology). If, however, reduction is construed according to Wimsatt's (1976b) model, a better case may be made for the claim that the pursuit of reductions is central to many scientific research programs. Though this point will not be pursued at length here – there will be many examples in later chapters – each of the three

important types of reduction [types (a), (c), and (e)] that were distinguished above has been actively pursued in genetics.[42] Liberated from the formal models, the value of reduction as a research strategy in genetics cannot be seriously denied.

Suffice it here only to note that in physics the search for strong reductions was the critical motivation for Maxwell, Boltzmann, and other physicists who pursued mechanistic explanations for the laws of thermodynamics in the latter half of the nineteenth century. In contemporary behavioral and psychiatric genetics the search for reductions, especially through linkage analysis, is the dominant research strategy; this point will be elaborated in Chapter 5, § 5.4. Perhaps the most interesting – and, in many ways, odd – use of reductionist research strategies was in some of the work of the Phage Group in molecular biology in the 1940s, when the reduction of biology to physics was pursued even when some members, including Delbrück and Stent, were expecting and hoping that reduction would ultimately fail.[43] Such a possibility, of course, illustrates how the pursuit of reduction as a research strategy is a different issue from whether a given explanation is reductionist. Reductionist research programs may fail to provide explanations at all (as will be argued for one kind of attempt to reduce phenotypic traits to genotypic ones in Chapter 4), or they may come up with perfectly respectable explanations that are not reductions (of whatever type is desired).

A second and more important role for reduction in methodology arises from the fact that successful abstract hierarchical and strong reductions [of types (c) and (e)] are routinely used to generate investigative tools where the continued success of new reductions is assumed and used to chart out a domain. In the abstract, the point is this: reductions of these types assume a particular hierarchical structure in the representation of a system. If it is assumed that a particular feature can be reduced according to such a pattern, then, from the existence of the feature, the existence of these internal structural features can be inferred. For instance, the reduction of some phenotypic traits to genetics between 1900 and 1912 generated the programs of segregation and, especially, linkage analysis, which assumed that such reductions would occur for other traits, and then used this assumption to map traits to factors and, in the case of linkage analysis, to specific linkage sets. (These processes will be analyzed in more detail in Chapter 5, § 5.3 and § 5.4.) Further, when these linkage sets were interpreted as corresponding to chromosomes, this procedure led to the systematic mapping of genes (loci) onto specific regions of chromosomes, starting with the work of the Morgan group at Columbia University in the 1910s and 1920s and continuing to this day.[44]

3.11. THE VALUE OF REDUCTIONS

Chapter 2 was quite critical of Nagel's and other formal models of reduction that regarded it necessarily as a relation between theories. However, unlike Schaffner (1967b, 1993a) and most other proponents of formal models of reduction, Nagel (1961) did not limit his attention to formal issues but coupled his formal model to a discussion of nonformal issues about the value of reductions in scientific practice. That discussion is characteristically illuminating. The formal model, Nagel correctly noted, "[did] not suffice to distinguish trivial from noteworthy scientific achievements."[45] Reductions had to satisfy at least three conditions to be of significance.

(i) The assumptions of the reducing theory (its laws and theoretical postulates) should not be ad hoc in the sense of having been introduced only to carry out a particular reduction. Rather, they must be "supported by empirical evidence possessing some degree of probative force" (p. 358).

(ii) The reducing theory "must also be fertile in usable suggestions for developing the [reduced theory], and must yield theorems referring to the latter's subject matter which augment or correct its currently accepted body of laws" (p. 360).

(iii) The reduced and reducing theories must be at a stage of their development where such a reduction aids the development of the reduced theory and, at the very least, does not frustrate its development by shifting interest away from it to the reduction or the reducing theory.

The continual references to theories (and to theorems) is no more than an expected artifact of Nagel's formal model. However, once those references are removed and replaced by the references to the different realms under consideration, little needs to be added to Nagel's account. Nagel shows no tendency to eliminativism of either kind, has a good appreciation of the scientific process as being dynamic, and has a healthy respect for the pragmatic component in the evaluation of scientific developments. Perhaps most importantly, he does not offer a blanket endorsement of reductionism, that is, of the thesis that (ultimately) reductions are bound to be successful in a given field. As the discussion in Chapter 6, especially § 6.6, will show, part of the defense of (strong) reductionism in molecular genetics will be to point out that attempts at reduction have been remarkably fertile in generating fruitful fields of inquiry.[46]

The only one of Nagel's conditions that requires some systematic elaboration is the first. Nagel basically restricts what he calls an ad hoc assumption to what would, in the framework developed in this chapter, be called

context-dependent assumptions (though not all context-dependent assumptions should be regarded as ad hoc) generally involved in approximations. However, these are not the only assumptions that would decrease the value of a reduction. As the discussion of § 3.4 showed, significant incorrigibility and an inability to estimate the effects of an approximation would both (independently) decrease the value of a reduction. Lack of mathematical justification could also be problematic and, of course, lack of **F**-justification may make a putative reduction fail to satisfy the fundamentalist criterion.

3.12. THE UNITY OF SCIENCE

The demand that all science, if not all knowledge, be unified into a single structure has been a popular and recurrent feature of the Western philosophical tradition since at least the seventeenth century. When the logical empiricists attempted their reformulation of philosophy, which was heavily dependent on their interpretation of science, the unity of science served as one of their most important regulative principles. It is easy enough to see how systematic reduction could lead to the unity of science. As Quine ([1977] 1979, p. 169) puts it: "Causal explanations of psychology are to be sought in physiology, of physiology in biology, of biology in chemistry, and of chemistry in physics – in the elementary physical states."[47] If (and, presumably, only if) all these explanations {which would at least be weak reductions [type (a)]} are forthcoming, then science would be unified, at least with respect to epistemological concerns, with the elementary physical states and the rules governing them providing the unifying framework. This is Quine's version, or at least one of his versions, of physicalism.

In general, however, with Feigl (e.g., 1963) being perhaps the most notable exception, the logical empiricists did not endorse reduction as the route to the unity of science. Rather, the unity of science was to be achieved through what they called "physicalism," which was a rather different position than Quine's version (as described above). Physicalism went through several changes as the logical empiricist program unfolded – and, as some would say, disintegrated – and it was interpreted rather differently by the different figures associated with it, including Neurath and Carnap.[48] However, it was almost always the thesis that the same language be chosen to describe the experimental domains of all the sciences. Originally, this was the language of (presumably theoretical) physics, later it became the language describing everyday physical objects such as chairs and tables, and finally simply any nonsolipsist language (in the sense that it is a language

in which all statements are intersubjectively confirmable (Carnap 1963).[49] Logical empiricism's demand for the unity of science is a rather innocuous doctrine; certainly, it makes no demand for any of the types of reduction being considered here.

Indeed, the thesis that the unity of science can be achieved through reduction seems only to have been clearly formulated by Oppenheim and Putnam (1958) in a manifesto written when logical empiricism was already on the wane. Oppenheim and Putnam have strong reductions in mind, and they assume that reductions involve the derivation of laws. They distinguish six levels of organization: those of elementary particles, atoms, molecules, cells, multicellular living organisms, and social groups. They are, of course, fully aware that strong reductions between all these levels were far from forthcoming; their thesis was intended as no more than a "working hypothesis." But, if that working hypothesis is supposed to describe all of scientific practice then, not only is it descriptively false but – as Fodor (1974) has argued – there are sound methodological reasons to doubt its utility.

Reductions can be unilluminating, as Nagel (1961) realized. Counterfactual, context-dependent, and other problematic assumptions made during reductions may not even allow weak epistemological eliminativism in many situations, let alone strong epistemological eliminativism (or, for that matter, any form of ontological eliminativism). At best, most reductions establish some (weak and not very precise) form of consistency between various realms; an explicit example of this kind will be treated in detail in Chapter 5, § 5.2. It will not be assumed in this book that all types of reduction necessarily contribute to unification. Approximate hierarchical or strong reductions [of types (b) and (d)] clearly do not.[50] However, those reductions that permit epistemological eliminativism do make such a contribution. But these may well be rare. Moreover, to the extent that reductions contribute to added confidence in the reduced theories, laws, etc., but involve problematic assumptions, they may well contribute to the disunity of science in practice.

Note, moreover, that one of the most successful unificatory theories in science, evolutionary theory, is manifestly nonreductionist [in any sense except perhaps (i)]. Even molecular biology sometimes makes use of nonreductionist modes of explanation, including functional explanation (which relies on evolutionary theory for its warrant),[51] and, as will be discussed in detail in Chapter 6, § 6.7, may even need what will be called "topological explanation." These examples point out that what may, at least intuitively, be called "deeper understanding" often requires nonreductive modes of explanation. Finally, no position will be taken here on the question of whether

the unification of science is achievable or even desirable as a goal. Suffice it merely to note that while some generality seems to be required to distinguish an explanation from description in most circumstances, it does not follow that more generality alone guarantees better understanding in all circumstances. In general those who suggest the disunity of science and the need for special sciences will probably find many of the analyses of this book more congenial than their opponents.

Eric R. Scerri

Statement

and

Readings

Abstract for Seven Pines Meeting.

Eric R. Scerri,
UCLA

Can Chemistry be reduced to physics?

The title of my talk implies that such a question can be meaningfully addressed which is something I will discuss. I will first explore the more amenable question of the extent to which chemistry, and especially the periodic system, have in fact been reduced. This will involve an excursion into the wavefunction and density functional approaches in quantum chemistry.

I will then turn to the question in the title and will discuss two recent attempts by other authors, both of whom have answered the question affirmatively.

The two kinds of questions may be loosely classified as epistemological and ontological but this is also a matter of debate. I take it that if one adopts a Quinean approach to the philosophy of science then ontology is obtained by examining the findings of current scientific theories. If this is the case then my own work has been concerned with the ontological question all along.

But if one believes that there is still some scope for philosophical enquiry that does not boil down to what theories tell us then there is room for a more genuinely ontological approach such as that due to Le Poidevin one of the authors I will discuss.

Suggested reading.

Le Poidevin, R. [2005]: 'Missing Elements and Missing Premises, A Combinatorial Argument for the Ontological Reduction of Chemistry', *British Journal for the* Emergence', *Foundations of Chemistry*, **4**, pp. 183-200.

McLaughlin, B. [1992]: 'The Rise and Fall of British Emergentism', In *Emergence or Reduction? Essays on the Prospect of a Non-Reductive Physicalism*, A. Beckerman, H. Flohr, J. Kim (eds.), Berlin, Walter de Gruyter, pp. 49-93.

Scerri, E.R. [1994]: 'Has Chemistry Been at Least Approximately Reduced to Quantum Mechanics?', in *PSA 1994 vol 1*, D. Hull, M. Forbes, and R. Burian, eds., Philosophy of Science Association, East Lansing, MI, pp. 160-170.

Scerri, E.R. [2004]: 'How ab initio is ab initio quantum chemistry', in *Foundations of Chemistry*, **6**, pp. 93-116.

Scerri, E.R. [2006]: Proceedings of PSA 06, contributed papers, to appear.

Scerri, E.R. [2007]: The Periodic Table: Its Story and Its Significance,
Oxford University Press, New York.

(Scerri 1994 and 2006 are posted on the Philosophy of Science Archives site)

ERIC R. SCERRI

JUST HOW AB INITIO IS AB INITIO QUANTUM CHEMISTRY?*

1. INTRODUCTION

Quantum Mechanics has been the most spectacularly successful theory in the history of science. As is often mentioned the accuracy to which the gyromagnetic ratio of the electron can be calculated is a staggering nine decimal places. Quantum Mechanics has revolutionized the study of radiation and matter since its inception just over one hundred years ago. The impact of the theory has been felt in such fields as solid state physics, biochemistry, astrophysics, materials science and electronic engineering, not to mention chemistry, the subject of this conference.

Quantum Mechanics offers the most comprehensive and most successful explanation of many chemical phenomena such as the nature of valency and bonding as well as chemical reactivity. It has also provided a fundamental explanation of the periodic system of the elements that summarizes a vast amount of empirical chemical knowledge. Quantum Mechanics has become increasingly important in the education of chemistry students. The general principles provided by the theory mean that students can now spend less time memorizing chemical facts and more time in actually thinking about chemistry.

I hope that with these opening words I have succeeded in convincing the audience that I do not come before you to deny the power and influence of Quantum Mechanics in the field of chemistry.

* A previous version of this article appeared as 'Löwdin's Remarks on the Aufbau Principle and a Philosopher's View of Ab Initio Quantum Chemistry' in E.J. Brändas, E.S. Kryachko (Eds.) *Fundamental World of Quantum Chemistry*, Vol. II, 675–694, Kluwer, Dordrecht, 2003.

Foundations of Chemistry **6:** 93–116, 2004.

2. THE AIM OF THIS WORK

My project is somewhat different. With the triumph of quantum mechanics there has been an inevitable tendency to exaggerate its success, especially on the part of practicing quantum chemists and physicists. As a philosopher of chemistry I have the luxury of being able to examine the field as an outsider and of asking the kinds of questions which true practitioners might not even contemplate.

Quantum mechanics is part of the reductionist tradition in modern science, and the general claim, often just made implicitly as in any branch of reduction, is that the highest ideal one can aspire to is to derive everything from the theoretical principles. The less experimental data one needs to appeal to, the less one is introducing measured parameters the purer the calculation and the closer it approaches to the ideal of Ockham's razor of being as economical as possible (Hoffmann et al., 1996).

Of course there is no such thing as a completely *ab initio* calculation and if one looks far enough back at the history of any scientific theory one finds that it began with the assumption of at least some experimental data. But it is also fair to say that once the basic principles of a theory have been arrived at, the theorist may 'kick away' the historical-experimental scaffolding. The modern student of quantum mechanics, for example, is not obliged to follow the tortuous route taken by Planck, Einstein, De Broglie, Schrödinger and others. She can go directly to the postulates of quantum mechanics where she will find procedures for doing all kinds of calculations and she can safely ignore the historical heritage of the theory. Indeed many argue, and correctly in my view, that it is actually a hindrance for the practitioner to get too involved in the historical aspects of the theory although it may of course be culturally enriching to do so.

The epitome of the ab initio approach is something like Euclidean geometry where one begins with a number of axioms and one derives everything from this starting point without any recourse whatsoever to empirical data. Needless to say geometry, Euclidean or otherwise, has its origins in the dim distant past when agrarian man needed to think about lines and angles and areas of land. But once the concepts of line, angle and distance had been sufficiently abstracted the agrarian heritage could be completely forgotten.

In a similar way my question today is going to be to ask to what extent the periodic table of the elements can be *explained* strictly from first principles of quantum mechanics without assuming any experimental data whatsoever. I suspect that some physicists and chemists in the audience might well experience some irritation at the almost perverse demands which I will make on what should be derivable from the current theory. If so then I apologize in advance.

By adopting a perspective from the philosophy of science we will cross levels of complexity from the most elementary explanations based on electron shells to frontier ab initio methods. Such a juxtaposition is seldom contemplated in the chemical literature. Textbooks provide elementary explanations that necessarily distort the full details but allow for a more conceptual or qualitative grasp of the main ideas. Meanwhile the research literature focuses on the minute details of particular methods or particular chemical systems and does not typically examine the kind of explanation that is being provided. To give a satisfactory discussion of explanation in the context of the periodic table we need to consider both elementary and supposedly deeper explanations within a common framework.

One of the virtues of philosophy of science is that it can bridge different levels in this way since it primarily seeks the 'big picture' rather than the technical details. In fact supposedly elementary explanations often provide this big picture in a more direct manner but what is also needed is to connect the elementary explanation to the technical details in the deeper theories.

The question of whether or not different levels of explanation for any particular scientific phenomenon are in fact consistent and whether they form a seamless continuum has been the subject of some debate. For example in her first book Nancy Cartwright goes to some lengths to argue that many different explanations can be found for the action of lasers and suggests that these explanations are not necessarily consistent with each other (Cartwright, 1983). In other writings she has expressed some support for the thesis that the various special sciences are dis-unified (Cartwright, 1996).

My own view differs from Cartwright's in that I am of the opinion that the sciences are unified and that explanations given for the same scientific phenomenon at different levels are essentially consistent, although the connection if frequently difficult to elaborate in full

(Scerri, 2000). In this paper I will attempt to draw such connections for the various explanations of the periodic table given at different levels of sophistication.[1]

3. FIRST AN ELEMENTARY APPROACH

Let us start at an elementary level or with a typically 'chemical' view. Suppose we ask an undergraduate chemistry student how quantum mechanics explains the periodic table. If the student has been going to classes and reading her book she will respond that the number of valency or outer-shell electrons determines, broadly speaking, which elements share a common group in the periodic table. The student might possibly also add that the number of outer-shell electrons *causes* elements to behave in a particular manner.

Suppose we get a little more sophisticated about our question. The more advanced student might respond that the periodic table can be explained in terms of the relationship between the quantum numbers which themselves emerge from the solutions to the Schrödinger equation for the hydrogen atom.[2]

This more sophisticated explanation for the periodic system is provided in terms of the relationship between the four quantum numbers that can be assigned to any electron in a many-electron atom. The first quantum number n can adopt any integral value starting with 1. The second quantum number which is given the label ℓ can have any of the following values related to the values of n,

$$\ell = \mathrm{n} - 1, \ldots 0$$

In the case when n = 3 for example, ℓ can take the values 2, 1 or 0. The third quantum number labeled m_ℓ can adopt values related to those of the second quantum numbers by the relationship,

$$\mathrm{m}_\ell = -\ell, -(\ell + 1), \ldots 0 \ldots (\ell - 1), \ell$$

For example if ℓ = 2 the possible values of m_ℓ are,

$$-2, -1, 0, +1, +2$$

Finally, the fourth quantum number labeled m_s can only take two possible values, either +1/2 or −1/2 units of spin angular

momentum. We thus have a hierarchy of related values for the four quantum numbers, which are used to describe any particular electron in an atom. These relationships are derived theoretically and do not involve the use of any experimental data.[3]

For example, if the first quantum number is 3 the second quantum number ℓ can take values of 2, 1 or 0. Each of these values of ℓ will generate a number of possible values of m_ℓ and each of these values will be multiplied by a factor of two since the fourth quantum number can adopt values of 1/2 or $-1/2$. As a result there will be a total of $2 \times (3)^2$ or 18 electrons in the third shell. This scheme thus explains *why* there will be a maximum total of 2, 8, 18, 32 etc. electrons in successive shells as one moves further away from the nucleus.

4. HOW DOES THIS EXPLAIN THE FORM OF THE PERIODIC TABLE?

But does the fact that the third shell can contain 18 electrons also explain why some of the periods in the periodic system contain eighteen places? Actually not exactly. If electron shells were filled in a strictly sequential manner there would be no problem and the explanation would in fact be complete. But as anyone who has studied high school chemistry is aware, the electron shells do not fill in the expected sequential manner. The configuration of element number 18, or argon is,

$$1s^2, 2s^2, 2p^6, 3s^2, 3p^6$$

This might lead one to think that the configuration for the subsequent element, number 19, or potassium, would be

$$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 3d^1$$

since up to this point the pattern has been to add the new electron to the next available orbital in the sequence of orbitals at increasing distances from the nucleus. However experimental evidence shows quite clearly that the configuration of potassium should be denoted as,

$$1s^2, 2s^2, 2p^6, 3s^2, 3p^6, 4s^1$$

As many textbooks state this fact can be explained from the fact that the 4s orbital has a lower energy than the 3d orbital. In the case of element 20 or calcium the new electron also enters the 4s orbital and for the same reason.

5. TRANSITION METAL CONFIGURATIONS

The interesting part is what happens next. In the case of the next element, number 21, or scandium, the orbital energies have reversed so that the 3d orbital has a lower energy, as shown in Figure 1. Textbooks almost invariably claim that since the 4s orbital is already full there is no choice but to begin to occupy the 3d orbital. This pattern is supposed to continue across the first transition series of elements, apart from the elements Cr and Cu where further slight anomalies are believed to occur.

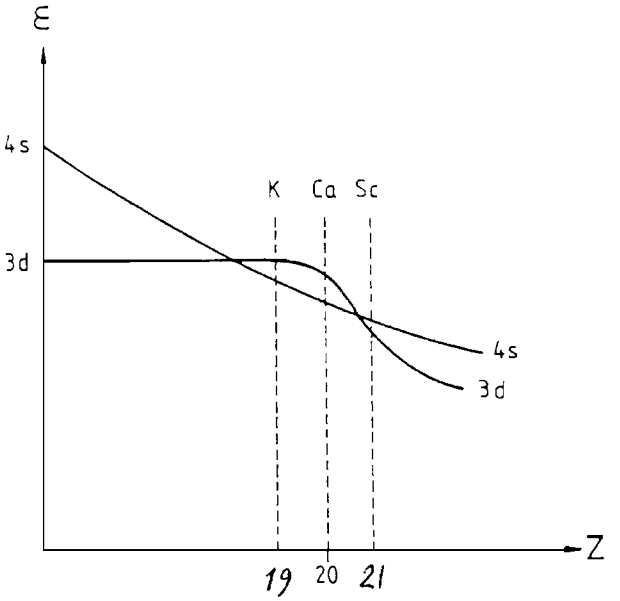


Figure 1. Variation of 4s and 3d orbital energies as a function of Z, atomic number.

In fact this explanation for the configuration of the scandium atom and most other first transition elements is inconsistent. If the 3d orbital has a lower energy than 4s starting at scandium then if

TABLE I

Table of configurations of first transition series

Sc	Ti	V	Cr	Mn	Fe
$4s^2 3d^1$	$4s^2 3d^2$	$4s^2 3d^3$	$4s^1 3d^5$	$4s^2 3d^5$	$4s^2 3d^6$
	Co	Ni	Cu	Zn	
	$4s^2 3d^7$	$4s^2 3d^8$	$4s^2 3d^9$	$4s^1 3d^{10}$	

one were really filling the orbitals in order of increasing energy one would expect that all three of the final electrons would enter 3d orbitals. The argument which most textbooks present is incorrect since it should be possible to predict the configuration of an element from a knowledge of the order of its own orbital energies (Scerri, 1989; Vanquickenborne et al., 1994). It is incorrect to consider the configuration of the previous element and assume that this configuration is carried over intact on moving to the next element, especially in cases where orbital energies cross over each other as they do in this case. It should be possible to predict the order of orbital filling for the scandium atom on its own terms. If one tries to do so, however, one predicts a configuration ending in $3d^3$, contrary to the experimental facts.

The full explanation of why the $4s^2 3d^1$ configuration is adopted in scandium, even though the 3d level has a lower energy, emerges from the peculiarities of the way in which orbital energies are defined in the Hartree–Fock procedure. The details are tedious but have been worked out and I refer anyone who is interested in pursuing this aspect to the literature (Melrose and Scerri, 1996).[4]

6. HOW ARE CONFIGURATIONS DERIVED FROM THE THEORY?

But let me return to the question of whether the periodic table is fully and deductively explained by quantum mechanics. In the usually encountered explanation one assumes that at certain places in the periodic table an unexpected orbital begins to fill as in the case of potassium and calcium where the 4s orbital begins to fill before the 3d shell has been completely filled (Scerri, 1989). This information itself is not derived from first principles. It is justified *post facto* and

TABLE II

Calculated energy levels for two scandium atom configurations

Sc	$4s^2 3d^1$		
	Non-Relativistic	−759.73571776	(atomic units or Hartrees)
	Relativistic	−763.17110138	
	$4s^1 3d^2$		
	Non-Relativistic	−759.66328045	
	Relativistic	−763.09426510	

by some very tricky calculations at that (Melrose and Scerri, 1997; Vanquickenborne et al., 1994).

But if we ignore the conceptual paradox of why 4s fills preferentially even though it has a higher energy than 3d we can just concentrate on calculations aimed at determining the ground state configuration. Suppose we were to use the most widely used method for calculating the energies of atoms and molecules in an ab initio fashion. The Hartree–Fock method[5] can be used to compare the energies of the scandium atom with two alternative configurations,

$$[\mathrm{Ar}]4s^2 3d^1 \quad \text{and} \quad [\mathrm{Ar}]4s^1 3d^2$$

This can be carried out using ordinary non-relativistic quantum mechanics or alternatively by including relativistic effects. The results of using a readily available program on the Internet, created by Froese Fischer[6] one of the leaders in the field of Hartree–Fock calculations, shown in Table II (http://hf5.vuse.vanderbilt.edu/hf.html).[7]

In each case the more negative the calculated value of the energy the more stable the configuration. Clearly the inclusion of relativistic effects serves to reduce the energy from the non-relativistic value. In the case of scandium it appears that both non-relativistic and relativistic *ab initio* calculations correctly compute that the $4s^2$ configuration has the lowest energy in accordance with experimental data. But these calculations, including the ones for subsequent elements must be done on a case-by-case basis. There is not yet a general derivation of the formula which governs the order of filling, sometimes called the $n + \ell$, or Madelung rule, which states that given

TABLE III

Calculated energy levels for two chromium atom configurations

Cr	$4s^1 3d^5$	
	Non-Relativistic	−1043.14175537
	Relativistic	−1049.24406264
	$4s^2 3d^4$	
	Non-Relativistic	−1043.17611655
	Relativistic	−1049.28622286

a choice of filling any two orbitals the order of filling goes according to increasing values of $n + \ell$. For example, 4s where $n + \ell = 4$, fills before 3d where $n + \ell = 5$. But similar calculations do not fare as well in other atoms. Consider the case of the chromium atom for example.

It appears that both non-relativistic and relativistic calculations fail to predict the experimentally observed ground state which is the $4s^1 3d^5$ configuration, as seen in Table III. Of course I do not deny that if one goes far enough in a more elaborate calculation then eventually the correct ground state will be recovered. But in doing so one knows what one is driving at, namely the experimentally observed result. This is not the same as strictly predicting the configuration in the absence of experimental information. In addition, if one goes beyond the Hartree–Fock approximation to something like the configuration interaction approach there is an important sense in which one has gone beyond the picture of a certain number of electrons in a set of orbitals.[8] Rather than just having every electron in every possible orbital in the ground state configuration we now have every electron in every one of thousands or even millions of configurations each of which is expressed in terms of orbitals.

7. COPPER ATOM

Let me consider the case of the copper atom calculated to the same degree of accuracy via the Hartree–Fock method. For this atom the experimentally observed ground state configuration is $4s^1 3d^{10}$.

TABLE IV

Calculated energy levels for two copper atom configurations

Cu	$4s^1 3d^{10}$	
	Non-Relativistic	−1638.96374169
	Relativistic	−1652.66923668
	$4s^2 3d^9$	
	Non-Relativistic	−1638.95008061
	Relativistic	−1652.67104670

From Table IV, we see that sometimes a non-relativistic calculation gives the correct result ($4s^1 3d^{10}$), in terms of which configuration has the lower energy, and yet carrying out the calculation to a greater degree of accuracy by including relativistic effects, gives the wrong prediction. Relativistically one predicts the opposite order of stabilities than what is observed experimentally. Clearly some observed electronic configurations cannot yet be successfully calculated from first principles, at least at this level of approximation. The fact that copper has a $4s^1 3d^{10}$ configuration rather than $4s^2 3d^9$ is an experimental fact. Similarly it is from experimental data that the lengths of the periods are known and not from ab initio calculations.

The development of the period from potassium to krypton is not due to the successive filling of 3s, 3p and 3d electrons but due to the filling of 4s, 3d and 4p. It just so happens that both of these sets of orbitals are filled by a total of 18 electrons.

As a consequence the explanation for the form of the periodic system in terms of how the quantum numbers are related is semi-empirical since the order of orbital filling is obtained from experimental data. Consider now the cumulative total number of electrons which are required for the filling successive shells and periods, respectively,

Closing of shells,
Occurs at Z = 2, 10, 28, 60, 110 (cumulative totals)

Closing of periods,
Occurs at Z = 2, 10, 18, 36, 54, etc.

It is the second sequence of Z values which really embodies the periodic system and not the first. For all we know, electron shells may not even exist or may be replaced by some other concept in a future theory. But the fact that chemical repetitions occur at Z = 3, 11 and 19, if we focus on the alkali metals, for example are chemical facts which will never be superceded.

Only if shells filled sequentially, which they do not, would the theoretical relationship between the quantum numbers provide a purely deductive explanation of the periodic system. The fact that the 4s orbital fills in preference to the 3d orbitals is not predicted in general for the transition metals but only rationalized on a case by case basis as we have seen. In some cases the correct configuration cannot even be rationalized, as in the cases of chromium and copper, at least at this level of approximation. Again, I would like to stress that whether or not more elaborate calculations finally succeed in justifying the experimentally observed ground state does not fundamentally alter the overall situation.[9]

To sum-up, we can to some extent recover the order of filling by calculating the ground state configurations of a sequence of atoms but still nobody has deduced the $n + \ell$ rule from the principles of quantum mechanics. Perhaps this should be a goal for quantum chemists and physicists if they are really to explain the periodic system in terms of electronic configurations of atoms in ab initio fashion.

8. NICKEL ATOM

The case of nickel turns out to be interesting for a different reason. According to nearly every chemistry and physics textbook the configuration of this element is given as

$$4s^2 3d^8$$

However the research literature on atomic calculations (e.g., Bauschlicher et al., 1988) always quotes the configuration of nickel as

$$4s^1 3d^9$$

TABLE V

Quantum mechanical calculations for the nickel atom

Ni	$4s^2 3d^8$	
	Non-Relativistic	−1506.87090774
	Relativistic	−1518.68636410
	$4s^1 3d^9$	
	Non-Relativistic	−1506.82402795
	Relativistic	−1518.62638541

The difference occurs because in more accurate work one considers the average of all the components arising from a particular configuration and not just the lowest possible component of the ground state term. Nickel is somewhat unusual in that although the lowest energy term arises from the $4s^2 3d^8$ configuration it turns out that the average of the energies of all the components arising from this configuration lies higher in energy than the average of all the components arising from the configuration of $4s^1 3d^9$. As a consequence the $4s^2 3d^8$ configuration is regarded as the ground state in research work and it is this average energy which is compared with experimental energies as in Table V. When this comparison is carried out it emerges that the quantum mechanical calculation using either a non-relativistic or a relativistic Hartree–Fock approach gives the wrong ground state.

Of course the calculations can be improved by adding extra terms until this failure is eventually corrected. However, these additional measures are only taken after the facts are known. In addition, the lengths to which theoreticians are forced to go to in order to obtain the correct experimental ordering of terms does not give one too much confidence in the strictly predictive power of quantum mechanical calculations in the context of the periodic table. For example, the very accurate calculations on nickel include the use of basis sets which extend up to 14s, 9p, 5d as well as f orbitals (Raghavachari and Trucks, 1989).[10]

9. CHOICE OF BASIS SET

There is yet another general problem which mars any hope of claiming that electronic configurations can be fully predicted theoretically and that quantum mechanics thus provides a purely deductive explanation of what was previously only obtained from experiments. In most of the configurations we have considered, with the exception of cases mentioned above, it has been possible to use a quantum mechanical method to calculate that this particular configuration does indeed represent the lowest energy possibility. However, in performing such calculations the candidate configurations which are subjected to a variation procedure are themselves obtained from the aufbau principle and other rules of thumb such as Hund's principle or by straightforward appeal to experimental data.

There is a very simple reason for this state of affairs. The quantum mechanical calculations on ground state energies involve the initial selection of a basis set, which in its simplest, or minimal, form is the electronic configuration of the atom in question. Quantum mechanical calculations are not capable of actually generating their own basis sets that must instead be put in 'by hand'. So whereas the correct ground state electronic configurations can in many cases be selected among a number of plausible options, the options themselves are not provided by the theory. I suggest this is another weakness of the present claims to the effect that quantum mechanics explains the periodic system and it is an aspect that might conceivably corrected by future developments.

I will now attempt to take stock of the various senses of the claim that the periodic system is reduced, or fully explained, by quantum mechanics and to extend the scope of this work to more elaborate theoretical approaches.

10. QUALITATIVE EXPLANATION OF PERIODIC TABLE IN TERMS OF ELECTRONS IN SHELLS

The usually given 'explanation' for the period table takes a qualitative form. In broad terms the approximate recurrence of elements after certain regular intervals is explained by the possession of a certain number of outer-shell electrons. This form of explanation

appears to be quantitative to some people because it deals in number of electrons but in fact turns out to be rather qualitative in nature. It cannot be used to predict quantitative data on any particular atom with any degree of accuracy.

Whereas the crude notion of a particular number of electrons in shells or orbitals does not produce very accurate calculations the process can be refined in several ways. The first refinement is perhaps the use of the Hartree method of calculating self-consistent orbitals while at the same time minimizing the energy of the atom.[11] The next refinement lies in making the method consistent with the notion that electrons are indistinguishable. This requirement is met by performing a permutation of all the electrons in the atom so that each electron finds itself simultaneously in all occupied orbitals at once. It is represented mathematically as a determinant that includes all possible permutations within it.

The third refinement is to include any number of excited state configurations for the atom, in a procedure called configuration interaction or the C.I. method. One now has a sum of determinants each of which represents a particular configuration and which is included in the overall atomic wavefunction with a particular weighting determined by a coefficient which is multiplied by the appropriate determinant.

$$\Psi = c_1 D_1 + c_2 D_2 + \ldots$$

The calculation consists in finding the optimum weighting which all the determinants must have in order to minimize the energy of the atom. Having reached this level of abstraction we have really left behind the homely picture of electrons in particular shells. If one still insists on visualization, each electron is now in every orbital of every single configuration that we choose to consider.

Clearly there is still a connection with the elementary homely model but it is also fair to say that the move towards greater abstraction has somewhat invalidated the naïve model. This now raises the question as to whether the elementary model really does have explanatory power. I would argue that it does not. It may have led historically to these more sophisticated approaches but it has been rendered vastly more abstract in the process.

But if we are considering the general question of explanation it is not essential to retain the homely picture that can be grasped by the

general chemist or the beginning student of physical chemistry. We must move on to enquire about how the more abstract approaches actually fare. The short answer is much better but still not in strictly ab initio fashion.

11. NECESSARY AND SUFFICIENT CONDITIONS

But in any case even within the elementary model it emerges that the possession of a particular number of outer-shell electrons is neither a necessary nor a sufficient condition for an element's being in any particular group. It is possible for two elements to possess exactly the same outer electronic configuration and yet not to be in the same group of the periodic system. For example, the inert gas helium has two outer-shell electrons and yet is not usually placed among the alkaline earth elements such as magnesium, calcium or barium, all of which also display two outer-shell electrons.[12] The possession of a particular number of outer-shell electrons is therefore not sufficient grounds for placing it in a particular group.

Conversely, there are cases of elements that do belong in the same group of the periodic table even though they do not have the same outer-shell configuration. In fact this occurrence is rather common in the transition metal series. To take one interesting example,[13] consider the nickel group in which no two elements show the same outer shell configuration!

Ni	[Ar] $4s^2 3d^8$
Pd	[Kr] $4s^0 3d^{10}$
Pt	[Xe] $4s^1 3d^9$

In addition the very notion of a particular number of electrons in a particular shell stands in strict violation of the Pauli Principle, arguably one of the most powerful principles in the whole of science. This states that electrons cannot be distinguished, which implies that we can never really state that a particular number belong in one shell and another number in a different shell, although there is no denying the usefulness of making this approximation. The independent-electron approximation, as it is known, represents one of the central

paradigms in modern chemistry and physics and of course I am not denying its usefulness but am focusing on its ontological status.

But all this talk of electrons in shells and orbitals is just naïve realism. The lesson from quantum mechanics is the need to abandon naïve realism, to abandon picturing waves or particles or picturing spinning electrons.[14] The standard, or Copenhagen, interpretation of quantum mechanics urges us to just do the mathematics and adopt an instrumental approach to the theory. Of course this is hard especially for chemists since most of their work consists in shapes, structures, diagrams, pictures, representations and observable changes. Let us finally consider explanations of the periodic table that do not involve picturing electrons in shells or orbitals.[15]

12. AB INITIO CALCULATIONS BASED ON WAVEFUNCTIONS

Some of the more abstract ab initio approaches have already been described above. They are the Hartree–Fock method and the configuration interaction approach.

Indeed, such approaches fare much better, and are serious contenders for the claim to a full explanation of the periodic system. In order to illustrate both the power and the pitfalls of the methods I will focus for simplicity on the ab initio calculation of ionization energies of atoms. In this approach the notion of electrons in shells is used instrumentally with the knowledge that such an approximation only represents a first order approach to calculations. If one is doing a Hartree–Fock calculation then all the electrons are simultaneously in all the orbitals of a particular chosen configuration. As mentioned earlier this results from the permutation procedure. If one is doing C.I. then many thousands if not millions of configurations are considered in the wavefunction expansion.

Within these ab initio approaches the fact that certain elements fall into the same group of the periodic table is not explained by recourse to the number of outer-shell electrons. The explanation lies in calculating the magnitude of a property such as the first ionization energy and seeing whether the expected periodicity is recovered in the calculations. Figure 2 below shows schematically the experimental ionization energies for the first 53 elements in the periodic table, along with the values calculated using ab

initio quantum mechanical methods. As can readily be seen, the periodicity is captured remarkably well, even down to the details of the sections of the graph occurring between elements in groups II and III in each period of the table. Clearly the accurate calculation of atomic properties can be achieved by the theory. The quantum mechanical explanation of the periodic system within this approach represents a far more impressive achievement than merely claiming that elements fall into similar groups because they share the same number of outer-electrons.

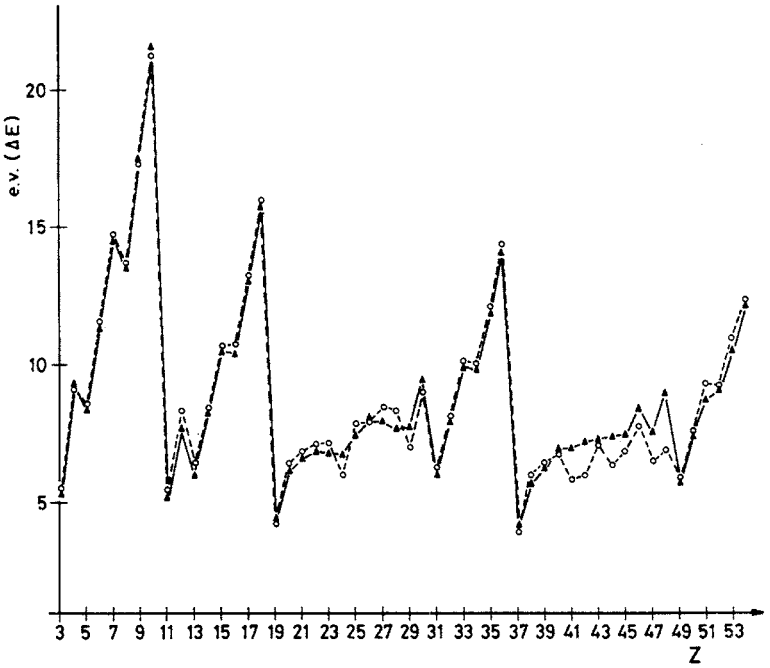


Figure 2. Comparison of computed and experimental first ionization energies for Z = 1–53.

And yet in spite of these remarkable successes such an ab initio approach may still be considered to be semi-empirical in a rather specific sense. In order to obtain calculated points shown in Figure 2 the Schrödinger equation must be solved separately for each of the 53 atoms concerned in this study. The approach therefore represents a form of 'empirical mathematics' where one calculates 53 individual Schrödinger equations in order to reproduce the well-known

pattern in the periodicities of ionization energies. It is as if one had performed 53 individual experiments, although the 'experiments' in this case are all iterative mathematical computations. This is still therefore not a general solution to the problem of the electronic structure of atoms.

13. DENSITY FUNCTIONAL APPROACH

In 1926 the physicist Llewellyn Thomas proposed treating the electrons in an atom by analogy to a statistical gas of particles. No electron-shells are envisaged in this model which was independently rediscovered by Italian physicist Enrico Fermi two years later, and is now called the Thomas–Fermi method.[16] For many years it was regarded as a mathematical curiosity without much hope of application since the results it yielded were inferior to those obtained by the method based on electron orbitals. The Thomas–Fermi method treats the electrons around the nucleus as a perfectly homogeneous electron gas. The mathematical solution for the Thomas–Fermi model is 'universal', which means that it can be solved once and for all. This should represent an improvement over the method that seeks to solve Schrödinger equation for every atom separately. Gradually the Thomas–Fermi method, or density functional theories, as its modern descendants are known, have become as powerful as methods based on orbitals and wavefunctions and in many cases can outstrip the wavefunction approaches in terms of computational accuracy.

There is another important conceptual, or even philosophical, difference between the orbital/wavefunction methods and the density functional methods. In the former case the theoretical entities are completely unobservable whereas electron density invoked by density functional theories is a genuine observable. Experiments to observe electron densities have been routinely conducted since the development of X-ray and other diffraction techniques (Coppens, 2001).[17] Orbitals cannot be observed either directly, indirectly or in any other way since they have no physical reality, a state of affairs that is dictated by quantum mechanics (Scerri, 2000). Orbitals as used in ab initio calculations are mathematical figments that exist, if anything, in a multi-dimensional Hilbert space.[18] Electron density

is altogether different, as I have indicated, since it is a well-defined observable and exists in real three-dimensional space a feature that some theorists point to as a virtue of density functional methods.[19]

14. DENSITY FUNCTIONAL THEORY IN PRACTICE

Most of what has been described so far concerning density theory applies in theory rather than in practice. The fact that the Thomas–Fermi method is capable of yielding a universal solution for all atoms in the periodic table is a potentially attractive feature but is generally not realized in practice. Because of various technical difficulties, the attempts to implement the ideas originally due to Thomas and Fermi have not quite materialized. This has meant a return to the need to solve a number of equations separately for each individual atom as one does in the Hartree–Fock method and other ab initio methods using atomic orbitals. In addition most of the more tractable approaches in density functional theory also involve a return to the use of atomic orbitals in carrying out quantum mechanical calculations since there is no known means of directly obtaining the functional that captures electron density exactly.[20] Researchers therefore fall back on using basis sets of atomic orbitals which means that conceptually we are back to square one and that the promise of density functional methods to work with observable electron density has not materialized.

To make matters worse, the use of a uniform gas model for electron density does not enable one to carry out accurate calculations. Instead, 'ripples' or a density gradient, to use the more technical term, must be introduced into the uniform electron gas distribution. The way in which this has been implemented has typically been in a semi-empirical manner by working backwards from the known results on a particular atom, usually the helium atom (Gill, 1998). In this way it has been possible to obtain an approximate set of functions which often give successful approximate calculations in many other atoms and molecules. There is no known way of yet calculating, in an ab initio manner, the required degree of density gradient that must be introduced into the calculations.

By carrying out this combination of semi-empirical procedures and retreating from the pure Thomas–Fermi notion of a uniform

electron gas it has actually been possible to obtain computationally *better* results in many cases of interest than with conventional ab initio methods. True enough, calculations have become increasingly accurate but if one examines them more closely one realizes that they include considerable semi-empirical elements at various levels. From the purist philosophical point of view this means that not everything is being explained from first principles.

As time has progressed the best of both approaches (DFT and ab initio orbital methods) have been blended together with the result that many computations are now performed by a careful mixture of wavefunction and density approaches within the same computations (Hehre, 1986). This feature brings with it advantages as well as disadvantages. The unfortunate fact is that, as yet, there is really no such thing as a pure density functional method for performing calculations and so the philosophical appeal of a universal solution for all the atoms based on electron density rather than ficticious orbitals has not yet borne fruit.[21]

15. CONCLUSION

My aim has not been one of trying to decide whether or not the periodic system is explained *tout court* by quantum mechanics. Of course broadly speaking quantum mechanics does provide an excellent explanation and certainly one better than was available using only classical mechanics. But the situation is more subtle.

Whereas most chemists and educators seem to believe that all is well, I think that there is some benefit in pursuing the question of how much is strictly explained from the theory. After all, it is hardly surprising that quantum mechanics cannot yet fully *deduce* the details of the periodic table that gathers together a host of empirical data from a level far removed from the microscopic world of quantum mechanics. As Roald Hoffmann's title at this memorial meeting stated, "Most of what's interesting in chemistry is not reducible to physics" It is indeed something of a miracle that quantum mechanics explains the periodic table to the extent that it does at present. But we should not let this fact seduce us into believing that it is a complete explanation. One thing that is clear is that the attempt to explain the details of the periodic table

continues to challenge the ingenuity of quantum physicists and quantum chemists. For example, a number of physicists are trying to explain the periodic table by recourse to group theoretical symmetries in combination with quantum mechanics (Ostrovsky, 2000). Meanwhile the theoretical chemist Herschbach and colleagues have worked on a number of approaches which also aim at obtaining a global solution to the energies of the atoms in the periodic table (Kais et al., 1994)

Perhaps philosophers of chemistry have a role to play here. Unconstrained by what can presently be achieved, or even what might be achieved in the foreseeable future, one can point out the limitations of the current state of the art and one can place the research in the wider context of scientific reductionism in general and what it might mean for a calculation to be really ab initio. This is not a denial of the progress achieved in quantum chemistry or a reproach of the current work. It is more of an unrestrained look at what more could conceivably be done. Of course this might require a deeper theory than quantum mechanics or maybe a cleverer use of the existing theory. There is really no way of telling in advance.

ACKNOWLEDGEMENT

I would like to thank John Bloor for his highly incisive comments on many aspects of the work discussed in this article. Nevertheless I am sure he will not agree with all that I write here. I also thank Roald Hoffmann and other participants at the Rosenfeld memorial meeting for making some interesting comments from which I have greatly benefited.

NOTES

[1] Another way of regarding the same question is to consider typical 'chemical explanations', full of visualizations and sometimes naïve realism, and contrast them with the more abstract mathematical explanations favored by the physicist.

[2] In fact the fourth quantum number does not emerge from solving Schrodinger's equation. It was initially introduced for experimental resons by Pauli, as a fourth degree of freedom possessed by each electron. In the later treatment by Dirac the fourth quantum number emerges in a natural manner.

[3] The fourth quantum number does not emerge from solving the Schrödinger equation.

[4] It is gratifying to see that this article has now been cited by about twelve chemistry textbooks including those by Atkins, Huheey, Levine etc.

[5] It should be noted that the Hartree–Fock method uses four quantum numbers which are given the same labels as those in the hydrogen atom. However these are not identical but only analogous. This fact is often overlooked in elementary presentations which imply that the two sets are identical.

In a recent paper Ostrovsky has criticized my claiming that electrons cannot strictly have quantum numbers assigned to them in a many-electron system (Ostrovsky, 2001). His point is that the Hartree–Fock procedure assigns all the quantum numbers to all the electrons because of the permutation procedure. However this procedure still fails to overcome the basic fact that quantum numbers for individual electrons such as l in a many-electron system fail to commute with the Hamiltonian of the system. As a result the assignment is approximate. In reality only the atom as a whole has quantum numbers, not individual electrons.

[6] Charlotte Froese Fischer was a PhD student of Hartree's in Cambridge and pioneered accurate calculations using the method initially devised by Hartree.

[7] Admittedly Hartree–Fock calculations whether relativistic or not omit correlation effects in atoms since they involve time averages of electron repulsions.

[8] Broadly speaking it is still an orbital based method of course but not one that corresponds to the elementary concept of a particular number of electrons in the shells of an atom.

[9] In fact given that the C.I. approach involves a mixture of so many different configurations it is capable of calculating the energy of the entire atom but not specifically of the ground state configuration.

[10] The CISD method produces typical errors of 0.4–0.7 eV for the ground states of elements from manganese to copper even after the inclusion of relativistic effects. The Coupled Cluster method called CPF produces an error of 0.4 eV for the d^8s^2 to d^9s^1 splitting in nickel. The basis set cited in the main text comes from a study in which an elaborate quadratic CI method was used in which the already large basis set was augmented with numerous 'diffuse' orbitals (Raghavachari and Trucks, 1989). The use of M-P perturbation theory produced what the authors of this article describe as "wild oscillations" for the same excitation energy.

[11] I am doing a certain amount of back-tracking given that this method was mentioned above when some results were quoted for transition metals.

[12] In fact there are some other good reasons to support the placement of helium in the alkaline earths, contrary to popular opinion among chemists as I will be exploring in a forthcoming article.

[13] Although as noted the configuration of Ni is actually $4s^13d^9$ contrary to what is stated in most textbooks.

[14] The question for realism is altogether different if taken in the sense of the belief in unobservable scientific entities. In fact many philosophers of science currently favor some form of scientific realism in the context of quantum mechanics (Cao, 2003).

[15] So I advocate realism about chemical reactions that can be observed macroscopically without being a realist about electrons in shells.

[16] But Teller showed that the Thomas–Fermi model cannot predict binding in atoms.

[17] This is why I and some others have been agitating about the recent reports, starting in Nature magazine in September 1999, that atomic orbitals had been directly observed. This is simply impossible (Scerri, 2000).

[18] I have tried to stress the educational implications of the claims for the observation of orbitals in other articles and will not dwell on the issue here (Scerri, 2000, both articles cited for that year).

[19] Of course it is a matter of taste whether one uses ficticious orbitals or real and observable electron density.

[20] Promise due to theorems proved by Hohenberg and Sham and Kohn.

[21] Some preliminary work aimed at developing pure density methods has been carried out (Wang and Carter, 2000).

REFERENCES

C.W. Bauschlicher, P. Siegbahn and G.M. Petterson. The Atomic States of Nickel. *Theoretica Chimica Acta* 74: 479–491, 1988.

T.Y. Cao (Ed.). Structural Realism and Quantum Field Theory, Special Issue, *Synthese* 136, 2003.

N.C. Cartwright. *How the Laws of Physics Lie*. Oxford: Clarendon Press, 1983.

P. Galison and D. Stump (Eds.). *The Disunity of Science*. Stanford, California: Stanford University Press, 1996.

P.M.W. Gill. In v.R. Schleyer (Ed.), *Encyclopedia of Computational Chemistry*, Vol. 1, pp. 678–689. Chichester: Wiley, 1998.

W.J. Hehre, L. Radom, P.v.R. Schleyer and J. Pople. *Ab Initio Molecular Orbital Theory*. New York: John Wiley, 1986.

R. Hoffmann, B. Minkin, and B. Carpenter. Ocham's Razor and Chemistry. *Bulletin de la Societée Chimique Française* 133: 117–130, 1996.

S. Kais, S.M. Sung and D.R. Herschbach. Large-Z and -N Dependence of Atomic Energies of the Large-Dimension Limit. *International Journal of Quantum Chemistry* 49: 657–674, 1994.

M.P. Melrose and E.R. Scerri. The Authors Reply to "Why the 4s Orbital Is Occupied before the 3d". *J. Chem. Educ.* 74: 616–616, 1997.

C.A. Moore. Tables of Atomic Energies . . . .

V.N. Ostrovsky. How and What Physics Contributes to Understanding the Periodic Law. *Foundations of Chemistry* 3: 145–182, 2001.

K. Raghavachari and G.W. Trucks. Highly Correlated Systems. Ionization Energies of First Row Transition Metals Sc-Zn, *Journal of Chemical Physics* 91: 2457–2460, 1989.

E.R. Scerri. Transition Metal Configurations and Limitations of the Orbital Approximation. *J. Chem. Educ.* 66: 481–483, 1989.

E.R. Scerri. How Good Is the Quantum Mechanical Explanation of the Periodic System? *Journal of Chemical Education* 75: 1384–1385, 1998.

E.R. Scerri. The Failure of Reduction and How to Resist the Disunity of Science in Chemical Education, *Science and Education* 9: 405-425, 2000.

E.R. Scerri. Have Orbitals Really Been Observed? *Journal of Chemical Education* 77: 1492–1494, 2000.

L.G. Vanquickenborne, K. Pierloot and D. Devoghel. Electronic Configurations and Orbital Energies. *Inorganic Chemistry* 28, 1805–1813, 1989.

L.G. Vanquickenborne, K. Pierloot and D. Devoghel. Transition Metals and the Aufbau Principle. *Journal of Chemical Education* 71: 469–471, 1994.

A. Wang and E.A. Carter. In S.D. Schwartz (Ed.), *Theoretical Methods in Condensed Phase Chemistry*, pp. 117–184. Dordrecht: Kluwer, 2000.

Department of Chemistry and Biochemistry
UCLA
Los Angeles, CA 90095-1569
USA

CORRIGENDUM

E.R. Scerri. Just How Ab Initio is Ab Initio Quantum Chemistry? *Foundations of Chemistry* 6: 93–116, 2004.

p. 99. The configurations for Cu and Zn shown in Table I are incorrect. They should read:

Cu $4s^1 3d^{10}$ Zn $4s^2 3d^{10}$

p. 107. The configurations for Pd and Pt are incorrect and should read:

Pd [Kr] $5s^0 4d^{10}$

Pt [Xe] $6s^1 5d^9$

Reduction and Emergence in Chemistry - Two Recent Approches*

Eric Scerri
Department of Chemistry & Biochemistry,
UCLA,
Los Angeles, CA 90095
scerri@chem.ucla.edu

Abstract

Two articles on the reduction of chemistry are examined. The first, by McLaughlin, claims that chemistry is reduced to physics and that there is no evidence for emergence or for downward causation between the chemical and the physical level. In a more recent article Le Poidevin maintains that his combinatorial approach provides grounding for the ontological reduction of chemistry, which also circumvents some limitations in the physicalist program.

1. Introduction.

In recent years there the reduction of chemistry has been discussed in a variety of ways. Many studies have concentrated on inter-theoretical reduction between theories of chemistry and theories of physics (Bunge 1982). Others have discussed the reduction of chemistry in a naturalistic manner, by examining the question of how some typically molecular properties can be deduced from quantum mechanics in an ab initio fashion or whether the periodic system can be deduced from quantum mechanics (Scerri 1994, 2004). More recently a number of authors have turned to discussing the ontological reduction of chemistry (McLaughlin 1992; Le Poidevin 2005). The present article examines the claims regarding emergence and the ontological reduction of chemistry in the last two cited articles.

2. McLaughlin on British Emergentism and the relationship of chemistry to physics.

Brian McLaughlin has written a frequently cited paper in which he seeks to give an overview of the philosophical school that he dubs 'British Emergentism' which includes the work of J.S. Mill, Bain, Morgan and most recently C.D. Broad. I begin with a brief summary of McLaughlin's characterization of these philosophers, especially of C.D. Broad.

Emergentists held, rather uncontroversially, that the natural kinds at each scientific level are wholly composed of kinds of lower levels, and ultimately of kinds of elementary particles. However, they also maintained that,

> Some special science kinds from each special science can be wholly composed of the types of structures of material particles that endow the kinds in question with fundamental causal powers (McLaughlin 1992, 50-51).

These powers were said to 'emerge' from the types of structures in question. One example given repeatedly by the British emergentists was that of chemical elements which have the power to bond to other elements by virtue of their internal microscopic structures. According to the emergentists, when these causal powers operate they bring about the movement of particles. The striking part, as McLaughlin calls it, about the emergentist claim, is that the kinds pertaining to a special science, such as chemistry, are said to have the power to influence microscopic motions of particles in ways that are not anticipated by the laws governing the microscopic particles. Emergentism is thus committed to the possibility of 'downward causation'.

For example, emergentists such as Broad believed that chemical bonding represents an example of emergence and the operation of downward causation. Indeed he went as far as to declare,

> The situation with which we are faced in chemistry…seems to offer the most plausible example of emergent behaviour (Broad 1925, 65).

Broad believed that emergent and mechanistic chemistry (non-emergent chemistry) agree in the following respect,

> That all the different chemical elements are composed of positive and negative electrified particles in different numbers and arrangements; and that these differences of number and arrangement are the only ultimate difference between them (Broad 1925, 69).

However, he also stressed that if mechanistic chemistry were true it should be possible to deduce the chemical behavior of any element from the number and arrangement of such particles, without needing to observe a sample of the element in question, which is something that is clearly not the case.

Against this position McLaughlin maintains that the coming of quantum mechanics and the quantum mechanical theory of bonding has rendered these emergentist claims untenable. In fact he is very categorical about the prospects for modern day emergentism.

> It is, I contend, no coincidence that the last major work in the British Emergentist tradition coincided with the advent of quantum mechanics. Quantum mechanics and the various scientific advances made possible are arguably what led to British Emergentism's downfall…quantum mechanical explanations of chemical bonding in terms of electromagneticism [sic], and various advances this made possible in molecular biology and genetics – for example the discovery of the structure of DNA – make the main doctrines of British emergentism, so far as the chemical and the biological are concerned at least, seem enormously implausible. Given the advent of quantum mechanics and these other scientific theories, there seems not a scintilla of evidence that there are emergent causal powers or laws in the sense in question… and there seems not a scintilla of evidence that there is downward causation from the psychological, biological and chemical levels (McLaughlin 1992, 54-55).

These anti-emergentist claims can be criticized on several different fronts. Granted that the quantum mechanical theory of bonding that McLaughlin appeals to does, provide a more fundamental account of chemical bonding than the classical, or Lewis theory. Nevertheless, it does not permit one to predict in advance the behavior of elements or the properties that a compound might have once any two or more elements have combined together. Moreover, it is not as though there was a complete absence of any theoretical understanding of chemical bonding before the quantum theory was introduced. Lewis's theory, whereby covalent bonds occur when elements share pairs of electrons, gave a good account of the bonding in most compounds. Lewis arrived at his theory through the crucial realization that most stable molecules have an even number of electrons, while unstable ones such as nitrogen monoxide (NO) possess an odd number of electrons. Lewis thus naturally assumed that bonding to form stable molecules involved the pairing of electrons in bonds or as lone pairs.

Admittedly the quantum mechanical theory, devised by Heitler, London, Pauling, Millikan and others goes beyond this 'homely picture' of pairs of electrons, mysteriously holding atoms together. However, Lewis' concept of bonds as pairs of electrons is not thereby refuted but rather given a deeper physical mechanism. According to the quantum mechanical account electrons are regarded as occupying bonding and anti-bonding orbitals. To a first approximation, if the number of bonding electrons exceeds the number of anti-bonding electrons the molecule is predicted to be a stable one.[1] Moreover, the electrons occupy these orbitals, two by two, in pairs. The deeper understanding lies in the fact that the electrons are regarded as spinning in opposite directions within all such pairs. Indeed it is the exchange energy associated with electron spin which accounts quantitatively for the bonding in any compound and it is in this last respect that the quantum mechanical theory goes beyond Lewis's theory.

Linus Pauling, one of the chief architects of the quantum mechanical account of chemical bonding was quick to point out the continuity with Lewis' concept when he wrote,[2]

> It may be pointed out that this theory is in simple cases entirely equivalent to G.N. Lewis's successful theory of the shared electron pair, advanced in 1916 on the basis of purely chemical evidence. Lewis's electron pair consists now of two electrons which are in identical states except that their spins are opposed (Pauling 1928, 359).

There is another aspect of McLaughlin's above cited passage that is entirely incorrect, namely his claim that the discovery of the structure of DNA owes something to the quantum mechanical theory of bonding. As a matter of fact there is no connection whatsoever between these two developments. All I can think of to explain McLaughlin's statement is that Pauling was involved in both developments.[3] But of course Pauling rather famously failed to find the structure of DNA and was beaten to it by Crick and Watson.

The discovery of the structure of DNA was driven almost entirely by the X-ray diffraction evidence that became available to Crick and Watson, courtesy of Wilkins and Franklin. It did not rest on any quantum mechanical calculations or indeed any insights

provided by the theory. It involved model building and cardboard-cut outs of bases.

McLaughlin does not say anything whatsoever about pre-quantum mechanical theories of bonding, except to imply that they were completely inadequate. At the same time he suggests that the quantum mechanical theory has provided a complete answer to the question of bonding. Neither of these extreme positions are correct.

It is not clear whether it is the superior quantitative nature of the quantum mechanical theory that McLaughlin is so impressed by, since he does not say. The only argument offered is that the quantum mechanical theory led directly to the elucidation of the structure of DNA and so on. If one puts aside these false arguments as I am urging, it raises the question of why McLaughlin believes that quantum mechanics was so overwhelmingly successful in chemistry, to the extent of rendering emergentism about bonding *completely* untenable. McLaughlin offers us no such argument for the superiority of the quantum mechanical account of bonding over the earlier classical theory of Lewis. McLaughlin implies that the quantum mechanical theory provides what the classical theory could not, namely the power to predict how two elements might react together. Or is McLaughlin suggesting that using quantum mechanics we can predict the properties of an element from a knowledge of the number of fundamental particles that its atoms possess?

Unfortunately, as anyone who is aware of the current state of quantum chemistry knows well, neither of these feats are possible. In the case of elements we can predict particular properties perhaps such as ionization energies but not chemical behavior. In the case of compounds what can be achieved is an accurate estimate, and in many cases even predictions, regarding specific properties in compounds that are known to have formed between the elements in question. Quantum mechanics cannot yet predict what compounds will actually form. Broad's complaint about the inability of mechanistic or classical chemistry to predict the properties of elements, or the outcome of chemical reactions between any two given elements, remains unanswered to this day. Why then should we accept McLaughlin's claim that pioneer quantum chemistry, or even today's version of the theory of bonding, can so decisively deal a death-blow to any notions of emergence and downward causation?

In any case, as McLaughlin himself seems to concede, the advent of a quantum mechanical theory of bonding did not in fact kill off emergentism completely since some prominent biologists and neurophysiologists such as Roger Sperry, whom he cites, continued to work in this tradition. Moreover, if one surveys the literature one cannot fail to be struck by the 're-emergence of emergence', as it has aptly been termed (Cunningham 2001). This is equally true of the humanities as it is of the physical sciences. For example, the prominent Harvard chemist George Whitesides has been showing increasing support for claims for the emergence of chemical phenomena from physical ones, precisely the example of emergence which McLaughlin wishes to deny so strenuously (Whitesides, Ismagilov 1999). Rather than being 'killed off' by the quantum mechanical account of chemical bonding, emergence is alive and well. McLaughlin's attempt to assert the reduction of chemistry by appealing to the non-existence of emergence of the chemical from the physical, and his associated denial of downward causation are thus entirely unconvincing at least to the present author.

Finally, as Kim has pointed out in another context, the notion of emergence is a perfectly respectable one that bears some striking similarities to the currently popular

notion of non-reductive physicalism that prevails in the philosophy of mind.[4] I do not believe that a straightforward appeal to the quantum mechanical account of chemical bonding can be taken as signaling the demise of emergence of chemistry from physics.

3. Another approach to the Reduction of Chemistry - Le Poidevin.

The second article under consideration also raises the question of the ontology of chemistry. To what extent can we avail ourselves of knowledge obtained through theories such as quantum mechanics? Robin Le Poidevin, contrary to McLaughlin's approach, believes that we need to separate ontology from epistemology rather sharply. He claims to have given an argument in favor of the ontological reduction of chemistry, which does not appeal to the fortunes of any particular physical or chemical theory. He also hopes to bypass the kinds of problems that beset a physicalist approach to ontological reduction. As he explains, these problems apply to the reduction of the mental, as much as they do to the reduction of the biological or chemical levels to fundamental physics.

Le Poidevin makes special mention of the periodic system and of Mendeleev's prediction of new elements. He sets out to discover why Mendeleev was so confident that the elements he predicted actually existed. Le Poidevin claims that this is not a question about Mendeleev's confidence in the periodic law but rather about an implicit conceptual move. If one grants that the gaps in the periodic table represented genuine possibilities, elements that could exist, why did Mendeleev assume that the possibilities would actually be realized?

Le Poidevin then draws the following distinction.

> Even if some elements in the table are merely possible, there is a genuine difference between the physical possibility of an element between, say, zinc and arsenic (atomic numbers 30 and 33), and the mere logical possibility of an element between potassium and calcium (19 and 20) (Le Poidevin 2005, 119).

I refer to this passage because the discreteness in the existence of elements goes on to play a pivotal role in Le Poidevin's eventual argument in favor of the ontological reduction of chemistry. Le Poidevin agrees with those who in recent years have claimed that chemistry is not reduced to physics in an epistemological sense but, to repeat, his real goal is to examine the ontological question without appeal to theories.

> There is, I think, a strong intuition that ontological reduction is true, whatever the fortunes of epistemological reduction. But what is the source of this intuition? Can ontological reduction be defended independently of epistemological reduction? (Le Poidevin 2005, 120-121).

Le Poidevin's answer to the last question is that it can. In addition he is well aware that the frequent appeal to physicalism that is made, especially in the philosophy of mind, is plagued by some rather serious problems. The author reminds us that the claim that chemical properties supervene on those properties described by the complete science is just as trivial as the thesis that mental properties do. Secondly he brings up the so-

called ‘symmetry problem’. Even if we suppose a one-to-one correspondence between a given chemical property and one described by physics, that correspondence would not by itself suggest that one is more fundamental than the other.

Le Poidevin considers the relationship between valence and electronic configuration in an effort to cast further light on these issues,

> Suppose, for example, valency to supervene on electronic configuration. At first sight, the relation appears to be asymmetric because of a valency of 1, for example, can be realized by a number of distinct configurations, but nothing can differ in terms of valency without also differing in terms of electronic configuration. However, the relevant part of the configuration--the part that determines valency--will not vary among elements of the same valency. The determination therefore goes both ways (Le Poidevin 2005, 123-124).

Is Le Poidevin correct in his assertion that " nothing can differ in terms of valency without also differing in terms of electronic configuration"? In fact this is not the case since, as is well known, most non-metal elements can show variable valences in spite of possessing a single electronic configuration. Sulfur, to take just one example, has the electronic configuration of $1s^2$, $2s^2$, $2p^6$, $3s^2$, $3p^4$. Nevertheless, it commonly shows valences of +2, +4 or +6 such as in the compounds SCl_2, SO_2 and SO_3 respectively.

But Le Poidevin is nevertheless still correct in pointing out that in general the symmetry problem is a pressing one. The grounding of reduction requires something more than the physicalist prejudice, or the hope, that physical levels determine chemical levels and not vice versa.

Le Poidevin proposes to circumvent both this problem and the problem of vacuity, mentioned above, by an approach that he terms combinatorialism.

> The central contention of combinatorialism is this: possibilities are just combinations of actually existing simple items (individuals, properties, relations). Let us call this the *principle of recombination*. To illustrate it, suppose the actual world to contain just two individuals, a and b, and two monadic properties, F and G, such that (Fa & Gb). Assuming F and G to be incompatible properties, and ignoring the possibility of there being nothing at all, then the following is an exhaustive list of the other possibilities:
>
> 1. Fa
> 2. Fb
> 3. Ga
> 4. Gb
> 5. Fa & Fb
> 6. Ga & Gb
> 7. Ga & Fb
>
> (Le Poidevin 2005, 124).

Le Poidevin explains that combinatorialism is a form of reductionism about *possibilia*. He claims that the talk of non-existent *possibilia* is made true by virtue of

actual objects and their properties, just as the inhabitants of his model world is made possible by virtue of a and b and the properties F and G. The idea is that we should consider Mendeleev's predicted elements in this way. According to Le Poidevin's approach, the elements that are as yet non-existent but physically possible are those that can be regarded as combinations of some undefined basic objects and/or basic properties.

Le Poidevin suggests that this approach provides a means of establishing the required asymmetry in order to ground the reduction of the chemical to the physical or the mental to the physical, and a means of countering the symmetry problem alluded to earlier.

> A property-type F is ontologically reducible to a more fundamental property-type G is the possibility of something's being F is constituted by a recombination of actual instances of G, but the possibility of something's being G is not constituted by a recombination of actual instances of F (Le Poidevin 2005, 129).

I come now to the crucial argument in Le Poidevin's paper,

> But since the thesis of ontological reduction is about properties, we do have to have a clear conception of what is to count as a chemical property. I shall take the identity of an element, as defined by its position in a periodic ordering, and its associated macroscopic properties (capacity to form compounds of a given composition with other elements, solubility etc.) to be paradigmatically chemical properties…The question of the ontological reduction of chemistry (or at least the question I am interested in) is the question of whether these paradigmatically chemical properties reduce to more fundamental properties (Le Poidevin 2005, 131).

Let me say something about the second sentence since I think this will turn out to be Le Poidevin's undoing. In his brief list of what he terms paradigmatically chemical properties the author has lumped together (a) the identity of elements, (b) their capacity to form compounds of a certain composition and (c) their solubilities. But there is a long-standing philosophical view whereby elements should be regarded as having a dual nature consisting of basic substances and of simple substances (Paneth 1962). If one takes this dual view seriously it casts doubt on Le Poidevin's lumping together of the existence of elements and their properties such as solubilities.

As Mendeleev, and more recently Paneth among others have stressed, the notion of an element as a basic substance concerns just its identity and its ability to act as the bearer of properties. A basic substance does not however possess any properties.[5] The 'properties' of an element however reside in the simple substance and not in the element as a basic substance. According to this view, the identity of an element and its properties are regarded as being quite separate. If we consider le Poidevin's three examples, namely identity, capacity to form compounds and solubility we see a conflation of basic substance aspects (identity) with simple substance aspects (solubility). It is only by failing to distinguish between the identity of elements and their possessing properties, such as solubility, that Le Poidevin is able to give the impression that he has provided an argument for the ontological reduction of chemistry as a whole.

He then adds,

> We might, just accept it as a brute fact about the world that the series of elements was discrete. But if there were a finite number of properties, combinations of which generate the physical possibilities represented by the periodic table, then variation would necessarily be discrete rather than continuous…The point is that, given the principle of recombination, unless those more fundamental properties exist, unactualized elements would not be physical possibilities (Le Poidevin 2005, 131-132).

Let me try to rephrase the argument. We assume that the combination of a finite number of fundamental properties, via a combinatorial approach, leads to a discrete set of macroscopic physical possibilities. We also know empirically that the chemical elements occur in a discrete manner since there are no intermediate elements between, say, hydrogen and helium. Le Poidevin is thus claiming that his combinatorial approach can be taken as an explanation for the discreteness in the occurrence of elements and furthermore that it justifies the fact that Mendeleev regarded the yet undiscovered elements like gallium as being physical possibilities rather than merely logical ones.

4. Further comments on Le Poidevin

One might even grant that Le Poidevin's argument provide the sought after justification for the ontological reduction of the chemical elements to fundamental physical properties. But has Le Poidevin provided any grounding for the ontological reduction of chemistry *tout court*? I think not. For example, the solubilities of elements which the author included in his list of paradigmatically chemical properties does not occur in a discrete manner. A particular ionic compound can have a solubility of 5 grams per liter. Another one might have a solubility of 6 grams per liter of water. But there is nothing discrete about solubility. It is quite possible that other salts will display solubilities falling *anywhere* between these two values.

Unlike the existence of chemical elements, which does appear to be a discrete phenomenon, solubility or acidity or indeed almost every "paradigmatically chemical property" does not form a discrete set. As a result one cannot invoke a combinatorial argument of the type suggested by le Poidevin in order to provide an ontological grounding for these properties.

As to whether Le Poidevin has separated the question of ontological reduction as fully from that of epistemological reduction as he seemed to promise in his article, I have some doubts. Admittedly, the ordering of the chemical elements may not be in any sense theoretical, as he states, but there is no denying that ordering the elements by way of atomic number, or by whatever other means, is dependent on our *knowledge* of the elements. It is just that this knowledge takes the form of a classification or ordering rather than a theory as Le Poidevin correctly points out. But surely this does not render the act of classification any less epistemological.

Finally, I would like to point out some specific points concerning Le Poidevin's analysis. Let me return to the question of the discrete manner in which the elements occur. Le Poidevin takes this fact to support a combinatorial argument whereby a finite number of fundamental entities combine together to give a discrete set of composite

elements. But what if we consider the combination of quarks (charge = 1/3), instead of protons (charge = 1)? In the former case a finite number of quarks would also produce a discrete set of atoms of the elements only the discreteness would involve increments of one-third instead of integral units. In fact chemists and physicists have been actively searching for such 'quark matter' (Jørgensen 1978).

And if this matter were found, then it would be physically possible for there to be two elements between say Z = 19 and Z = 20 to use Le Poidevin's example. Let us further suppose that a future theory might hold that the fundamental particles are some form of sub-quarks with a charge of 0.1 units. Under these conditions combinatorialism would lead to the existence of nine physical possibilities between elements 19 and 20, and so on. It would appear that Le Poidevin's distinction between a physical possibility, as opposed to a merely logical one, is dependent on the state of knowledge of fundamental particles at any particular epoch in the history of science which is surely not what he intends. Indeed the distinction proposed by Le Poidevin would appear to be susceptible to a form of vacuity, not altogether unlike that faced by physicalism, and which was supposed to be circumvented by appeal to combinatorialism.

Finally there is a somewhat general objection to the use of combinatorialism in order to ground the ontological reduction of chemistry. It would seem that the assumption that fundamental entities combine together to form macroscopic chemical entities ensures from the start that the hoped for asymmetry is present. If one assumes that macroscopic chemical entities like elements are comprised of sub-atomic particles then of course it follows that the reverse is not true. The hoped for asymmetry appears to have been written directly into the account, I claim, rather than deduced.

5. Conclusion

After many years during which philosophers of chemistry concentrated on the question of the epistemological reduction of chemistry, and had perhaps dismissed the question of ontological reduction as a foregone conclusion, there has been a recent resurgence of interest in the ontological question. McLaughlin has used the success of the quantum theory of chemical bonding to conclude *incorrectly* that the emergence of chemistry from physics is entirely ruled out. Le Poidevin claims to have given an ontological argument in favor of the reduction of chemistry which does not appeal to any physical theories and yet it appears to do just that.

My own conclusion is that one should exercise moderation between an extreme Quinean approach of attending mainly to scientific theories and Le Poidevin's approach of dispensing altogether with the findings of scientific theories. Surely a more subtle approach is required in trying to uncover the ontology of chemistry or any other special science. Of course one needs to consult the findings of the empirical sciences in question, but there is still scope for philosophical consideration, perhaps along the general lines offered by Le Poidevin. Philosophical positions such as reductionism, atomism and emergence cannot be judged only on the basis of some contemporary theory or other. In addition if one does consult the findings of scientific theories to draw ontological lessons it is essential for one to do so in an accurate manner and not in the way that these two authors appear to have done. Nevertheless, it is encouraging to now see mainstream philosophers now taking an interest in chemistry.

REFERENCES

Cunningham, B. (2001), "The Reemergence of Emergence", *Philosophy of Science*, 68 (Proceedings): S62-S75.

Jørgensen, Christian (1978), "Predictable Quarkonium Chemistry", *Structure & Bonding*, 34: 19-38.

Kim, Jagwon (1999), "Making Sense of Emergence", *Philosophical Studies*, 95: 3-36.

Le Poidevin, Robin (2005), "Missing Elements and Missing Premises, A Combinatorial Argument for the Ontological Reduction of Chemistry", *British Journal for the Philosophy of Science*, 56: 117-134.

Lightman, Alan (2005), *The Discoveries*, New York, Pantheon Books.

McLaughlin, Brian (1992), "The Rise and Fall of British Emergentism", In *Emergence or Reduction? Essays on the Prospect of a Non-Reductive Physicalism*, A. Beckerman, H. Flohr, J. Kim (eds.), Berlin, Walter de Gruyter, 49-93.

Paneth, Freidrich (1962), "The Epistemological Status of the Concept of Elements", *British Journal for the Philosophy of Science*, 13: 1-14, 144-160. Reprinted in *Foundations of Chemistry*, 5: 113-145, 2003.

Pauling, Linus (1928), "The Shared-Electron Chemical Bond", *Proceedings of the National Academy of Science*, USA, 14: 359-362.

Scerri, Eric R. (1994), "Has Chemistry Been at Least Approximately Reduced to Quantum Mechanics?", in *PSA 1994 vol 1*, D. Hull, M. Forbes, and R. Burian, eds., Philosophy of Science Association, East Lansing, MI, 160-170.

Scerri, Eric R. (2004), "How ab initio is ab initio quantum chemistry", in *Foundations of Chemistry*, 6: 93-116.

Whitesides, George, Ismagilov, R.F. (1999), "Complexity in Chemistry", *Science*, 284: 89-92.

Notes

1. I am referring here to molecular orbital theory as developed by Mulliuken, Hund and others which is mathematically equivalent to the valence bond method to which Pauling made seminal contributions. The notion of bonds as pairs of electrons is also retained in the valence bond method that in many senses is closer to Lewis' classical theory.
2. This article is singled out, and reproduced, in a recent book by Alan Lightman as one of the 22 most influential scientific articles of the twentieth century. (Lightman, 2005).
3. Admittedly Pauling discovered that protein molecules have the structure of an α helix and this was a step towards the realization by Crick and Watson that DNA has a double helical structure. But no quantum mechanics went into Pauling's discovery. Furthermore, Pauling was involved in the race to find the structure of DNA but by his own admission was working on altogether the wrong track. Neither he nor Crick and Watson employed any quantum mechanics in their search for the structure of DNA.
4. This is not to say that Kim supports either emergence or non-reductive physicalism. In fact he argues that non-reductive physicalism in particular represents an unstable position (Kim, 1999).
5. Except for possessing an atomic weight which is the characteristic property of an element as a basic substance for Mendeleev. In modern terms, the characteristic property becomes atomic number.

Eörs Szathmáry

Statement

and

Readings

How to organize inanimate processes into living systems?

Eörs Szathmáry

Although definitions, unlike theories, cannot be falsified, they can be more useful or less so. Useful definitions aid conceptualization and foster good research. Most important for the topic of emergence is the concept of minimal life. According to Ganti's theory, minimal living systems consist of three coupled autocatalytic subsystems: (1) a metabolic cycle (energy and material supply), (2) template replication (informational processes), and (3) a fluid membrane (container). The theory, first conceived in 1971, is more timely than ever. Any two of the above three autocatalytic systems can form a so-called infrabiological system, with interesting properties but no full-fledged capacity for life. Theoretical results, showing the emergence of qualitatively new properties, and attempts at experimental realization, will be discussed.

References:

Ganti, T. (2003) *The Principles of Life*. Oxford Univ. Press.

Maynard Smith, J. & Szathmáry, E. (1995) *The Major Transitions in Evolution.* Freeman & Co., Oxford.

Szathmáry, E. & Maynard Smith, J. (1995) The major evolutionary transitions. *Nature* **374,** 227-232.

Szabó, P., Scheuring, I., Czárán, T. & Szathmáry, E. (2002) *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420,** 360-363.

Fernando, C., Santos, M. & Szathmáry, E. (2005) Evolutionary potential and requirements for minimal protocells. *Top. Curr. Chem.* **259,** 167-211.

Szathmáry, E (2005) Life: in search of the simplest cell. *Nature* **433,** 469-470.

Szathmáry, E (2006) The origin of replicators and reproducers. *Phil. Trans. R. Soc. Lond. B. Biol. Sci.* **361,** 1761-1776.

Life

In search of the simplest cell

Eörs Szathmáry

Top-down, bottom-up; RNA-based, lipid-based; theory, experiment — there are many different ways of investigating what constitutes a 'minimal cell'. Progress requires finding common themes between them.

In investigating the origin of life and the simplest possible life forms, one needs to enquire about the composition and working of a minimal cell that has some form of metabolism, genetic replication from a template, and boundary (membrane) production. Approaches to this intriguing problem are discussed in Tibor Gánti's *The Principles of Life* (Oxford Univ. Press, 2003), and were also debated at a meeting last December*.

Identifying the necessary and sufficient features of life has a long tradition in theoretical biology. But living systems are products of evolution, and an answer in very general terms, even if possible, is likely to remain purely phenomenological: going deeper into mechanisms means having to account for the organization of various processes, and such organization has been realized in several different ways by evolution. Eukaryotic cells (such as those from which we are made) are much more complicated than prokaryotes (such as bacteria), and eukaryotes harbour organelles that were once free-living bacteria. A further complication is that multicellular organisms consist of building blocks — cells — that are also alive. So aiming for a general model of all kinds of living beings would be fruitless; instead, such models have to be tied to particular levels of biological organization.

Basically, there are two approaches to the 'minimal cell': the top-down and the bottom-up. The top-down approach aims at simplifying existing small organisms, possibly arriving at a minimal genome. Some research to this end takes *Buchnera*, a symbiotic bacterium that lives inside aphids, as a rewarding example (A. Moya, Univ. Valencia). This analysis is complemented by an investigation of the duplication and divergence of genes (A. Lazcano, Univ. Mexico). Remarkably, these approaches converged on the conclusion that genes dealing with RNA biosynthesis are absolutely indispensable in this framework. This may be linked to the idea of life's origins in an 'RNA world', although such an inference is far from immediate.

Top-down approaches seem to point to a minimum genome size of slightly more than 200 genes. Care should be taken, however, in blindly accepting such a figure. For example, although some gene set A and gene set B may not be common to all bacteria, that does not mean that (A and B) are dispensable. It may well mean that (A or B) is essential, because the cell has to solve a problem by using either A or B. Only experiments can have the final word on these issues.

There was general agreement that a top-down approach will not take us quite to the bottom, to the minimal possible cells in chemical terms. All putative cells, however small, will have a genetic code and a means of transcribing and translating that code. Given the complexity of this system, it is difficult to believe, either logically or historically, that the simplest living chemical system could have had these components.

The bottom-up approach aims at constructing artificial chemical supersystems that could be considered alive. No such experimental system exists yet; at least one component is always missing. Metabolism seems to be the stepchild in the family: what most researchers in the field used to call metabolism is usually a trivial outcome of the fact that both template replication and membrane growth need some material input. This input is usually simplified to a conversion reaction from precursors to products.

Even systems missing one or the other component can, of course, advance our understanding. Such systems could be called 'infrabiological', because they are not quite biological but are similar to living systems in some crucial respects: elementary combinatorics suggests that out of metabolism (M), boundary (B) and template (T) three dual systems can be built — MT, MB, TB. In particular, coupling of compartment formation with some form of template replication (TB) is the subject of many experiments.

Following earlier work on liposomes (P. Walde, Univ. Zurich), protein expression in these entities has become a viable prospect: liposomes are tiny bags with walls made of layers of phospholipids, like the phospholipids that make up cell membranes. Even composite systems incorporating gene transcription and translation are now possible in liposomes. For example, an artificial stretch of DNA can harbour the gene for T7 RNA polymerase, an enzyme that catalyses the production of RNA from DNA, which in turn induces the expression of green fluorescent protein as an indicator of translation (T. Yomo, Univ. Osaka;

**Towards the Minimal Cell.* Erice International School on Complexity, Erice, Sicily, 7–10 December 2004.

100 YEARS AGO

The Preparation of the Child for Science. **A great change in the character of the books concerned with the teaching of science has taken place during the last twenty years or so. A quarter of a century ago the claims of science to a place in the school curriculum were being advocated vigorously, and men of science had still to convince reigning schoolmasters that no education was complete which ignored the growth of natural knowledge and failed to recognise that an acquaintance with the phenomena of nature is necessary to intelligent living. Speaking broadly, it may be said that most classicists even admit now that there are faculties of the human mind which are best developed by practice in observation and experiment. One consequence of the success which has followed the persistent efforts of Huxley and his followers – to secure in the school an adequate recognition of the educative power of science – has been that modern books on science teaching are concerned almost entirely with inquiries into the best methods of instructing young people, by means of practical exercises, how to observe accurately and to reason intelligently.**
From *Nature* 2 February 1905.

50 YEARS AGO

Principles of Geomorphology. **Geomorphology as a science has grown up in the railway age. A hint of what was coming might be espied in those eighteenth-century travellers who, like Gilpin, began very haltingly to display an interest in the form of landscape rather than its formalized versions. A hundred years later and the trains have reached Lucerne; soon we are well into the age of physiography, that pleasant ill-defined compost which made an agreeable part of the later Victorian education. A further hundred years, and this lively branch of science has given birth to a remarkable variety of new and odd words such as pediplains, steptoes and fluviraption… Progress has been rapid; yet the discussion of the characteristics, origin and development of land-forms will long continue to provide an attractive and challenging mental discipline and a valuable education. Geomorphology not only gives scope for the exploratory and cartographical type of mind but also allows abundant opportunity to increase with time the precision of measurement, examination and analysis. Probing, indeed, may gradually replace mapping in this as in other fields.**
From *Nature* 5 February 1955.

K. Tsumoto, Mie Univ., Tsu). The snag is, of course, that these systems contain components taken from contemporary cells, and are far from being self-sufficient.

Replication can also happen in liposomes. RNA from the phage Qβ (a virus infecting bacteria) can be incorporated in liposomes (T. Yomo) and be replicated by a replicase enzyme provided by the experimenter. A common by-product of RNA replication is the advent of smaller, faster-replicating mutant RNA molecules, which take over the population. This apparently failed to happen in these experiments, but the reason is debatable. Maybe self-association of template and copy strands reduced competition to such an extent that coexistence is guaranteed (G. von Kiedrowksi, Univ. Bochum). Or perhaps the efficient mutants simply failed to arise owing to the small number of replication cycles (E. Szathmáry).

Experimental work is increasingly being complemented by computational investigations. For example, it is possible to account for the growth and fission of compartments in simulations of molecular-assembly dynamics (T. Igekami, Univ. Tokyo). On the genetic side, the origin of heredity was demonstrated in a simulated system of cross-catalytic autocatalytic networks (K. Kaneko, Univ. Tokyo). Kaneko argued that 'minority control' is a possible origin of heredity in a bag of genes that constitutes a primordial genome, in that genes with a lower copy number have a more decisive influence on the protocell's simulated behaviour. It is difficult to assess the importance of this finding, as there is no example of the particular network modelled. But the idea may prove helpful in attempts to produce more realistic constructions.

According to the 'composome' model, in which micelles or vesicles are formed from amphiphilic compounds — those having one end that is hydrophilic and the other hydrophobic — there is the prospect of constructing a 'lipid world'. Here, a hereditary component arises from alternative autocatalytic sets of lipids (D. Segré, Harvard Med. School).

Clearly, there is a divide between the top-down and bottom-up approaches, and between theoretical and experimental investigations. In the future, for example, one would like to see more realistic models of the primordial genome and, conversely, an experimental approach to the lipid world. An aim in the coming years will be to bridge those gaps — hence the great value of meetings such as this. ■

Eörs Szathmáry is at the Collegium Budapest (Institute for Advanced Study), 2 Szentháromság utca, H-1014 Budapest, Hungary.
e-mail: szathmary@colbud.hu

Surface chemistry

Approximate challenges

Greg Sitz

There is growing evidence that the usual approach to modelling chemical events at surfaces is incomplete — an important concern in studies of the many catalytic processes that involve surface reactions.

To describe all the transformations through which a molecule must go during a chemical reaction is a daunting task. The intermediate transition states of a reaction are hard to examine directly, and theory is needed to obtain a full understanding of all the relevant interactions. In 1927, Born and Oppenheimer formulated an 'approximation', which greatly simplified such calculations. Their theory has been crucial to advances in theoretical and chemical physics. It is therefore of great interest when the Born–Oppenheimer approximation breaks down, which may be the case particularly for reactions that take place at surfaces. On page 503 of this issue[1], Jason White and colleagues provide the clearest example to date of such a case.

The break-up of a chemical bond involves a large bond vibration — in other words, a large relative motion of the two atoms that make up the bond. Rather than taking into account all the interactions involved, the Born–Oppenheimer approximation treats the motion of atomic nuclei separately from electronic excitation. This is justified by the fact that nuclei are much heavier than electrons and move more slowly. Therefore — it is assumed — when nuclei move, as they do during the formation or breaking of a bond, electrons will simply readjust quickly.

Many theoretical methods use this approximation, and solve the Schrödinger equation (the fundamental equation that describes all such interactions) in terms of electrons moving in slowly changing, stationary frameworks of nuclear arrangements. The result can be visualized as a 'potential-energy surface', which plots the solutions of the Schrödinger equation as a function of a molecule's changing structure during a reaction — a popular method for describing chemical reactions. However, although the Born–Oppenheimer approximation has been widely tested for gas-phase reactions and its limitations are known, the situation for reactions at surfaces is much less clear.

In their experiments, White and colleagues[1] prepared nitric oxide molecules in highly excited vibrational states, so that the atoms were subjected to large motion, close to the limit at which the molecules will break up. The excited molecules were scattered from a specially prepared metal surface from which electrons could escape easily. A detector above the surface picked up any electron emission. The experiment's main observation was that when the vibrational energy of the incident nitric oxide molecule exceeded the binding energy of electrons in the surface, electrons were directly emitted from the surface. This finding points to a coupling between nuclear motion and electronic excitation, and therefore indicates that the Born–Oppenheimer approximation is invalid in this case.

The research by White *et al.* extends work in which electronic excitation was produced at metal surfaces by bombardment with various gas-phase species (mostly atoms such as oxygen, hydrogen and nitrogen, high-kinetic-energy rare gases and some molecules)[2,3]. In one of these experiments[3], electrons in the metal tunnelled through a potential-energy barrier to a semiconductor substrate as a result of the bombardment. The charge flow induced in the semiconductor as a result of the tunnelling electrons was termed a 'chemicurrent', to reflect the chemical cause of the electronic excitation.

Although these previous results also point to a breakdown of the Born–Oppenheimer approximation, the situation is somewhat harder to interpret because the electronic excitation is most probably mediated by 'phonons' — vibrational excitations in the substrate itself. White and colleagues' experiment bypasses this poorly defined intermediate step.

Experiments of the type presented by White *et al.* (and the closely related chemicurrent work[3]) serve as a warning over the widespread use of potential-energy surface models, and should act as an impetus for modifying the conceptual framework used in surface chemistry. There have been attempts to include electronic excitation in theoretical models, but the task is a daunting one and has been limited by a lack of clear experimental findings. The new experiments provide well-characterized results to guide further theoretical development. ■

Greg Sitz is in the Department of Physics, University of Texas, 1 University Station C1600, Austin, Texas 78712, USA.
e-mail: gositz@physics.utexas.edu

1. White, J. D., Chen, J., Matsiev, D., Auerbach, D. J. & Wodtke, M. *Nature* **433,** 503–505 (2005).
2. Amirav, A. & Cardillo, M. J. *Phys. Rev. Lett.* **57,** 2299–2302 (1986).
3. Nienhaus, H. *Surf. Sci. Rep.* **45,** 3–78 (2002).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2006) **361**, 1761–1776
doi:10.1098/rstb.2006.1912
Published online 7 September 2006

The origin of replicators and reproducers

Eörs Szathmáry[1,2,*]

[1]*Collegium Budapest (Institute for Advanced Study), 2 Szentháromság utca, 1014 Budapest, Hungary*
[2]*Department of Plant Taxonomy and Ecology, Institute of Biology, Eötvös University, 1/c Pázmány Péter sétany 1117 Budapest, Hungary*

Replicators are fundamental to the origin of life and evolvability. Their survival depends on the accuracy of replication and the efficiency of growth relative to spontaneous decay. Infrabiological systems are built of two coupled autocatalytic systems, in contrast to minimal living systems that must comprise at least a metabolic subsystem, a hereditary subsystem and a boundary, serving respective functions. Some scenarios prefer to unite all these functions into one primordial system, as illustrated in the lipid world scenario, which is considered as a didactic example in detail. Experimentally produced chemical replicators grow parabolically owing to product inhibition. A selection consequence is survival of everybody. The chromatographized replicator model predicts that such replicators spreading on surfaces can be selected for higher replication rate because double strands are washed away slower than single strands from the surface. Analysis of real ribozymes suggests that the error threshold of replication is less severe by about one order of magnitude than thought previously. Surface-bound dynamics is predicted to play a crucial role also for exponential replicators: unlinked genes belonging to the same genome do not displace each other by competition, and efficient and accurate replicases can spread. The most efficient form of such useful population structure is encapsulation by reproducing vesicles. The stochastic corrector model shows how such a bag of genes can survive, and what the role of chromosome formation and intragenic recombination could be. Prebiotic and early evolution cannot be understood without the models of dynamics.

Keywords: replicator; origin of life; ribozyme; autocatalysis; compartmentation; error threshold

1. INTRODUCTION

The replicator, as introduced by Dawkins (1976), has become one of the central concepts in evolutionary theory. He identified two types of replicator with unbounded evolutionary potential, namely genes and memes (memes were meant to be hereditary units of cultural rather than genetic evolution). These ideas have turned out to be extremely fruitful: they have elicited renewed interest in the philosophy of evolution (e.g. Hull 1980) and led to the recognition of other types of replicators with the most important role in evolution (Maynard Smith & Szathmáry 1993, 1995).

A classification of replicators was presented by Maynard Smith & Szathmáry (1995) and it has been refined a number of times (Szathmáry 1995, 2000). Most widely known replicators, including genes, are strongly tied to the world of chemistry: this is obviously not true for memes. Some replicators have only limited heredity (Maynard Smith & Szathmáry 1995), implying that the number of possible types is smaller than or roughly equal to the number of individuals (copies, sequences, etc.) in a plausible (realistic) system. Conversely, in the case of unlimited hereditary replicators, the number of types by far exceeds that of individuals in the population (Szathmáry & Maynard Smith 1997). This shows that a classification of replicators is not naturally hierarchical: there exist molecular and non-molecular replicators with limited or unlimited hereditary potential.

Oparin (1961) defined any system capable of replication and mutation as alive. Most evolutionary biologists would agree with this view. Systems with these properties can evolve complex adaptations (purposeful functions) in the natural world, highly characteristic of living beings. Yet some authors (including Gánti 1971, 1978) have raised doubts concerning such an approach. The acid test is whether viruses are alive or not. Gánti (1971) argued that to regard viruses as living amounts to a conceptual mistake equating programs with computers. In the full analogy, the virus corresponds to a program, written in a decodable language, which says to the computer: 'Print me again and again, even if you disintegrate as a result of doing so!' The active part is obviously the computer and not the program. The computer can do many things without such a malign program. In sharp contrast, the program cannot do anything on its own. The living cell is thus analogous to the computer. Since everyone regards the cell in its active state alive, life as such in the example rests with the cell rather than the virus.

Yet viruses evolve. In fact, they have become one of the most accessible test systems for evolutionary hypotheses (e.g. Poon & Chao 2004). Computer programs can also evolve (e.g. Bedau *et al.* 2000). What is the relationship between units of evolution and units of life? To give a tentative answer, both the concepts must be defined first with sufficient clarity,

*szathmary@colbud.hu

One contribution of 19 to a Discussion Meeting Issue 'Conditions for the emergence of life on the early Earth'.

and only after this the two notions can be compared. Units of evolution must: (i) multiply, (ii) have heredity and (iii) heredity must not be totally accurate (variability). Furthermore, some of the inherited traits must affect the chance of reproduction or of survival of the units. If all these criteria are met, then in a population of such entities, evolution by natural selection can take place (Maynard Smith 1986). Note that this definition does not refer to living systems. Any system satisfying these criteria can evolve in a Darwinian manner.

Units of life as such are less well studied, although cells and organisms are widely known and analysed. Gánti (1971, 1979, 1987, 2003) has refined his 'life criteria' that living systems must meet. He observed, correctly, that for the individual living state, reproduction is neither necessary nor sufficient. Many cells and organisms are commonly regarded alive even if they cannot reproduce (any longer). The so-called potential life criteria must be met only if the population of units is to be maintained and evolved. Then, the correct relationship between units of evolution and units of life is that of two partially overlapping sets (Szathmáry 2002).

Some regard the concept of a replicator more informational, detached from real processes of replication, reproduction and development. The elegant concept of a reproducer (Griesemer 2000, 2002) is meant to fill this gap. A reproducer is a unit of multiplication, hereditary variation and development. A reproducer must have at least a minimum developmental capacity required for further multiplication. There is not only an informational link but also material overlap between generations of reproducers. Thus, genes in an organism are replicators but not reproducers. Conversely, an organism is not a replicator but reproducer. In the course of prebiotic and early biological evolution, replicators ganged up to yield reproducers. We shall consider in detail how this could have happened.

2. SURVIVAL CRITERIA FOR INFORMATIONAL REPLICATORS

Informational replicators, such as genes, have unlimited heredity. The earliest informational replicators must have faced at least two severe constraints. Serious considerations suggest that primordial nucleic acids (or their analogues) must have been rather short molecules owing to excessive noise in their copying. Another consideration emphasizes the fact that replicators must have a growth rate high enough to compensate for spontaneous decay. I consider these two aspects in turn.

(a) *The error threshold*

Eigen (1971) called attention to the fact that the length of molecules (number of nucleotides) maintained in mutation–selection balance is limited by the copying fidelity. We recapitulate the simplified treatment by Maynard Smith (1983). Imagine two sequences with replication rate constants K and $k(<K)$, respectively. The first sequence mutates into the second with a mutation rate $(1-Q)$. If we assume that they are in a flow reactor where total concentration is kept constant, then the rate equations for growth and competition become

$$\mathrm{d}x/\mathrm{d}t = xKQ - x\Phi, \tag{2.1a}$$

$$\mathrm{d}y/\mathrm{d}t = yk + xK(1-Q) - y\Phi, \tag{2.1b}$$

where x and y are concentrations of wild-type and mutant, respectively, $\Phi = xK + yk$ and total concentration is (without loss of generality) unity. It is easy to see that in equilibrium, when both templates are present in non-zero concentration, it holds that

$$x = \frac{(KQ-k)}{(K-k)}, \tag{2.2}$$

where it must be true that $Q > k/K$. If there are ν digits in the sequence, $Q = q^{\nu}$ can be approximated by $\mathrm{e}^{-\nu(1-q)}$, where q is the copying fidelity per base per replication. From this we obtain

$$\nu < \frac{\ln(K/k)}{(1-q)}, \tag{2.3}$$

which is Eigen's *error threshold of replication*. Non-enzymatic replication implies low q, so $\nu < 100$ is probable for prebiotic chemistry, which is about the size of a tRNA molecule. Therefore, early genomes must have consisted of independently replicating entities. But they would compete with each other and the one with the highest fitness would win (Eigen 1971). Hence, the 'Catch-22' of molecular evolution: no enzymes without a large genome and no genome without enzymes (Maynard Smith 1983).

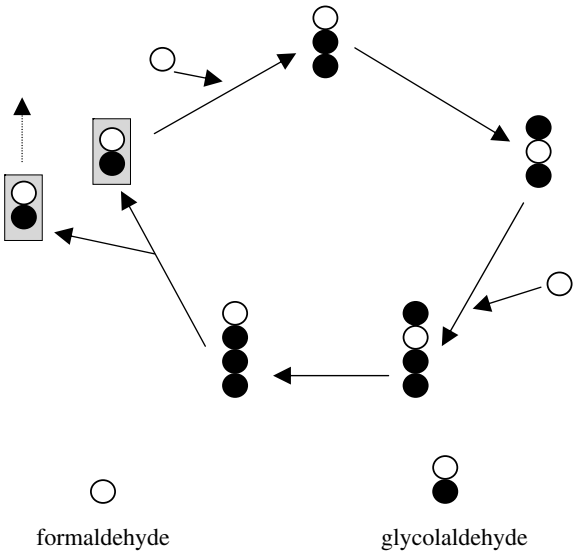


Figure 1. The autocatalytic core or seed of the formose reaction (Fernando *et al.* 2005). Each circle represents a chemical group including one carbon atom.

(b) *The decay threshold*

Consider, for a change, a non-informational replicator, such as any intermediate in the formose reaction (figure 1). Note that such an autocatalytic cycle differs markedly from Kauffman's (1993) reflexively autocatalytic protein nets: in the former, each elementary reaction is stoichiometric rather than catalytic. There is a severe problem with the formose reaction: deadly side reactions drain it to such an extent that the

intermediates of the cycle disappear ultimately (e.g. Shapiro 1986). This may have been different for cycles on surfaces, but we do not know (yet). As King (1982, 1986) pointed out, the smaller the cycle, the better the chances for its propagation. Suppose that there is a simple autocatalytic cycle of p steps (similar to the system in figure 2, where $p=4$). At each step, the legitimate reaction leads to the next cycle intermediate, and a number of side reactions drain the system. The latter give rise to all sorts of unwanted by-products. Let the specificity of a reaction at step i be s_i, which is the rate of legitimate reaction divided by the total rate of all (legitimate + side) reactions. Successful growth of the cycle is guaranteed if

$$2\prod_{i=1}^{p} s_i > 1, \tag{2.4}$$

or if we calculate with the geometric mean σ of the specificities

$$\sigma^p > 1/2, \quad \text{i.e.} \quad p < -\log(2)/\log(\sigma). \tag{2.5}$$

This shows that the viable system size p increases hyperbolically with specificity. Let us apply Eigen's (1971) full dynamical formalism to this problem (Szathmáry 2002) by assuming that there can be a number of *alternative* cycles such as the formose reaction that occasionally can produce each other's intermediates:

$$\dot{x}_i = (R_i Q_i - D_i)x_i + \sum_{j\neq i}^{n} w_{ij}x_j - x_i F, \tag{2.6}$$

where x_i is the concentration of species i; R_i, the rate of replication irrespective of the correctness of the offspring; Q_i, the fidelity of replication; D_i, the rate of spontaneous decomposition; w_{ij}, the mutation rate from species j to species i; and F, an outflow ensuring that the total concentration remains unity. Here, the different 'species' mean the catalytic seeds of different alternative cycles (if their existence is feasible, see below), and 'mutation' refers to the 'macromutation', producing an intermediate of another autocatalytic cycle. Spontaneous decay corresponds to irreversible side reactions; in the case of DNA, it means damage (rather than mutation; damaged DNA is chemically no longer DNA).

When is species i viable? It means that it can increase in concentration when rare. If we forget about selection of, and mutations to, this species for a moment, from equation (2.6) we obtain

$$R_i Q_i - D_i > 0, \quad \text{or} \quad R_i Q_i > D_i, \tag{2.7}$$

which after rearrangement yields

$$1 > Q_i > D_i/R_i > 0, \tag{2.8}$$

where it also holds that

$$R_i > D_i. \tag{2.9}$$

Lack of enzymatic catalysis implies that the decay rate is rather high. Inequalities (2.8) and (2.9) suggest that copying fidelity must be high. Fortunately, this fits, since mutations are expected to be very rare in the systems composed of cycles of small molecules (most fluctuations cannot propagate their own kind). Thus for autocatalytic cycles, damage is the most severe hurdle (Szathmáry 2000). The same considerations necessarily apply to the fittest cycles. If they coexist, ecology tells us that they must occupy different niches in abstract space, such as requiring different combination of raw materials.

An alternative way of maintaining a variety of cycles is a high mutation rate (low copying fidelity). This is true, but low copying fidelity does not allow the selection for the fittest, because the system gets below the error threshold of replication (see §2*a*). In such a case, the cycles would cease to be selectable individuals: they would rather form a single, un-evolvable network.

Orgel (1992) called attention to the fact that the intermediates of formose reaction are not informational replicators. In the prebiotic context, Wächtershäuser (1992) called attention to the possibility that there could be, in principle, a limited set of metabolic replicators. These replicators could have limited heredity, allowing some evolution by natural selection. This possibility is intriguing, but it is without any direct experimental support at present: nobody has seen a metabolic replicator, other than the formose reaction, that would run without enzymes. In contemporary systems, such cycles (the Calvin cycle, the reductive citric acid cycle) are well above the damage threshold outlined here, owing to the rate-enhancing effect of evolved enzymes. Thus, *the requisite degree of metabolic channelling is one of the biggest (if not the biggest) hurdles of the origin of life.*

3. INFRABIOLOGICAL SYSTEMS AND THE LIPID WORLD SCENARIO

We do not know where RNA came from. Some people think that the first replicators were not even template-based; as we shall see reproducing compartments (vesicles, micelles) are favoured by some. Others see the crucial steps in the linking of different autocatalytic systems that ultimately could evolve into primitive living systems.

(a) *Infrabiological systems*

Gánti (e.g. 2003) emphasized that contemporary living systems always have: (i) some metabolic subsystem, (ii) some systems for heritable control and (iii) some boundary system to keep the component together. So I consider it unlikely that a chemical system satisfying all the constraints from this abstraction could have appeared just out of chemical chaos. This observation led to the formulation of the concept of infrabiological systems (Szathmáry 2005; Fernando *et al.* 2005). Infrabiological systems always lack one of the key components just listed. For example, in the original formulation of Ganti (1971), a model of minimal life did not include a boundary system. The combination of a metabolic cycle and a membrane was conceived also by Gánti (1978), and called a self-reproducing microsphere. In contrast, Szostak *et al.* (2001) conceived a protocell-like entity with a boundary and template replication but no metabolic subsystem. Such systems show a crucial subset of necessary biological phenomena. The three subsystems can be combined to yield three different doublet systems (figure 2).

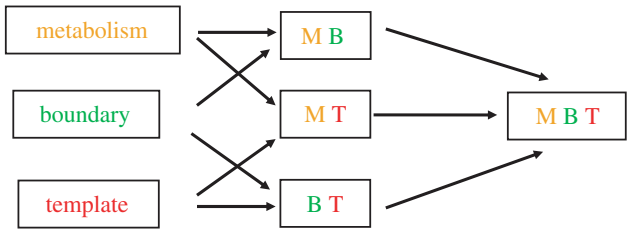


Figure 2. Elementary combinatorics of infrabiological systems (Fernando *et al.* 2005). The chemoton is a biological minimal system comprising three qualitatively different subsystems (metabolism, membrane and template).

(b) *Composomes and the graded autocatalytic replication domain model*

An interesting line of research has been initiated by Doron Lancet with his group, conveniently referred to as the 'lipid world' scenario (Segré *et al.* 2001*a*). The basic idea is as follows. We know that lipids (more generally, amphiphilic compounds with a hydrophobic tail and a hydrophilic head) tend to form supramolecular structures, such as bilayers, micelles and vesicles. They can grow autocatalytically. Now imagine that we have a mixture of molecules in any one vesicle. Some of them may act as catalysts of certain reactions. It is theoretically possible that some will catalyse their own incorporation (direct autocatalysis), or there will be a gang of molecules each exerting some catalytic function; thus as a net result, the incorporation of all members of the gang is ensured by the gang (reflexive autocatalysis). If this idea holds water, membrane heredity in the lipid world, and natural selection of vesicles without a genetic subsystem, would be feasible. The different, reflexively autocatalytic gangs would constitute compositional genomes or 'composomes' (Segré *et al.* 2001*b*). Note that the model does not deal with the formation of the lipid constituents: they are assumed to be there in the surrounding soup.

Now, there is nothing mysterious about compositional genomes in the first place. Although relying on direct autocatalysis at the molecular level, the genome of the stochastic corrector (see §7) is also a compositional genome in which the genes are unlinked and the genome is characterized by gene composition. Formally, each protocell can be characterized by a genome vector with entries denoting the number of copies of the *i*th gene in that vesicle. The change in this number is a stochastic process, which can be characterized by mean and variance. A crucial difference is that, in the stochastic corrector model, we are dealing with a bag of template replicators: there are no genes in Lancet's model.

A similar approach is possible while considering questions in the lipid world; however, the issue is complicated by the fact that we need to tackle the problem of reflexive autocatalysis. This has also precedence in the literature: the reflexively autocatalytic protein networks (e.g. Kauffman 1993) are perhaps the best-known example. I hasten to point out that nobody has seen real reflexively autocatalytic protein sets. Let us see whether one can be more hopeful regarding autocatalytic lipid sets.

The process imagined is shown in figure 3. It displays a reflexively autocatalytic micelle with many components. The incorporation of amphiphile L_i may be catalysed by amphiphile L_j at rate enhancement β_{ij} (the ratio of catalysed and uncatalysed reaction rates). The crucial question is this: where can one obtain the values of β_{ij}, considering the fact that no such system has been realized so far (the experimental cases are all directly autocatalytic and show no heredity; see Fernando *et al.* 2005 for review)? The authors suggest translating the model developed for molecular recognition between receptors and ligands (Segré *et al.* 1998). If catalysis depends on recognition of substrate by catalyst, the reasoning is sound implying that catalysis is a graded phenomenon. From this empirically constrained theoretical distribution, the authors obtain the distribution of β_{ij} values in their model.

It is imagined that every micelle (or vesicle) is a sample with replacement of a set of possible lipid molecules. Some samples will contain mutually autocatalytic gangs, but not others. The latter ones will not be able to grow. The former will grow and then fragment/divide by some spontaneous process. Micelles containing more efficient gangs (characterized by higher β_{ij} values) will take over. Such sets have some heredity; the gangs maintain and propagate their identity by virtue of their mutual catalytic activity.

What are the major concerns apart from the lack of an experimental basis (at this moment) of this model? In the light of the foregoing, I see the following difficulties:

(i) This model works only if the β_{ij} values are drawn from a lognormal, rather than a normal distribution. In the latter case, there is no interesting composome population.
(ii) The absolute magnitude of the β_{ij} values will also matter. Side reactions, as in many other prebiotic models, are neglected in the lipid world scenario. If the catalytic values are too low, then composomes may shrink below the decay threshold, even if without decay very interesting dynamics may unfold.
(iii) Even if the decay threshold is not reached, composomal replication may be so inaccurate that fitter composomes cannot be maintained by selection; thus the system may be above the corresponding error threshold.

I hope the fascinating scenario of the lipid world scenario will be complemented by theoretical investigations along these lines. Experimental validation is another formidable problem.

(c) *Limited heredity in composomes*

Contemporary DNA-based organisms have an unlimited hereditary potential, since the number of types that one can construct from the purely informational point of view greatly exceeds the number of individuals that the Earth can maintain. What is the hereditary potential of composomes? They can have limited heredity only (Szathmáry 2000). First of all, it is only the composition rather than the steric configuration of the system that is maintained. In order to appreciate

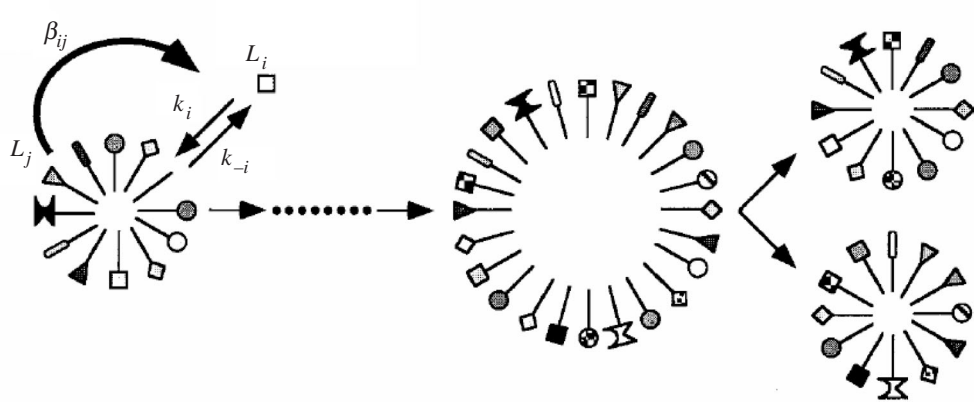


Figure 3. The graded autocatalytic replication domain or composome model: catalysed micelle growth and fission (Segré *et al.* 2001*a*,*b*). L_i and L_j molecules are different amphiphilic compounds, k_i and k_{-i} are rate constants for spontaneous insertion and emigration of amphiphile L_i, and β_{ij} is the rate enhancement of getting in and out of this molecule from the micelle, catalysed by L_j. Note that the model does not deal with the primary origin of L_i molecules *per se*.

this point, consider n types of molecules that we use to build our replicator of size k. In the case of template (digital, see later) replication, all possible sequences are potential replicators; Hence, their number is given by

$$N_s = n^k, \tag{3.1}$$

as it follows from elementary combinatorics. In the case of ensemble replicators, the positions do not matter and hence the upper bound for the number of possible types is

$$N_c = \binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!k!}. \tag{3.2}$$

This is clearly an upper bound since every possible subset cannot be realized by the alternative attractors associated with the system. For the same n and k, N_s is always larger than N_c, usually by orders of magnitude. Indeed, by the application of the Stirling formula for factorials, one can deduce an approximate equation for the proportion of the number of types

$$\frac{N_s}{N_c} \approx k^{k+1/2}(n-1)^{n-1/2}n^k(n+k-1)^{1/2-k-n}\sqrt{2\pi}, \tag{3.3}$$

which, for sufficiently large n and k, further approximates to

$$\frac{N_s}{N_c} \approx k^k n^{k+n}(n+k)^{-k-n}\sqrt{2\pi}. \tag{3.4}$$

Note that the number of attractors for such collective replicators has not been analytically calculated yet. In any case, the ratio (3.4) showing the advantage of modular template replicators is definitely underestimated. A satisfactory answer must take two considerations into account: (i) the number of attractors in sets of unlimited size (Kauffman 1993) and (ii) finite size k for realistic systems (Segré *et al.* 1998).

4. PARABOLIC GROWTH, SURVIVAL OF EVERYBODY AND THE APPEARANCE OF DARWINIAN SELECTION

In the field of prebiotic evolution, non-conventional growth laws, such as hyperbolic and parabolic, have been widely discussed. Both represent departures from simple Malthusian growth: hyperbolic and parabolic growth are faster and slower than Malthusian growth, respectively. Hyperbolic growth was thought to be relevant for hypercycles (mutualistic molecular replicators), whereas parabolic growth was experimentally demonstrated to happen with small synthetic replicators. The consequences for selection in a competitive setting are remarkable: survival of the common for hyperbolic growth and survival of everybody for parabolic growth. In this section, I focus mainly on parabolic growth and its consequences.

(a) *Growth laws and selection consequences*

The simplest reproduction process is the binary fission of the parent object, of which the formal stoichiometry is

$$A + S \rightarrow 2A + W,$$

where A is a replicator, and S and W are source and waste materials, respectively (here I follow the treatment of Szathmáry & Maynard Smith, 1997). The associated kinetic equation describes a Malthusian growth process

$$\frac{dx}{dt} = \dot{x} = kx, \tag{4.1}$$

which means that growth of x (the concentration of A) is exponential with a per capita rate constant k, provided the concentration of S is kept stationary. When two replicators with different rate constant grow together, the one with larger k will outgrow the other. This is, of course, elementary. For didactic purposes, let us express this outcome through the ratios of the growing concentrations

$$\frac{x_1(t)}{x_2(t)} = \frac{x_1(0)e^{k_1 t}}{x_2(0)e^{k_2 t}} = Ce^{gt}, \quad g = k_1 - k_2 > 0, \tag{4.2}$$

showing that even in a freely growing system, the worse growing population is diluted out in the limit. This is a very simple demonstration of differential survival.

Departures from this simple scheme are easily imaginable. A minimum complication is that two individuals are necessary to produce a third one

(akin to sexual reproduction), such as:

$$2A + X \rightarrow 3A + Y,$$

and the associated growth equation reads

$$\dot{x} = kx^2, \tag{4.3}$$

which is called hyperbolic growth, the selection consequences of which are very interesting (Eigen 1971). In order to see this, let us replace the exponent 2 by p and solve the equation by separation to obtain

$$x(t) = [kt - kpt + x(0)^{1-p}]^{1/(1-p)}. \tag{4.4}$$

When $p > 1$, defining hyperbolic growth, the system has a finite escape time, i.e. it reaches infinite concentration in finite time. As it is easy to check, for $p = 2$ the asymptote lies at $t = 1/[x(0)k]$. The smaller the time of unbounded explosion, the larger $x(0)k$. Among the competitors, the one with *the highest initial concentration times the growth rate constant* wins. Thus, initial conditions also determine the outcome of selection and this phenomenon has been called the 'survival of the common', where intrinsic fitness is masked by the growth law (Michod 1983, 1984).

The relevance of hyperbolic growth and survival of the common may be as follows. Eigen (1971) proposed that the hypercycle might have been a link between solitary genes and bacterial genomes. It is a cycle of replicators in which any member catalyses the replication of the next. Each member undergoes a replication cycle as an autocatalyst, and there is the superimposed cyclic network of heterocatalytic aid, hence the term hypercycle. Under simplifying kinetic assumptions, the members of the hypercycle grow coherently and hyperbolically (e.g. Eigen 1971; Eigen & Schuster 1977). Thus, among a set of rival hypercycles, the already common is likely to win. This dynamics was claimed to have been important in the fixation of chirality and the genetic code (e.g. Küppers 1983). Yet this assumption is unwarranted (Szathmáry 1989*a*), briefly because: (i) parallel simple autocatalytic replication modifies invadability, (ii) stochastic effects allow uncommon, but intrinsically fitter hypercycles to invade and (iii) spatially distinct habitats would have allowed for diversity anyway. Thus, although hypercyclic systems may have played some role in prebiotic evolution, it is unlikely that their hyperbolic growth was very important (cf. Szathmáry *et al.* 1988).

Parabolic growth ensues when in the equation

$$\dot{x} = kx^p, \quad 0 < p < 1, \tag{4.5}$$

the solution of which is also given by equation (4.4). When $p = 1/2$, it is reduced to

$$x(t) = [kt/2 + x^{1/2}(0)]^2, \tag{4.6}$$

which is why this type of growth is called parabolic.

Parabolic growth entails survival of everybody in a competitive situation. To see this, consider the relative concentration of two parabolically growing replicators in the same environment

$$\frac{x_1(t)}{x_2(t)} = \frac{\left[\sqrt{x_1(0)} + k_1 t/2\right]^2}{\left[\sqrt{x_2(0)} + k_2 t/2\right]^2}, \tag{4.7}$$

and in the limit

$$\lim_{t\to\infty} \frac{x_1(t)}{x_2(t)} = \frac{k_1^2}{k_2^2}. \tag{4.8}$$

Thus, 'survival of everybody' (Szathmáry 1991*a*) is guaranteed, as shown by selection equations in Szathmáry & Gladkih (1989).

But what kind of molecular mechanism could underlie such an odd type of growth? von Kiedrowski (1986) and Zielinski & Orgel (1987) were the first to show that oligonucleotide analogues follow a square-root growth law in the appropriate medium. The reason, in a simplified form, is as follows. A template molecule A reacts with the source materials whereby a new copy of A is made, which remains associated with the template.

$$A + A \underset{b}{\overset{a}{\rightleftarrows}} AA,$$
$$A + X \xrightarrow{c} AA.$$

Crucial is the ordering of the rate constants $a \gg b > c$, i.e. association of two template molecules is faster than their dissociation, and replication *per se* is rate limiting. Note that the immediate product of copying is the replicationally inert AA complex. Thus, replication in this way is self-limiting. The higher the concentration of A, the stronger this self-limitation is. Note also that this type of replication is conservative: there is no material overlap between copy and template, and template and copy are exactly identical as well as complementary (this can be achieved by palindromes).

As it is apparent from the above reaction scheme, the rate of replication is determined by the concentration of free A, and at high enough total concentration of A (denoted by x) and AA (denoted by y), the former is negligible since association is stronger than dissociation. The formation and dissociation of AA are in quasi-equilibrium, thus

$$ax^2 \approx by, \quad x \approx \sqrt{by/a} \approx \rho\sqrt{z}, \quad z = x + y, \tag{4.9}$$

and therefore,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \dot{z} \approx kz^{1/2}, \tag{4.10}$$

which is formally identical with equation (4.5).

Owing to self-limitation based on molecular complementarity, AA and BB complexes (where A and B are two different replicators) are stronger than AB complexes. Hence, each species limits its own growth more strongly: this condition for joint survival is also found in traditional Lotka–Volterra competitive systems. This is the ultimate cause for survival of the common in parabolic systems (Szathmáry 1991*a*).

In the meantime, several more replicators obeying the same type of growth dynamics have been constructed among others by Rebek (1994) and Sievers & von Kiedrowski (1995). (In the latter case, the single-stranded templates are not self-complementary.) A detailed kinetic theory for parabolic growth of minimal replicators was worked out by von Kiedrowski (1993). It seems that parabolic growth is a rather robust phenomenon among these replicators, although with

the appropriate 'molecular gymnastics' nearly exponential growth can be achieved (Kindermann *et al.* 2005).

One of the important steps of prebiotic evolution must have been the emergence of replicators with exponential growth. Incidentally, this is very likely to have opened up the possibility of a transition from limited to unlimited heredity as well.

(b) *A nontrivial consequence of exponential decay*

Szathmáry & Gladkih (1989) realized that parabolic growth as expressed in equation (4.5) results in coexistence whenever replicators are in a competitive situation. The system they used was:

$$\dot{x}_i = k_i x_i^p - x_i \sum_j k_j x_j^p, \tag{4.11}$$

which implies a constraint of constant total population size (cf. Eigen 1971). The strange result of the analysis of this system was 'survival of everybody' (Szathmáry 1991) in contrast to the classical (Darwinian) case of exponential growth ($p=1$), where survival of the fittest prevails. This result was mathematically confirmed by Varga & Szathmáry (1997) who, by finding an appropriate Liapunov function, demonstrated that there was a single internal, globally stable rest point of the system (4.11).

Lifson & Lifson (1999) recently extended these findings by demonstrating that if single strands decompose by spontaneous (exponential) decay, coexistence is not possible any more and 'selection of the unfittest' sets in. Independently, von Kiedrowski (1998) announced that in a simulated chromatographic system of competing self-replicators natural selection could happen, despite the fact that this would not be possible in the spatially homogeneous case, modelled by equation (4.11).

Let us first point out that it is not the system (4.11) that the Lifsons modified. If you introduce decay rates into the model, you get

$$\dot{x}_i = k_i x_i^p - d_i x_i - x_i \sum (k_j x_j^p - d_j x_j), \tag{4.12}$$

for which survival of everybody is still guaranteed, despite the specific decay rates d_i. Using essentially the original rationale of Szathmáry & Gladkih (1989) one finds that

$$\dot{x}_i = x_i^p \left[k_i - x_i^{1-p} \left(d_i + \sum_j (k_j x_j^p - d_j x_i) \right) \right] > x_i^p \left(k_i - x_i^{1-p} k_{\max} \right), \tag{4.13}$$

wh4ich means that the time derivative is positive if the concentration x_i is sufficiently low (Scheuring & Szathmáry 2001).

In their model, the Lifsons assume that 'double strands do not replicate and are resistant to decomposition' (cf. their equations (3.2) and (4.15)). Their assumption that double strands do not decompose at all is unrealistic. In the following, I review results by von Kiedrowski & Szathmáry (2000) that competitive coexistence is still possible under a range of parameter values for self-replicators with a parabolic growth tendency, even if decay of strands is taken into account.

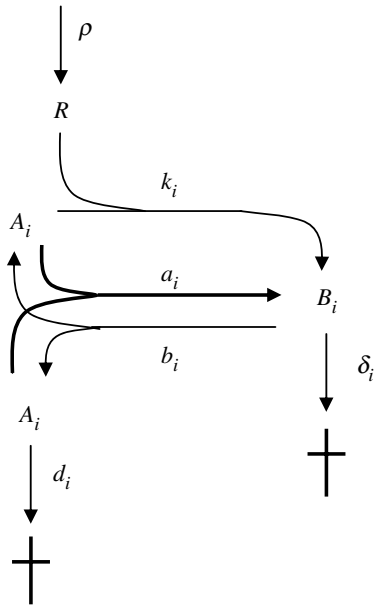


Figure 4. Stoichiometric scheme of the simplified system with differential decay rates for the double and single strands (von Kiedrowski & Szathmáry 2000). The resource R is fed into the system at a constant rate ρ. The assumption $d \gg \delta$ corresponds to that of the more complicated case when the double strand is retained much more strongly than the single strand by the chromatography column.

(c) *Theory before experiment: the chromatographized replicator model*

A common problem of non-enzymatic artificial replicator systems is product inhibition leading to parabolic instead of exponential amplification. Exponential chemical replication of oligonucleotides was achieved by an iterative stepwise procedure, which employs the surface of a solid support and was called Surface Promoted Replication and Exponential Amplification of DNA analogues (SPREAD; Luther *et al.* 1998). I review theoretical insights (von Kiedrowski & Szathmáry 2000) into the design of an autonomous variant of the SPREAD procedure. The corresponding program simulates a given set of chemical reactions coupled to a chromatographic process, where the chromatographic column is treated as a series of connected cells. The crucial step is a template-directed reaction occurring at the surface: thus it is assumed that two parabolic replicators compete for their building blocks in the chromatographic column. A simplified semi-analytic treatment confirms that competing parabolic replicators, which spread on mineral surfaces are amenable for Darwinian selection under a wide range of parameter values.

Now my aim is to demonstrate by a semi-analytically soluble simplified model that differential retention can lead to competitive exclusion (von Kiedrowski & Szathmáry 2000). Consider a single compartment with a constant nutrient (raw material) inflow and assume that single strands have a higher decay rate than double strands. This is meant to substitute for the higher retention of double strands on the chromatography column. The scheme of reactions is displayed in figure 4. For two species, we have the following

ordinary differential equation system:

$$\left.\begin{aligned} \mathrm{d}R/\mathrm{d}t &= \rho - R(k_1A_1 + k_2A_2) \\ \mathrm{d}A_i/\mathrm{d}t &= 2\left(b_iB_i - a_iA_i^2\right) - A_i(k_iR + d_i) \\ \mathrm{d}B_i/\mathrm{d}t &= a_iA_i^2 - b_iB_i + k_iRA_i - \delta_iB_i, \end{aligned}\right\}, \tag{4.14}$$

where R is the common resource and A_i, B_i are the single and double strands of species i, respectively ($i=1, 2$). We are interested in the conditions under which invasion by the inferior species when rare is *not* possible, i.e. we have competitive exclusion. A crucial relation is the following:

$$R > \frac{d_2}{k_2}. \tag{4.15}$$

Thus, when R_1 maintained by species 1 alone satisfies condition (4.15), invasion by species 2 is possible, otherwise it is impossible. Obviously, if A_2 is to invade, then the rate of its template ligation must be large and that of its decay must be small. A symmetric treatment applies to invasion by species 1 if species 2 is the resident one. The significant fact is that the threshold R_1 depends on the decay rates of the single strand (d_1) and the double strand (δ_1) of the resident species 1 as well.

Competitive exclusion (survival of the fittest) is compatible with

$$d \gg \delta, \tag{4.16}$$

but not the other way round. In the chromatographic case, this corresponds to a high retention factor for the double strand and low for the single strand. Note that an increase in δ easily throws the system into the region of coexistence.

I believe that the chromatographized replicator model is relevant to the origin of life on Earth. The chromatographic column is equivalent to a tunnel or a riverbed of minerals in which water containing the resources is continuously running through. Although our model, so far, refers to an isothermal reaction system, it can be easily extended to account for a gradient of increasing temperature along the direction of the column. As long as parabolic replicators need high temperatures whereas short replicators work at low temperatures (von Kiedrowski 1993), long replicators may grow from the consumption of shorter ones synthesized at the entry of the column where the temperature is low. The chromatographized replicator model can be simplified by means of attributing individual desorption rates to individual decay rates. Moreover, the findings from the simplified reaction model, viz. that both selection and coexistence can occur, has been independently confirmed by simulations based on the original model.

The case presented is an unusual one in that theory makes a clear prediction for experiment. Moreover, experimental realization of the model should be relatively straightforward.

5. REAL RIBOZYMES AND A RELAXED ERROR THRESHOLD

The error threshold—the critical copying fidelity below which the fittest genotype deterministically disappears—for replication limits the length of the genome that can be maintained by selection; see equation (2.3). Primordial replication must have been error-prone, so early replicators are thought to have been necessarily short (Eigen 1971). The error threshold also depends on the fitness landscape. In an RNA world (Gilbert 1986), there will be many neutral and compensatory mutations that can raise the threshold, below which the functional phenotype, rather than a particular sequence, is still present. A comparative analysis of two extensively mutagenized ribozymes has shown that with a copying fidelity of 0.999 per digit per replication, the phenotypic error threshold rises well above 7000 nucleotides, which permits the selective maintenance of a functionally rich ribo-organism with a genome over 100 different genes the size of a tRNA (Kun *et al.* 2005*a*,*b*). This 'only' requires an order of magnitude improvement in the accuracy of *in vitro* generated polymerase ribozymes (Johnston *et al.* 2001; Müller & Bartel 2003). Incidentally, this genome size coincides with that estimated for a minimal cell achieved by top-down analysis (comparative analysis of the genomes of reduced organisms: Gil *et al.* 2004) minus the genes dealing with translation.

Eigen's insight of an error threshold quantifies the problem. Following (2.3), we have

$$\nu < \frac{\ln s}{(1-q)}, \tag{5.1}$$

where $s=K/k$ is the so-called selective superiority of the fittest (master) sequence. In this simplified treatment, all mutants share the same replication rate, neutral mutations of and back mutations to the master are ignored.

The error threshold was first defined in relation to a particular genotype. However, it is obvious that in an RNA world there will be many neutral and compensatory mutations, which allow the preservation or the restoration of the fittest phenotype rather than of a single genotype. Other things being equal, this will modify the error threshold by increasing it (thus longer genomes will become maintainable). Since in an RNA world the functional ribozymes will have the strongest effect on fitness, one should gather the pertinent data from known ribozymes. As we shall see, there is just enough empirical evidence to formulate an encouraging statement.

To construct a fitness/functionality landscape of a ribozyme: (i) its secondary structure has to be experimentally determined, (ii) this secondary structure cannot contain a pseudo-knot, a special structural element that conventional RNA folding algorithms cannot satisfactorily cope with, (iii) mutagenesis experiments have to reveal all important sites and nucleotides and (iv) the size of the ribozyme should not be very long, otherwise any calculation would be practically unfeasible. The first requirement excludes most of the known ribozymes, since apart from the function only the sequence has been determined. The naturally occurring ribozymes generally fulfil the third requirement, but Hepatitis Delta Virus fails to meet the second requirement and Group I and II introns, as well as RNAase P,

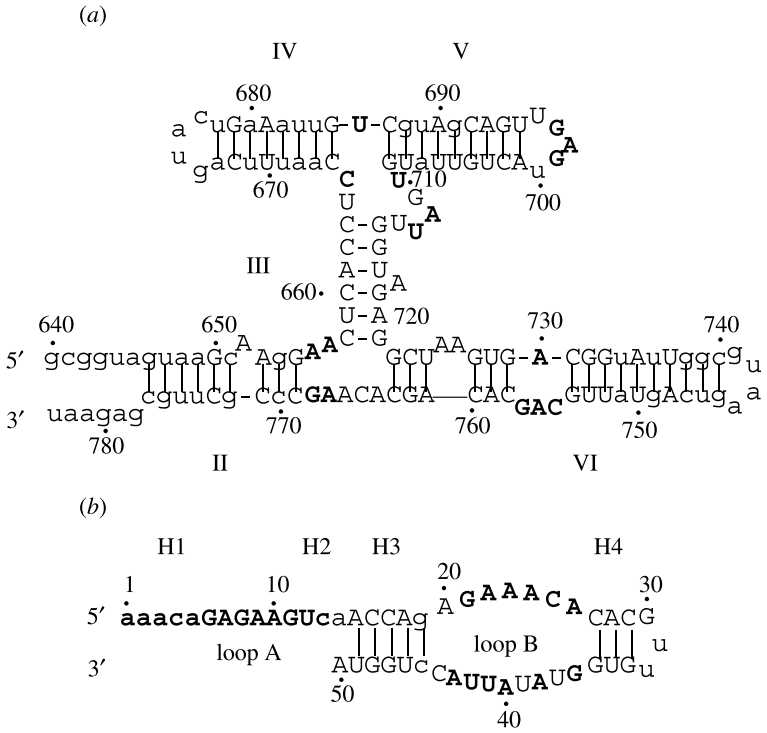


Figure 5. Secondary structures of (*a*) *Neurospora* VS ribozyme and (*b*) hairpin ribozyme indicating different regions (Kun *et al.* 2005*a*,*b*). Position numbering follows standard convention. Capitalized nucleotides specify those sites that have been subjected to mutagenesis experiments, and enzymatic activities of mutants are available. A total of 183 mutants for the VS ribozyme affecting 83 out of 144 positions, excluding insertions and deletions, were considered. For the hairpin ribozyme, the survey was based on 142 mutants affecting 39 out of 50 positions of the ribozyme and some part of the substrate region. Nucleotides marked in bold are the critical sites.

fail to meet the fourth. This leaves the hammerhead, the hairpin and the *Neurospora* VS ribozymes as possible candidates. Kun *et al.* (2005*a*) chose the hairpin and the *Neurospora* VS ribozymes for our study (figure 5). Both are relatively short, naturally occurring self-cleaving ribozymes, which can be divided into a *trans*-acting enzyme/substrate system where the *trans*-acting enzyme part does not contain a pseudo-knot.

The construction of the fitness/functionality landscape is based on four general observations: (i) the maintenance of the secondary structure is a major factor in retaining enzymatic activity, but the nature of most individual base pairs is not important and many can be reversed or replaced by a different pair without major loss of activity so long as a base pair is retained at a given position, (ii) the structure can have slight variations which in most cases manifest in some mismatch base pairs and/or some deletions or elongation in a helical region, (iii) there are critical regions in the molecule, where the nature of the base located there is also important and (iv) the effect of multiple mutations is multiplicative, i.e. the product of the activities of single mutants provides the activity of the multiple mutants.

From the fitness/functionality landscapes, the estimated phenotypic error thresholds are $\dot{\mu} = 0.0533$ and $\dot{\mu} = 0.144$ for the VS and hairpin ribozymes, respectively, where $\dot{\mu}$ is the effective mutation rate per nucleotide per replication. As expected, these figures are substantially higher than those inferred from fitness landscapes that do not take into account the secondary structure of the ribozymes but include information on single mutational effects.

This is the first time that the fitness landscape in terms of functionality has been inferred from real ribozymes (see also Kun *et al.* 2005*b*). The phenotypic error threshold thus inferred alleviates Eigen's paradox. This relates to the finding that the fitness landscapes are sufficiently similar. Inequality (5.1) cannot be used to assess the effect of the landscape on the error threshold owing to its restrictive preconditions. A recently derived expression (Takeuchi *et al.* 2005) offers a much more pertinent approximation:

$$\nu < \frac{-\ln s}{\ln(q + \lambda - q\lambda)}, \tag{5.2}$$

where λ is the fraction of neutral single substitutions. For the VS ribozyme $\nu = 144$, $q = 0.947$, $\lambda = 0.26$; and for the hairpin ribozyme $\nu = 50$, $q = 0.856$, $\lambda = 0.22$. Thus, for $\ln s$ we obtain 5.761 and 5.957, respectively.

The fitness values obtained allow us to reconsider Eigen's paradox. Although it was shown that within-gene recombination could raise the error threshold to some extent, it has been unknown until recently what would be the required accuracy of a sufficient replicase ribozyme in a ribo-organism. Substituting an accuracy of $q = 0.999$ in the lower bound of viral RNA replicases into inequality (5.2), and using the two obtained values for λ, we find that $\nu = 7000$–8000; namely, such a ribozyme could replicate a genome consisting of more than 100 different genes each of length 70 nucleotides or more than 70 different genes each of length 100. This would be sufficient to run a functionally rich ribo-organism, estimated to harbour about this number of genes (Jeffares *et al.* 1998). Incidentally, a recent

analysis of a core minimal bacterial gene set gives about 200 genes (Gil *et al.* 2004). This shows that if we take away the genes coding for the whole contemporary translation system, we are again in the same ballpark.

The artificial template-dependent RNA polymerase ribozyme selected by Johnston *et al.* (2001) has an average fidelity $q=0.97$. Using formula (5.2) and the fitness/functionality landscape obtained for the VS and hairpin ribozymes (an admitted leap), it was concluded that the accuracy of this ribozyme would allow the maintenance of replicators with length around 250, which means that this ribozyme could replicate itself if other conditions (such as processivity) were favourable. In order to eliminate the burden of Eigen's paradox, a replicase with an error rate of 10^{-3} per nucleotide per replication might have been sufficient to provide the minimal life requirements in the RNA world.

6. REPLICATOR EVOLUTION ON THE SURFACE

It is a common experience in theoretical ecology and evolutionary biology that population structure promotes coexistence and favours the spread of altruism. Importantly, theoretical investigations in the field of early evolution have paved the way for such investigations to a considerable extent. Without the aim of completeness, I survey some interesting relevant examples.

(a) *Metabolic ribozymes coexist on surfaces*

Imagine a non-hypercyclic, so-called 'metabolic' system (cf. figure 45 in Eigen & Schuster 1978). Undoubtedly, we are here comfortably in the RNA world: we assume that informational replication and selection for enzymatic function has already been achieved. The templates are assumed to contribute to metabolism via enzymatic aid; metabolic products are in turn used up by the templates for replication at different rates. Although all templates contribute to metabolism ('the common good'), they are able to use it with different efficiency. Thus in a spatially homogenous environment, competitive exclusion follows despite the metabolic coupling (Eigen & Schuster 1978).

Interesting selection dynamics occurs when molecules are bound to the surface without being washed away regularly. This problem was modelled by the use of 'cellular automata' (Czárán & Szathmáry 2000). Without becoming too technical, it suffices to say that each square of a grid is assumed to be occupied by a single molecule (template), or be empty. Templates can do two things: to replicate (put an offspring into a neighbouring empty cell if available) and hop away into empty sites nearby. Replication may depend on the composition of the few neighbouring cells. In the case of a hypercycle, for example, the template and a specimen of the preceding cycle member must be present in the same small area if replication of the former is to occur. This of course makes perfect chemical sense.

Boerlijst & Hogeweg (1991) simulated hypercycles on a surface exactly in this way. They found that rotating spirals on the surface appear, provided the hypercycle consists of more than four members. This is linked to the fact that such a hypercycle without population structure shows sustained oscillation in time. Each wing of a rotating spiral looks a bit like the arm of a galaxy, and is dominated by templates of the same membership in the hypercycle. Parasites are unable to kill the hypercycle in that system. This finding was attributed to the dynamics of spirals. Two questions emerge: Are spirals necessary? What happens if one models other systems in the same way (i.e. by cellular automata)?

The dynamics of the non-spatial version of the metabolic system looks as follows.

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i[k_i M(\boldsymbol{x}) - \Phi(\boldsymbol{x})], \tag{6.1}$$

where x_i stands for the concentrations of template I_i, and $\boldsymbol{x}$ is the vector of these concentrations. $M(\boldsymbol{x})$ is a multiplicative function of the concentrations of all the templates, and $\Phi(\boldsymbol{x})$ is an outflow term representing a selection constraint (constant total concentration). This formulation is formally identical to that given by Eigen & Schuster (1978) for a 'minimum model of primitive translation'. As they noted correctly, the fact that replication of any template is impossible without the presence of all the others does not prohibit the system from undergoing competitive exclusion: $M(\boldsymbol{x})$ is same in all the equations, hence the system essentially behaves as a collection of Malthusian competitors, whose dynamics are influenced by a common time-dependent factor.

It is assumed that the replicators I_i have dual functionality: as templates they are necessary for their own replication (autocatalysis), and as 'ribozymes' (RNAs able to act as enzymes) they contribute to metabolism producing the monomers.

Now we assume that replication takes place on the surface of a mineral (possibly pyrite) substrate. The replicator molecules themselves are of a finite size; therefore the number of replicators bound to a unit area of the substrate is constrained. We consider a two-dimensional square lattice of binding sites as the scene of the replication–diffusion process; each of the sites can harbour a single macromolecule at most. The lattice is toroidal (the opposite edges of the grid are merged in both dimensions) to avoid edge effects.

At $t=0$, half of the sites are occupied by n different types of macromolecules (we call n the system size). The replicator types are equally abundant in the initial pattern and individual molecules are randomly assigned to sites. The other half of the sites are empty initially. Time is discrete; replication, decay and diffusion take place in each generation of the simulation.

The effect of monomer-producing metabolism is implicit in the model, itself directly acting on the replication process through a *local metabolic function*. It is local in the sense that its arguments are the copy numbers $f(i)$ of replicator types i ($i=1, \ldots, n$) within certain localities (neighbourhoods) of the lattice. In accordance with the assumption that the presence of a complete set of replicators is necessary for metabolism to produce monomers for replication, the

metabolic function must be a multiplicative form of within-neighbourhood copy numbers $f(i)$. A simple option for the concrete form of the metabolic function $M(\boldsymbol{f}_s)$ at a site occupied by a replicator s is the geometric mean of the copy numbers $f_s(i)$ within the *metabolic neighbourhood* of s, i.e.

$$M(\boldsymbol{f}_s) = \left[\prod_{i=1}^{n} f_s(i)\right]^{1/n}. \qquad (6.2)$$

Note that $M(\boldsymbol{f}_s)$ is zero if any of the replicator types is missing from the metabolic neighbourhood of s, and that the larger and more uniform the copy numbers of the different replicator types within the metabolic neighbourhood, the more efficient the metabolism at the given locality. By choosing (6.2) as the metabolic function, we assume that the conspecific replicators within the same neighbourhood help replication and that the focal replicator supports its own replication. The first assumption can be interpreted as metabolism being somewhat faster locally in the presence of more catalysts. The actual effect should be rather weak and it should vanish with the copy number increasing; this feature is properly reflected in the metabolic function (6.2): if a replicator type is already present in a replication neighbourhood, then its successive copies do not add too much to the replication chance of the focal template. Implicit in the second assumption is that the time-scale of metabolite diffusion out of the neighbourhood in which it was produced is longer than that of the catalysed reactions of metabolism. The 'habitat' of the reaction-diffusion system being an absorptive mineral surface is again straightforward to assume. The size of the metabolically effective neighbourhood is an implicit measure of metabolite and monomer diffusivity: larger neighbourhoods represent faster diffusion of the intermediate metabolites and the monomers.

Czárán & Szathmáry (2000) managed to show that given such a spatial setting, non-hypercyclic systems are once again viable alternatives. The fundamental difference between their model and that of Boerlijst & Hogeweg (1991) is the following: the dynamical link among the replicators is realized through a common metabolism, instead of the direct, intransitive hypercyclic coupling. Using the cellular automaton model of the metabolic system, the aim was to show that

(i) metabolic coupling can lead to coexistence of replicators in spite of an inherent competitive tendency,
(ii) parasites cannot easily kill the whole system and
(iii) complexity can increase by natural selection.

The result that *there is coexistence without any conspicuous pattern* (i.e. something like spirals) is robust and counter-intuitive. It is owing to the inherent discreteness (i.e. the corpuscular nature of the replicator molecule populations) and spatial explicitness of the model, which grasp essential features of the living world in general, and macromolecular replicator systems in particular. An inferior (i.e. more slowly replicating) molecule type does not die out since there is an *advantage of rarity* in the system: a rare template is more likely to be complemented by a metabolically sufficient set of replicators in its neighbourhood than a common one.

(b) *Reciprocal altruism on the rocks and the evolution of replicases*

Although the question where the first RNA molecules came from is still unsolved, it is nevertheless assumed that catalytic RNA enzymes (ribozymes) with replicase function emerged at some stage of early evolution. Eigen's finding of the error threshold demonstrates that the length of templates maintained by selection is limited by the copying fidelity; therefore, other things being equal, an increase in template length is disadvantageous. On the contrary, longer molecules are expected to be better replicases—a feature not incorporated in the original model. An iterative scenario for longer and longer molecules with better and better replicase function has been suggested (James & Ellington 1999; Poole *et al.* 1999) and analysed mathematically (Scheuring 2000). A crucial open question is whether parasites (efficient templates that are inefficient replicases) can ruin the system. Absorption to mineral surfaces was hypothesized to help replicases find their useful colleagues in the immediate neighbourhood (Joyce & Orgel 1999). A cellular automaton simulation revealed that copying fidelity, replicase speed and template efficiency could increase by evolution, despite the presence of molecular parasites, essentially owing to reciprocal altruism on the surface, thus making the scenario for a gradual improvement of replicase function more plausible (Szabó *et al.* 2002).

Consider a population of macromolecules, adsorbed to a surface and built of four different monomers: A, B, C and D. Owing to their catalytic activity, macromolecules located on neighbouring sites of the surface can template-replicate each other, which means building a new macromolecule from free monomers by copying an existing one. In each replication process, two replicator molecules are involved: one is the template and the other acts as a replicase enzyme. We attribute two main properties to replication events, speed and fidelity, which in turn depend on three parameters of the two replicators involved in the process:

(i) *replicase activity* expresses how fast the molecule can add a monomer to a primer while acting as a replicase,
(ii) *replicase fidelity* measures the accuracy of replication per monomer when the molecule acts as a replicase and
(iii) *template efficiency* defines an average 'affinity' of the molecule behaving as a template against others.

The authors assumed that these traits are in a three-way tradeoff: there were no free lunches. Replication speed depends on the activity of the replicase and the quality of the template: higher replicase activity and template efficiency result in faster replication. Given two neighbouring replicator molecules, L and M, on the surface, one of the two different replication events can occur between them: either L as replicase copies

and M as a template, or the other way round. Mutations allowed not only point mutations but also additions and deletions of one nucleotide

The outcome was a bimodal population: efficient replicases evolved and short parasites could not ruin the system. This result, together with the chromatographized replicator model, emphasizes the importance of surface dynamics in prebiotic evolution. It also raises the idea that compartmentation offered by vesicles could have been an even more efficient means to evolve more efficient and accurate systems, a possibility to which I now turn.

7. BAGS OF GENES: THE STOCHASTIC CORRECTOR MODEL

It is true that the hypercyclic link ensures indefinite ecological survival of all member replicators. However, problems arise when mutations are taken into account. In order to consider them, it is worthwhile to look at a diagram where auto- and heterocatalytic aids are functionally clearly separate, such as in a hypercycle with protein replicases (figure 6). Mutants providing stronger heterocatalytic aid to the next member are not selected. In contrast, increased autocatalysis is always selected, irrespective of its concomitant effect on heterocatalytic efficiency. This is the well-known problem of parasites in the hypercycle (Maynard Smith 1979). As Eigen *et al.* (1981) observed, putting hypercycles into reproducing compartments helps, because 'good' hypercycles (with efficient heterocatalysis) can be favoured over 'bad' ones. The following two questions arise out of this:

(i) Are there other means whereby parasites can be selected against?
(ii) Are there non-hypercyclic systems that function well in a compartment context?

The answers turned out to be 'yes' to both of these questions; I discuss them below.

(a) *Group selection of early replicators*

The phase of evolution just outlined refers to the pre-cellular level. Later in evolution, protocells must have appeared. It turns out that cellularization offers the most natural, and at the same time most efficient, resolution to Eigen's paradox. It also leads to the appearance of linkage, i.e. the origin of chromosomes. The dynamics of genes encapsulated in a reproductive protocell is described by the *stochastic corrector* model (Szathmáry & Demeter 1987; Szathmáry 1989*a*,*b*; Grey *et al.* 1995; Zintzaras *et al.* 2002; Fontanari *et al.* 2006). It rests on the following assumptions (figure 7).

(i) Templates contribute to the fitness of the protocell as a whole and there is an optimal proportion of the genes. Concretely, we assume that the genes encode enzymatic aid given to the intracellular metabolism.
(ii) Templates compete with each other within the same protocell. As before, replication rates may differ from gene to gene.

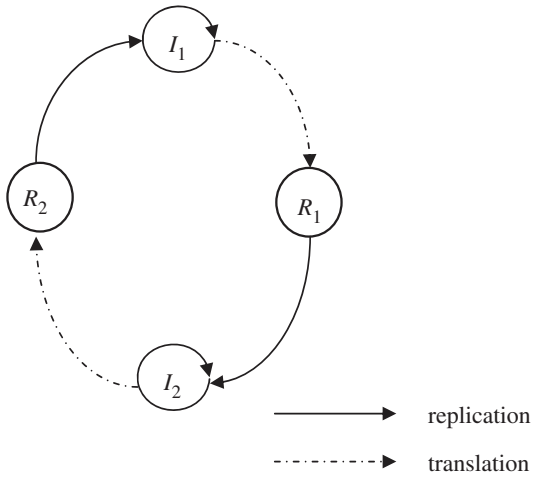


Figure 6. The hypercycle with translation. R_i is a replicase protein enzyme coded for by gene I_i.

(iii) Replication of templates is described by stochastic means. Since the number of genes in any compartment is small (up to a few hundred), their growth is affected by the luck of the draw. Ecologists would express this as demographic stochasticity.
(iv) There is no individual regulation of template copy number per protocell.
(v) Templates are assorted randomly into offspring cells upon protocell division.

I must emphasize that in the stochastic corrector model, the templates are not coupled to one another through a reflexive (intransitive) cycle of replicational aid, since it would be a hypercycle. Instead, we assume that they contribute to the 'common good' of the protocell by catalysing steps of its metabolism. Within each compartment, the templates are free to compete because they can reap the benefits of a common metabolism differently. (A similar situation can arise among chromosomes and plasmids in contemporary bacteria.) *Despite the fact that templates compete, the two sources of stochasticity generate between-cell variation in template copy number on which natural selection (between protocells) can act.* This is an efficient means of group selection of templates, since it is the protocells that are the groups obeying the stringent criteria: (i) there are many more groups than templates, (ii) each group has only one ancestor and (iii) there is no migration between the groups (cf. Leigh 1983). Grey *et al.* (1995) gave a fully rigorous re-examination of the stochastic corrector model. The two mentioned sources of stochasticity effectively lead to the *correction* of a malign within-protocell trend of harmful competition of the templates. It cannot be too strongly emphasized that the stochastic corrector is not, contrary to common misunderstanding, a hypercyclic system. *Hypercycles need compartments but compartments can live without hypercycles.* It is interesting to see that genuine group selection is likely to have aided a major transition from naked genes to protocells. Group structure is provided by the physical boundaries of cells.

Within the same context, the origin and establishment of chromosomes (linked genes) in the population have also been analysed (Maynard Smith &

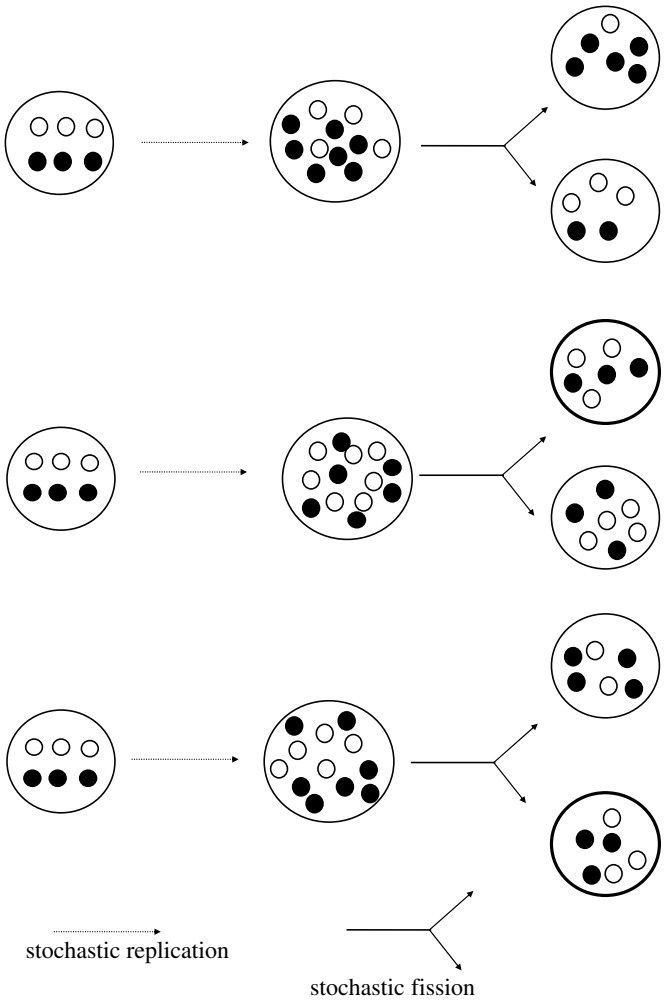


Figure 7. The stochastic corrector model. Different templates (labelled by open and closed circles) contribute to the well being of the compartments (protocells) in that they catalyse steps of metabolism, for example. During protocell growth, templates replicate at differential expected rates stochastically. Upon division, there is chance assortment of templates into offspring compartments. Stochastic replication and reassortment generate variation among protocells on which natural selection at the compartment level can act and oppose to (correct) internal deterioration owing to within-cell competition.

Szathmáry 1993). A chromosome consisting of two genes takes about twice as long to be replicated as the single genes. It turns out that chromosomes are strongly selected for at the cellular level even if they have this twofold within-cell disadvantage. Linkage reduces intracellular competition (genes are necessarily replicated simultaneously) as well as the risk of losing one gene by chance upon cell division (a gene is certain to find its complementing partner in the same offspring cell if it is linked to it). The molecular biology of the transition from genes to chromosomes has also been worked out (Szathmáry & Maynard Smith 1993).

(**b**) *Sex and protocells*

The results on coexistence leave one (one could say the original) question in the dark: does the error threshold increase or decrease in various systems? Although it was shown that the stochastic corrector model performs better than the compartmentalized hypercycle under a high error rate (Zintzaras *et al.* 2002), we still do not know the selectively maintainable genome size (or the number of different genes) in the stochastic corrector model. The results on real ribozymes (§5) alleviate, but do not solve, the problem. Lehman (2003) raised the issue that recombination, a frequently ignored player in models of early evolution, could have been crucial to build up primeval genomes of sizeable length. In the article that coined the phrase 'the RNA world', Gilbert (1986) already speculated that 'the RNA molecules evolve in self-replicating patterns, using recombination and mutation to explore new functions and to adapt to new niches'. In this context, Riley & Lehman (2003) have shown that *Tetrahymena* and *Azoarcus* ribozymes can promote RNA recombination.

This capability of RNA recombination to potentially reduce the burden imposed by the error threshold has been recently analysed by Santos *et al.* (2004). They assumed that the recombination in protocells took place via copy-choice means, i.e. the replicase switched between RNA-like templates, as occurs frequently in RNA viruses and is crucial for retroviral replication during reverse transcription. The numerical results showed that there is a quite intricate interplay between mutation, recombination and gene redundancy, but

the conclusion from the fitness function they used was that the informational content could have increased by 25% by keeping the same mutational load as that for a population without recombination.

The consequences of imperfect replication in vesicle models are puzzling. For small mutation rates, increased level of polyploidy favours the persistence of protocell lineages since the random loss of essential genes after fission is attenuated. However, for large mutation rates, the situation is reversed: those lineages with low levels of polyploidy are better able to cope with higher mutation rates, particularly when recombination is allowed. This means that gene redundancy was indeed costly. Therefore, selective forces favouring the linkage of genes to make the first chromosomes would eventually outweigh the advantage of faster replicating single genes, because linked genes are less likely to be lost by random assortment when protocells divide (Maynard Smith & Szathmáry 1993).

The role of the number of gene copies in a primitive cell was investigated by Koch (1984), who pointed out the existence of two conflicting forces: (i) higher copy numbers act as a safeguard against random loss of all copies of a gene but (ii) such copy numbers slow down adaptive evolution because a newly arisen favourable mutant is diluted out and cannot be 'seen' efficiently by natural selection acting on cells. He further observed that a moderately high (less than 100) copy number per gene is not only optimal, but also confers some additional evolvability by the 'duplication and divergence' scenario, as first emphasized by Ohno (1970).

This work was supported by the Hungarian Scientific Research Fund (OTKA T 047245) and the National Office for Research and Technology (NAP 2005/KCKHA005). Helpful comments by two anonymous referees are gratefully acknowledged.

REFERENCES

Bedau, M. A., McCaskill, J. S., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K. & Ray, T. S. 2000 Open problems in artificial life. *Artif. Life* **6**, 363–376. (doi:10.1162/106454600300103683)

Boerlijst, M. C. & Hogeweg, P. 1991 Spiral wave structure in pre-biotic evolution—hypercycles stable against parasites. *Physica* **D48**, 17–28.

Czárán, T. & Szathmáry, E. 2000 Coexistence of replicators in prebiotic evolution. In *The geometry of ecological interactions: simplifying spatial complexity* (ed. U. Dieckmann, R. Law & J. A. J. Metz), pp. 116–134. Wien, Austria: IIASA and Cambridge University Press.

Dawkins, R. 1976 *The selfish gene*. Oxford, UK: Oxford University Press.

Eigen, M. 1971 Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523. (doi:10.1007/BF00623322)

Eigen, M. & Schuster, P. 1977 The hypercycle: a principle of natural self-organization. Part A: emergence of the hypercycle. *Naturwissenschaften* **64**, 541–565. (doi:10.1007/BF00450633)

Eigen, M. & Schuster, P. 1978 The hypercycle: a principle of natural self-organization. Part C: the realistic hypercycle. *Naturwissenschaften* **65**, 341–369. (doi:10.1007/BF00439699)

Eigen, M., Schuster, P., Gardiner, W. & Winkler-Oswatitsch, R. 1981 The origin of genetic information. *Sci. Am.* **244**, 78–94.

Fernando, C., Santos, M. & Szathmáry, E. 2005 Evolutionary potential and requirements for minimal protocells. *Top. Curr. Chem.* **259**, 167–211.

Fontanari, J. F., Santos, M. & Szathmáry, E. 2006 Coexistence and error propagation in pre-biotic vesicle models: a group selection approach. *J. Theor. Biol.* **239**, 247–256. (doi:10.1016/j.jtbi.2005.08.039)

Gánti, T. 1971 *The principle of life (in Hungarian)*. Budapest, Hungary: Gondolat.

Gánti, T. 1978 *The principle of life (in Hungarian)*. Budapest, Hungary: Gondolat.

Gánti, T. 1979 *A theory of biochemical supersystems*. Baltimore, MD: University park press.

Gánti, T. 1987 *The principle of life*. Budapest, Hungary: OMIKK.

Gánti, T. 2003 *The principles of life*. Oxford, UK: Oxford University Press.

Gil, R., Silva, F. J., Peretó, J. & Moya, A. 2004 Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–537. (doi:10.1128/MMBR.68.3.518-537.2004)

Gilbert, W. 1986 The RNA world. *Nature* **319**, 818. (doi:10.1038/319618a0)

Grey, D., Hutson, V. & Szathmáry, E. 1995 A re-examination of the stochastic corrector model. *Proc. R. Soc. B* **262**, 29–35.

Griesemer, J. 2000 The units of evolutionary transition. *Selection* **1**, 67–80. (doi:10.1556/Select.1.2000.1-3.7)

Griesemer, J. 2002 What is "epi" about epigenetics? *Ann. NY Acad. Sci.* **981**, 97–110.

Hull, D. L. 1980 Individuality and selection. *Annu. Rev. Ecol. Syst.* **11**, 311–332. (doi:10.1146/annurev.es.11.110180.001523)

James, K. D. & Ellington, A. D. 1999 The fidelity of template-directed oligonucleotide ligation and the inevitability of polymerase function. *Orig. Life Evol. Biosph.* **29**, 375–390. (doi:10.1023/A:1006544611320)

Jeffares, D. C., Poole, A. M. & Penny, D. 1998 Relics from the RNA world. *J. Mol. Evol.* **46**, 18–36. (doi:10.1007/PL00006280)

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasen, M. E. & Bartel, D. P. 2001 RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325. (doi:10.1126/science.1060786)

Joyce, G. F. & Orgel, L. E. 1999 Prospects for understanding the origin of the RNA world. In *The RNA world* (ed. R. F. Gesteland, T. R. Cech & J. F. Atkins) 2nd edn., pp. 49–77. Plainview, NY: Cold Spring Harbor Lab. Press.

Kauffman, S. A. 1993 *The origins of order*. Oxford, UK: Oxford University Press.

von Kiedrowski, G. 1986 A self-replicating hexadeoxy nucleotide. *Angew. Chem. Int. Ed. Engl.* **25**, 932–935. (doi:10.1002/anie.198609322)

von Kiedrowski, G. 1993 Mimimal replicator theory I: parabolic versus exponential growth. *Bioorg. Chem. Frontiers* **3**, 113–146.

von Kiedrowski, G. 1998 120. *Versammlung Deutscher Ärtze und Nafurforscher*. Berlin, 21st September 1998.

von Kiedrowski, G. & Szathmáry, E. 2000 Selection versus coexistence of parabolic replicators spreading on surfaces. *Selection* **1**, 173–179. (doi:10.1556/Select.1.2000.1-3.17)

Kindermann, M., Stahl, I., Reimold, M., Pankau, W. M. & von Kiedrowski, G. 2005 Systems chemistry: kinetic and computational analysis of a nearly exponential organic replicator. *Angew. Chem.* **117**, 6908–6913. (doi:10.1002/ange.200501527) *Angew Chem. Int. Ed. Engl.* 44: 6750–6755.

King, G. A. M. 1982 Recycling, reproduction, and life's origin. *BioSystems* **15**, 89–97. (doi:10.1016/0303-2647(82)90022-3)

King, G. A. M. 1986 Was there a prebiotic soup? *J. Theor. Biol.* **123**, 493–498. (doi:10.1016/S0022-5193(86)80216-8)

Koch, A. L. 1984 Evolution vs the number of gene copies per primitive cell. *J. Mol. Evol.* **20**, 71–76. (doi:10.1007/BF02101988)

Kun, Á., Santos, M. & Szathmáry, E. 2005*a* Real ribozymes suggest a relaxed error threshold. *Nat. Genet.* **37**, 1008–1011. (doi:10.1038/ng1621)

Kun, A., Maurel, M.-C., Santos, M. & Szathmáry, E. 2005*b* Fitness landscapes, error thresholds, and cofactors in aptamer evolution. In *The Aptamer handbook* (ed. S. Klussmann), pp. 54–92. Weinheim, Germany: Wiley-Vch.

Küppers, B.-O. 1983 *Molecular theory of evolution*. Berlin, Germany: Springer.

Lehman, N. 2003 A case for the extreme antiquity of recombination. *J. Mol. Evol.* **56**, 770–777. (doi:10.1007/s00239-003-2454-1)

Leigh, E. G. 1983 When does the good of the group override the advantage of the individual? *Proc. Natl Acad. Sci. USA* **80**, 2985–2989. (doi:10.1073/pnas.80.10.2985)

Lifson, S. & Lifson, H. 1999 Models of prebiotic replication: survival of the fittest versus extinction of the unfittest. *J. Theor. Biol.* **199**, 425–433. (doi:10.1006/jtbi.1999.0969)

Luther, A., Brandsch, R. & von Kiedrowski, G. 1998 Surface-promoted replication and exponential amplification of DNA analogues. *Nature* **396**, 245–248. (doi:10.1038/24343)

Maynard Smith, J. 1979 Hypercycles and the origin of life. *Nature* **280**, 445–446. (doi:10.1038/280445a0)

Maynard Smith, J. 1983 Models of evolution. *Proc. R. Soc. B* **219**, 315–325.

Maynard Smith, J. 1986 *The problems of biology*. Oxford, UK: Oxford University Press.

Maynard Smith, J. & Szathmáry, E. 1993 The origin of chromosomes I. Selection for linkage. *J. Theor. Biol.* **164**, 437–446. (doi:10.1006/jtbi.1993.1165)

Maynard Smith, J. & Szathmáry, E. 1995 *The major transitions in evolution*. Oxford, UK: Freeman & Co.

Michod, R. 1983 Population biology of the first replicators: on the origin of genotype, phenotype, and organism. *Am. Zool.* **23**, 5–14.

Michod, R. 1984 Constraints on adaptation, with special reference to social behaviour. In *A new ecology* (ed. P. W. Price, C. N. Slobodchikoff & W. S. Gaud), pp. 253–278. New York, NY: Wiley.

Müller, U. F. & Bartel, D. P. 2003 Substrate 2′-hydroxyl groups required for ribozyme-catalyzed polymerization. *Chem. Biol.* **10**, 799–806. (doi:10.1016/S1074-5521(03)00171-6)

Ohno, S. 1970 *Evolution by gene duplication*. Berlin, Germany: Springer.

Oparin, A. I. 1961 *Life, its nature, origin and development*. New York, NY: Academic Press.

Orgel, L. E. 1992 Molecular replication. *Nature* **358**, 203–209. (doi:10.1038/358203a0)

Poole, A., Jeffares, D. & Penny, D. 1999 Early evolution: the new kids on the block. *BioEssays* **21**, 880. (doi:10.1002/(SICI)1521-1878(199910)21:10<880::AID-BIES11>3.0.CO;2-P)

Poon, A. & Chao, L. 2004 Drift increases the advantage of sex in RNA bacteriophage Phi6. *Genetics* **166**, 19–24. (doi:10.1534/genetics.166.1.19)

Rebek, J. 1994 Synthetic self-replicating molecules. *Sci. Am.* **271**, 34–40.

Riley, C. A. & Lehman, N. 2003 Generalized RNA-directed recombination of RNA. *Chem. Biol.* **10**, 1233–1243. (doi:10.1016/j.chembiol.2003.11.015)

Santos, M., Zintzaras, E. & Szathmáry, E. 2004 Recombination in primeval genomes: a step forward but still a long leap from maintaining a sizeable genome. *J. Mol. Evol.* **59**, 507–519. (doi:10.1007/s00239-004-2642-7)

Scheuring, I. 2000 Avoiding Catch-22 of early evolution by stepwise increase in copying fidelity. *Selection* **1**, 135–145. (doi:10.1556/Select.1.2000.1-3.13)

Scheuring, I. & Szathmáry, E. 2001 Survival of replicators with parabolic growth tendency and exponential decay. *J. Theor. Biol.* **212**, 99–105. (doi:10.1006/jtbi.2001.2360)

Segré, D., Lancet, D., Kedem, O. & Pilpel, Y. 1998 Graded autocatalysis replication domain (GARD): kinetic analysis of self-replication in mutually catalytic sets. *Orig. Life Evol. Biosph.* **28**, 501–514.

Segré, D., Ben-Eli, D., Deamer, D. W. & Lancet, D. 2001*a* The lipid world. *Orig. Life Evol. Biosph.* **31**, 119–145.

Segré, D., Shenhav, B., Kafri, R. & Lancet, D. 2001*b* The molecular roots of compositional inheritance. *J. Theor. Biol.* **213**, 481–491.

Shapiro, R. 1986 *A skeptic's guide to the creation of life on Earth*. New York, NY: Summit Books.

Sievers, D. & von Kiedrowski, G. 1995 Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224. (doi:10.1038/369221a0)

Szabó, P., Scheuring, I., Czárán, T. & Szathmáry, E. 2002 *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420**, 360–363.

Szathmáry, E. 1989*a* The emergence, maintenance, and transitions of the earliest evolutionary units. *Oxf. Surv. Evol. Biol.* **6**, 169–205.

Szathmáry, E. 1989*b* The integration of the earliest genetic information. *Trends Ecol. Evol.* **4**, 200–204. (doi:10.1016/0169-5347(89)90073-6)

Szathmáry, E. 1991 Simple growth laws and selection consequences. *Trends Ecol. Evol.* **6**, 366–370. (doi:10.1016/0169-5347(91)90228-P)

Szathmáry, E. 1995 A classification of replicators and lambda-calculus models of biological organization. *Proc. R. Soc. B* **260**, 279–286.

Szathmáry, E. 2000 The evolution of replicators. *Phil. Trans. R. Soc. B.* **355**, 1669–1676. (doi:10.1098/rstb.2000.0730)

Szathmáry, E. 2002 Units of evolution and units of life. In *Fundamentals of life* (ed. G. Pályi, L. Zucchi & L. Caglioti), pp. 181–195. Paris, France: Elsevier.

Szathmáry, E. 2005 Life: in search of the simplest cell. *Nature* **433**, 469–470. (doi:10.1038/433469a)

Szathmáry, E. & Demeter, L. 1987 Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128**, 463–486.

Szathmáry, E. & Gladkih, I. 1989 Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.* **138**, 55–58.

Szathmáry, E. & Maynard Smith, J. 1993 The origin of chromosomes II. Molecular mechanisms. *J. Theor. Biol.* **164**, 447–454. (doi:10.1006/jtbi.1993.1166)

Szathmáry, E. & Maynard Smith, J. 1997 From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* **187**, 555–571. (doi:10.1006/jtbi.1996.0389)

Szathmáry, E., Kotsis, M. & Scheuring, I. 1988 Limits of hyperbolic growth and selection in molecular and biological populations. In *Mathematical ecology* (ed. T. G. Hallam, L. Gross & S. Levin), pp. 46–68. Singapore, Singapore: World Scientific.

Szostak, J. W., Bartel, D. P. & Luisi, P. L. 2001 Synthesizing life. *Nature* **409**, 387–390. (doi:10.1038/35053176)

Takeuchi, N., Poorthuis, P. H. & Hogeweg, P. 2005 Phenotypic error threshold; additivity and epistasis in RNA evolution. *BMC Evol. Biol.* **5**, 9. (doi:10.1186/1471-2148-5-9)

Varga, Z. & Szathmáry, E. 1997 An extremum principle for parabolic competition. *Bull. Math. Biol.* **59**, 1145–1154. (doi:10.1016/S0092-8240(97)00048-7)

Wächtershäuser, G. 1992 Groundworks for an evolutionary biochemistry: the iron–sulphur world. *Prog. Biophys. Mol. Biol.* **58**, 85–201. (doi:10.1016/0079-6107(92)90022-X)

Zielinski, W. S. & Orgel, L. E. 1987 Autocatalytic synthesis of a tetranucleotide analogue. *Nature* **327**, 346–347. (doi:10.1038/327346a0)

Zintzaras, E., Mauro, S. & Szathmáry, E. 2002 Living under the challenge of information decay: the stochastic corrector model versus hypercycles. *J. Theor. Biol.* **217**, 167–181. (doi:10.1006/jtbi.2002.3026)

C. Kenneth Waters

Statement

and

Readings

Can Biology be Reduced to Chemistry and Physics?

I've been asked: can biology be reduced to chemistry and physics? Many philosophers have answered no on the basis of what's called the multiple realizability argument. I will use Elliott Sober's critique of this argument (included in the Symposium materials) as a point of departure. In my talk, I will suggest that we keep a number of distinctions in mind as we discuss the reduction question. The most basic of these is the distinction between metaphysics, on the one hand, and epistemology on the other. This gives rise to two kinds of questions. The metaphysical questions concern the relationship between the biological and the physical. Are, we might ask, organisms nothing but physical stuff organized in distinctive ways? The epistemological questions concern relations between the science of biology and the science of physics. Will, we might ask, biologists succeed in explaining all biological processes in terms of physical processes?

My talk will center on the connection between these two kinds of questions. Should we try to read metaphysics off the biology or biology off the metaphysics? That is, can we appeal to the best biological knowledge of the day to answer metaphysical questions such as the question of whether organisms are nothing but physical stuff organized in distinctive ways? Some argue along these lines by claiming that the remarkable success of DNA-centered research indicates that the fundamental theory of molecular biology, according to which all life processes are ultimately programmed in DNA and executed through the production of RNA and polypeptide molecules, is true. The metaphysical thesis of reductionism, they conclude, is vindicated.

Alternatively, we might think that we shouldn't read metaphysics off the science of biology, but appeal to the best metaphysics of the day to answer questions about the appropriate form of biological knowledge. Some argue on abstract considerations about notions such as emergence that today's DNA-centric biology is deeply problematic because it is based on the metaphysical falsehood of reductionism. According to this view, a more balanced research program that did not focus so much attention on genes and DNA would yield a truer, more holistic, and multi-leveled understanding of life.

I will argue that there are severe limitations in trying to read metaphysics off biology and dangers in drawing upon metaphysics to justify conclusions about the form biological knowledge should take. Hence, we should keep the two kinds of questions distinct. With regard to the epistemological questions about reduction, I will argue that the success of DNA-centered research depends as much on methodology as on representations, and the representations upon which it does rest are modest. There is no need to posit a fundamental biological theory of molecular biology to explain the success of DNA-centered sciences. I will leave the metaphysics and fundamental theorizing to others.

The Multiple Realizability Argument Against Reductionism*

Elliott Sober†‡
Department of Philosophy, University of Wisconsin-Madison

Reductionism is often understood to include two theses: (1) every singular occurrence that the special sciences can explain also can be explained by physics; (2) every law in a higher-level science can be explained by physics. These claims are widely supposed to have been refuted by the multiple realizability argument, formulated by Putnam (1967, 1975) and Fodor (1968, 1975). The present paper criticizes the argument and identifies a reductionistic thesis that follows from one of the argument's premises.

1. Introduction. If there is now a received view among philosophers of mind and philosophers of biology about reductionism, it is that reductionism is mistaken. And if there is now a received view as to why reductionism is wrong, it is the multiple realizability argument.[1] This

*Received March 1999.

†Send requests for reprints to the author, 5185 Helen C. White Hall, Department of Philosophy, University of Wisconsin, Madison, WI 53706.

‡My thanks to Martin Barrett, John Beatty, Tom Bontly, Ellery Eells, Berent Enç, Branden Fitelson, Jerry Fodor, Martha Gibson, Daniel Hausman, Dale Jamieson, Andrew Levine, Brian Mclaughlin, Terry Penner, Larry Shapiro, Chris Stephens, Richard Teng, Ken Waters, Ann Wolfe, and an anonymous referee for this journal for comments on earlier drafts.

1. Putnam (1967, 1975) and Fodor (1968, 1975) formulated this argument with an eye to demonstrating the irreducibility of psychology to physics. It has been criticized by Lewis (1969), Churchland (1982), Enç (1983), and Kim (1989), but on grounds distinct from the ones to be developed here. Their criticisms will be discussed briefly towards the end of this paper.

The multiple realizability argument was first explored in philosophy of biology by Rosenberg (1978, 1985), who gave it an unexpected twist; he argued that multiple realizability entails a kind of reductionism (both about the property of fitness and also about the relation of classical Mendelian genetics to molecular biology). In contrast, Sober (1984) and Kitcher (1984) basically followed the Putnam/Fodor line. The former work argues that the multiple realizability of fitness entails the irreducibility of theo-

Philosophy of Science, 66 (December 1999) pp. 542–564. 0031-8248/1999/6604-0003$2.00

argument takes as its target the following two claims, which form at least part of what reductionism asserts:

(1) Every singular occurrence that a higher-level science can explain also can be explained by a lower-level science.
(2) Every law in a higher-level science can be explained by laws in a lower-level science.

The "can" in these claims is supposed to mean "can in principle," not "can in practice." Science is not now complete; there is a lot that the physics of the present fails to tell us about societies, minds, and living things. However, a completed physics would not thus be limited, or so reductionism asserts (Oppenheim and Putnam 1958).

The distinction between higher and lower of course requires clarification, but it is meant to evoke a familiar hierarchical picture; it runs (top to bottom) as follows—the social sciences, individual psychology, biology, chemistry, and physics. Every society is composed of individuals who have minds; every individual with a mind is alive;[2] every individual who is alive is an individual in which chemical processes occur; and every system in which chemical processes occur is one in which physical processes occur. The domains of higher-level sciences are subsets of the domains of lower-level sciences. Since physics has the most inclusive domain, immaterial souls do not exist and neither do immaterial vital fluids. In addition, since the domains are (properly) nested, there will be phenomena that lower-level sciences can explain, but that higher-level sciences cannot. Propositions (1) and (2), coupled with the claim of nested domains, generate an asymmetry between higher-level and lower-level sciences.

Reductionism goes beyond what these two propositions express. Events have multiple causes. This means that two causal explanations of the same event may cite different causes. A car skids off the highway because it is raining, and also because the tires are bald (Hanson 1958).

---

retical generalizations about fitness; the latter argues for the irreducibility of classical Mendelian genetics to molecular biology. Waters (1990) challenges the specifics of Kitcher's argument; much of what he says is consonant with the more general criticisms of the multiple realizability argument to be developed here. Sober (1993) defends reductionism as a claim about singular occurrences, but denies that it is correct as a claim about higher-level laws.

2. If some computers (now or in the future) have minds, then the reducibility of psychology to biology may need to be revised (if the relevant computers are not "alive"); the obvious substitute is to have reductionism assert that psychology reduces to a physical science. Similarly, if some societies are made of mindless individuals (consider, for example, the case of the social insects), then perhaps the reduction will have to "skip a level" in this instance also.

Proposition (1) says only that if there is a psychological explanation of a given event, then there is also a physical explanation of that event. It does not say how those two explanations are related, but reductionism does. Societies are said to have their social properties *solely in virtue of* the psychological properties possessed by individuals; individuals have psychological properties *solely in virtue of* their having various biological properties; organisms have biological properties *solely in virtue of* the chemical processes that occur within them; and systems undergo chemical processes *solely in virtue of* the physical processes that occur therein. Reductionism is not just a claim about the explanatory capabilities of higher- and lower-level sciences; it is, in addition, a claim to the effect that the higher-level properties of a system are determined by its lower-level properties.[3]

These two parts of reductionism are illustrated in Figure 1. The circled *e* represents the relation of diachronic explanation; the circled *d* represents the relation of synchronic determination. Reductionism says that if (x) explains (y), then (z) explains (y); it also asserts that (z) determines (x). The multiple realizability argument against reductionism does not deny that higher-level properties are determined by lower-level properties. Rather, it aims to refute propositions (1) and (2)—(z) does not explain (y), or so this argument contends.

2. Multiple Realizability. Figure 2 is redrawn from the first chapter, entitled "Special Sciences," of Fodor's 1975 book, *The Language of Thought*. It describes a law in a higher-level science and how it might be related to a set of laws in some lower-level science. The higher-level

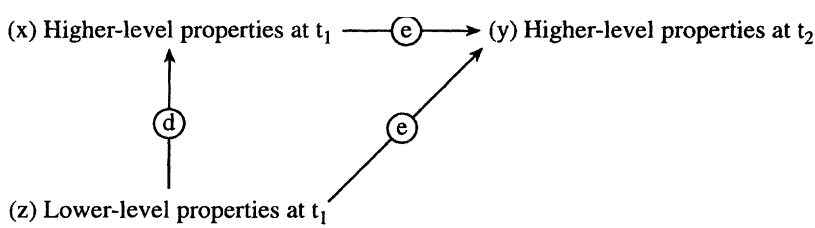


Figure 1. Relations of synchronic determination (*d*) and diachronic explanation (*e*) that may connect higher- and lower-level properties.

3. Reductionism should not be formulated so that it is committed to individualism of the sort discussed in philosophy of mind. For example, if wide theories of content are correct, then the beliefs that an individual has at a time depend not just on what is going on inside the skin of that individual at that time, but on what is going on in the individual's environment, then and earlier.

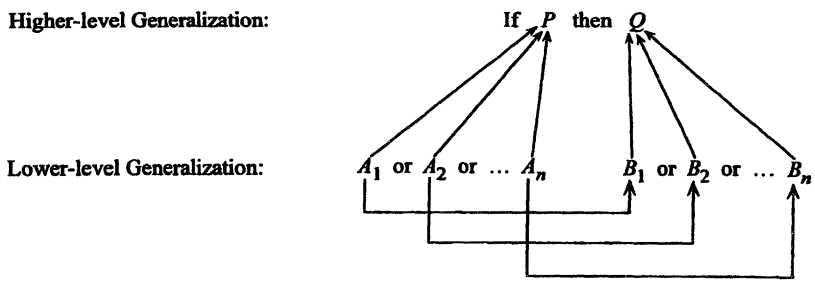


Figure 2. The lower-level properties A_i and B_j provide multiple realizations of the higher-level properties P and Q, respectively. One higher-level law and n lower-level laws are depicted, following Fodor 1975.

law is couched in its own proprietary vocabulary; P and Q are higher-level properties and the higher-level law says that everything that has P also has Q. The lower-level science provides n laws, each of them connecting an A predicate to a B predicate; the lower-level laws say that everything that has A_i also has B_i (for each $i = 1, 2, \ldots, n$).

The higher-level property P is said to be multiply realizable; A_1, A_2, . . . , A_n are the different (mutually exclusive and collectively exhaustive) realizations that P might have. Similarly, Q has B_1, B_2, . . . , B_n as its alternative realizations. What does multiple realizability mean? First, it entails the relation of simultaneous determination; necessarily, if something has A_i at time t, then it has P at t, and if it has B_i at time t, then it has Q at t. But there is something more, and it is this second ingredient that is supposed to ensure that the multiple realizability relation is anti-symmetric. An individual that has P has that property *solely in virtue* of the fact that it has whichever A_i it possesses. Because the higher-level properties are multiply realizable, the mapping from lower to higher is many-to-one. You cannot tell which of the A_i properties is exhibited by a system just from knowing that it has property P, and you cannot tell which of the B_j properties the system has just from knowing that it has Q.[4]

Two examples will make the intended meaning of multiple realizability sufficiently clear. Suppose that different types of physical system can have minds; minds can be built out of neurons, but perhaps they also can be built out of silicon chips. An individual mind—you, for

4. Although multiple realizability induces an asymmetry between P and each A_i, it does not entail that there is an asymmetry between P and the disjunctive property (A_1 or A_2 or . . . or A_n). Fodor would say that this disjunctive predicate fails to pick out a natural kind, a point that will be discussed later.

example—will have its psychological properties in virtue of the physical properties that the system possesses. But if you and someone else have some psychological properties in common, there is no guarantee that the two of you also will share physical properties; you and this other person may deploy different physical realizations of the same psychological properties. The same point can be made with respect to biological properties—you have various biological properties, and each of these is present in virtue of your possessing this or that set of physical properties. However, you and some other organism may share a given biological property even though you are physically quite different; this will be true if you and this other organism deploy different physical realizations of the same biological properties.

Since the multiple realizability relation obtains between simultaneously instantiated properties, the relation is not causal (assuming as I will that cause must precede effect). However, the diachronic laws I want to consider *are* causal—they say that a system's having one property at one time causes it to exhibit another property sometime later. The reason I will focus on causal diachronic laws is not that I think that all diachronic laws are causal, but that these provide the clearest cases of scientific explanations.[5] Thus, returning to propositions (1) and (2), we can ask the following two questions about the multiple realizability relations depicted in the second figure:

(1′) If an individual's having property P explains its having property Q, is it also true that its having property A_i explains its having property Q?

(2′) Do lower-level laws of the form "if A_i then B_i" explain the higher-level law "if P then Q"?

Let us assume that the properties described in higher-level sciences are multiply realized by properties discussed in a lower-level science. What consequences follow from this concerning reductionism?

3. The Explanation of Singular Occurrences—Putnam's Peg. Suppose a wooden board has two holes in it. One is circular and has a 1-inch diameter; the other is square and is 1 inch on a side. A cubical peg that is 15/16ths of an inch on each side will fit through the square hole, but not the circular one. What is the explanation? Putnam (1975) says that the explanation is provided by the *macro*-properties just cited of the peg and the holes. He denies that the *micro*-properties of molecules or atoms or particles in the peg and the piece of wood explain this fact.

5. Here I waive the question of whether *all* explanations are causal explanations, on which see Sober 1983 and Lewis 1986.

The micro-description is long and complicated and it brings in a welter of irrelevant detail. To explain why the peg goes through one hole but not the other, it does not matter what micro-properties the molecules have, as long as the peg and board have the macro-properties I mentioned. The macro-properties are explanatory; the micro-properties that realize those macro-properties are not. Hence, reductionism is false.

This is a delightfully simple example and argument, but it is possible to have one's intuitions run in the opposite direction. Perhaps the micro-details do not interest *Putnam*, but they may interest *others*, and for perfectly legitimate reasons. Explanations come with different levels of detail. When someone tells you more than you want to hear, this does not mean that what is said fails to be an explanation. There is a difference between explaining too much and not explaining at all.

Compare the micro-story that Putnam derides with a quite different story. Suppose someone suggested that the reason the peg goes through one hole but not the other is that the peg is *green*. Here it is obvious that a mistake has been made. If we demand that explanations be *causal* explanations, it will be quite clear why the color of the peg is not explanatory. It is causally irrelevant. This is an objective feature of the system under consideration and has nothing to do with our desire for brevity or detail.

It is possible to be misled by a superficial similarity that links the micro-story about the particles in the peg and board and the pseudo-explanation that cites the peg's color. Both of the following counterfactuals are true:

> If the particles in the peg and board had been different, the peg still would have passed through one hole but not the other, as long as the macro-dimensions were as described.
>
> If the peg had not been green, it still would have passed through one hole but not the other, as long as the macro-dimensions were as described.

If we say that causes are *necessary* for their effects (as does Lewis 1973a), we might be tempted to use these counterfactuals to conclude that the system's micro-features and the peg's color are both causally irrelevant, and hence should not be cited in a causal explanation. This proposal should be understood to mean that the effect would not have happened if the cause had not, *in the specific circumstances that actually obtained*; striking a match is not always necessary to get the match to light, but it may be necessary in various specific circumstances.

There are general questions that may be raised about the adequacy of this account of causation.[6] However, even if we waive these ques-

6. I will mention two. The first concerns how this theory of causation analyzes putative

tions, it is important to examine more closely how the counterfactual test connects with Putnam's argument. Let us suppose that the micro-properties of the peg and board's molecules are not necessary for the peg to go through one hole but not the other, if we hold fixed the macro-dimensions. But are the macro-dimensions necessary, if we hold fixed the micro-properties? That is, are we prepared to affirm the following counterfactual?

> If the macro-dimensions of the peg and board had been different, while the micro-properties were as described, the peg would not have passed through the one hole but not the other.

This counterfactual has a nomologically impossible antecedent. Many of us simply draw a blank when asked to assign a truth value to such assertions. The semantics of Stalnaker (1968) and Lewis (1973b) does not; it says that the counterfactual is vacuously true. However, before we interpret this as vindicating Putnam's argument, we also should note that the same semantic theory says that the following counterfactual is true as well:

> If the macro-dimensions of the peg and board had been different, while the micro-properties were as described, the peg still would have passed through the one hole but not the other.

It is hard to see how such counterfactuals can vindicate the judgment that the macro-properties are causally efficacious while their micro-realizations are not.[7]

I very much doubt that the concept of explanatory relevance means what Putnam requires it to mean in this argument. When scientists discover why smoking causes cancer, they are finding out which ingredients in cigarette smoke are carcinogenic. If smoking causes cancer, this is presumably because the micro-configuration of cigarette smoke is doing the work. If there turn out to be several carcinogenic ingredients and different cigarettes contain different ones, this does not make the molecular inquiry explanatorily irrelevant to the question of why people get cancer. The fact that *P* is multiply realizable does not mean that *P*'s realizations fail to explain the singular occurrences that

---

cases of overdetermination by multiple actual causes. Suppose Holmes and Watson each simultaneously shoot Moriarty through the heart. The theory entails that Holmes did not cause Moriarty's death, and Watson did not either. Rather, the cause is said to be disjunctive—Holmes shot him or Watson did. The second question comes from thinking about the possibility of indeterministic causation. Just as the totality of the antecedent causal facts need not suffice for the effect to occur, so the effect could have happened even if the causes had been different.

7. I am grateful to Brian Mclaughlin for drawing my attention to this line of argument.

P explains. A smoker may not want to hear the gory details, but that does not mean that the details are not explanatory.[8]

Putnam says he does not care whether we call the micro-story about the peg and the board a non-explanation, or simply describe it as a "terrible" explanation (Putnam 1975, 296). He thinks that the "goodness" of an explanation "is not a subjective matter." According to the objective concept of good explanation that Putnam has in mind, "an explanation is superior if it is more general" and he quotes with approval a remark by Alan Garfinkel—that "a good explanation is invariant under small perturbations of the assumptions" (301). What makes a more general (more invariant) explanation *objectively* better than one that is less? Putnam's answer is that "one of the things we do in science is to look for laws. Explanation is superior not just subjectively, but *methodologically*, in terms of facilitating the aims of scientific inquiry, if it brings out relevant laws" (301). My reply is that the goal of finding "relevant" laws cuts both ways. Macro-generalizations may be laws, but there also may be laws that relate micro-realizations to each other, and laws that relate micro- to macro- as well. Although "if *P* then *Q*" is more general than "if A_i then B_i ,"[9] the virtue of the

8. It is worth considering a curious remark that Putnam makes in a footnote before he introduces the example of the peg and board. He says:

> Even if it were not physically possible to realize human psychology in a creature made of anything but the usual protoplasm, DNA, etc., it would still not be correct to say that psychological states are identical with their physical realizations. For, as will be argued below, such an identification has no *explanatory* value in *psychology*. (1975, 293)

He then adds the remark: "on this point, compare Fodor, 1968," presumably because Fodor thought that antireductionism depends on higher-level properties being *multiply* realizable.

If we take Putnam's remark seriously, we must conclude that he thinks that the virtue of higher-level explanations does not reside in their greater generality. If a higher-level predicate (*P*) has just one possible physical realization (A_1), then *P* and A_1 apply to exactly the same objects. Putnam presumably would say that citing A_1 in an explanation provides extraneous information, whereas citing *P* does not. It is unclear how this concept of explanatory relevance might be explicated. In any event, I have not taken this footnote into account in describing the "multiple realizability argument," since Putnam's point here seems to be that *multiple* realizability does not bear on the claims he is advancing about explanation. This is not how the Putnam/Fodor argument has been understood by most philosophers.

9. I grant this point for the sake of argument, but it bears looking at more closely. Intuitively, "if *P* then *Q*" is more general than "if A_i then B_i" because the extension of *P* properly contains the extension of A_i. However, each of these conditionals is logically equivalent with its contrapositive, and it is equally true that the extension of not-B_i properly contains the extension of not-*Q*. This point is not a mere logical trick, to be swept aside by saying that the "right" formulation of a law is one that uses predicates that name natural kinds. After all, some laws (specifically, zero force laws) are typically

micro-generalization is that it provides more details. Science aims for depth as well as breadth. Some good explanations are fox-like; others are hedgehogian (Berlin 1953). There is no objective rule concerning which is better.

The claim that the preference for breadth over depth is a matter of taste is consistent with the idea that the difference between a genuine explanation and a nonexplanation is perfectly objective. In fact, it also is consistent with Hempel's (1965) view that the concept of scientific explanation should be explicated in terms of the notion of an ideally complete explanation, and that this is an objective notion. Perhaps an ideally complete scientific explanation of a singular occurrence in which an individual (or set of individuals) exhibits a multiply realizable property (or relation) would include the macro-story, the micro-story, and an account of how these are connected. If this is right, then reductionists and antireductionists alike are mistaken if they think that only part of this multilevel account deserves mention. But whatever the merits are of the idea of an ideally complete scientific explanation, we need to recognize that science in its currently incomplete state still is able to offer up "explanations." Perhaps these should be termed "explanation sketches," since they fall short of the Hempelian ideal. In any case, it remains true that science provides a plurality of such accounts of a given event. They vary in how detailed they are and in the level of organization described.[10]

Returning to Putnam's example, let us imagine that we face *two* peg-plus-board systems of the type that he describes. If we opt for the macro-explanation of why, in each case, the peg goes through one hole but not the other, we will have provided a *unified explanation.* We will have

---

stated as conditionals but their applications usually involve the predicates that occur in the contrapositive formulation. For example, the Hardy-Weinberg Law in population genetics describes how gamete frequencies will be related to genotype frequencies when no evolutionary forces are at work; its typical applications involve noting a departure from Hardy-Weinberg genotype frequencies, with the conclusion being drawn that some evolutionary forces are at work (Sober 1984). To say that the Hardy-Weinberg law has zero generality because every population is subject to evolutionary forces is to ignore the standard way in which the law is applied, and applied frequently, to nature.

10. Putnam's argument also has implications about the explanatory point of citing distal and proximate causes of a given effect. Imagine a causal chain from C_d to C_p to E. Suppose that C_d suffices for the occurrence of C_p, but is not necessary, and that the only connection of C_d to E is through C_p. Then Putnam's argument apparently entails that C_p explains E, and that C_d is either not an explanation of E, or is a terrible explanation of that event. But surely there can be an explanatory point to tracing an effect more deeply into the past. And surely it does not automatically increase explanatory power to describe more and more proximate causes of an effect.

explained similar effects by describing similar causes. However, if we choose a micro-explanation, it is almost inevitable that we will describe the two systems as being physically different, and thus our explanation will be *dis*unified. We will have explained the similar effects by tracing them back to different types of cause. Putnam uses the terms "general" and "invariant" to extol the advantages of macro-explanation, but he might just as well have used the term "unified" instead. In claiming that it is a matter of taste whether we prefer the macro- or the micro-explanation, I am claiming that there is no objective reason to prefer the unified over the disunified explanation. Science has room for both lumpers and splitters. Some people may not be interested in hearing that the two systems are in fact different; the fact that they have the same macro-properties may be all they wish to learn. But this does not show that discerning differences is less explanatory. Indeed, many scientists would find it more illuminating to be shown how the same effect is reached by different causal pathways.

In saying that the preference for unified explanation is merely a matter of taste, I seem to be contradicting a fundamental fact about scientific inference—that it counts in favor of the plausibility of a theory that the theory unifies disparate phenomena. Actually, no such consequence follows from what I am saying. Here, it is essential to distinguish the *context of justification* from the *context of explanation*.[11] When two theories are evaluated in the light of the evidence available, the fact that one is unified and the other is disunified is epistemologically relevant. In a wide range of circumstances, the unified theory can be expected to be more predictively accurate than the theory that is disunified, when they fit the data about equally well (Forster and Sober 1994). Whether a theory is unified is relevant to deciding whether we should accept it. However, the problem addressed by the multiple realizability argument is not about acceptance. We are supposed to assume that the macro-story and the micro-story are both *true*. Given this, we now are asked to decide which provides the better explanation of why the systems behave similarly. Unification is relevant to acceptance, but unification is not objectively relevant to deciding which accepted statements we should use in formulating explanations. The latter is simply a matter of taste—do we want more details or fewer? The context of justification and the context of explanation are different.

11. The distinction between justification and explanation was clearly drawn by Hempel (1965), who points out that why-questions can be requests for evidence or requests for explanation. This distinction supplements the familiar logical empiricist distinction between the *context of discovery* and the *context of justification*.

4. The Explanation of Laws—Fodor's Horror of Disjunctions. Whereas Putnam discusses the explanation of singular occurrences, Fodor uses the idea of multiple realizability to argue that laws in a higher-level science are not explained by laws in a lower-level science. This shift introduces some new considerations. Although many, if not all, explanations of singular occurrences are causal, the most familiar cases of explaining laws do not involve tracing effects back to their causes. Laws are usually explained by deriving them from "deeper" laws and initial condition statements; the explained laws and the explaining laws are true at the same time, so it is hard to think of the one as causing the other.

To understand Fodor's antireductionist position, let us consider the following derivation of a higher-level law:

If A_i then B_i (for each $i = 1, 2, \ldots, n$).
If A_1 or A_2 or . . . or A_n , then B_1 or B_2 or . . . or B_n.
P iff A_1 or A_2 or . . . or A_n.
Q iff B_1 or B_2 or . . . or B_n.

If P then Q.

The first premise describes a set of lower-level laws; the second premise follows from the first. The third and fourth premises state bridge principles that connect a property discussed in a higher-level science with its multiple, lower-level, realizations. By assumption, the premises are true and the conclusion follows from the premises. Why, then, is this derivation not an explanation of the higher-level law?

Fodor's answer is not that the premises involve concepts that come from the higher-level science. Given that the higher-level science and the lower-level science use different vocabularies, any derivation of the one from the other must include bridge principles that bring those different vocabularies into contact (Nagel 1961). Rather, Fodor's reason is that laws cannot be disjunctive. Although he grants that each statement of the form "if A_i then B_i " is a law, he denies that the second premise expresses a law. For the same reason, the third and fourth premises also fail to express laws. To reduce a law, one must explain why the proposition is not just true, but is a law; this is supposed to mean that one must derive it solely from lawful propositions. This is why Fodor thinks that multiple realizability defeats reductionism.

Even if laws cannot be disjunctive, why does the above derivation fail to explain why "if P then Q" is a law? After all, the conclusion will be nomologically necessary if the premises are, and Fodor does not deny that the premises are necessary. Are we really prepared to say

that the truth and lawfulness of the higher-level generalization is *inexplicable*, just because the above derivation is peppered with the word "or"? I confess that I feel my sense of incomprehension and mystery palpably subside when I contemplate this derivation. Where am I going wrong?

It also is not clear that laws must be nondisjunctive, nor is it clear what this requirement really amounts to. Take a law that specifies a quantitative threshold for some effect—for example, the law that water at a certain pressure will boil if the ambient temperature exceeds 100°C. This law seems to be disjunctive—it says that water will boil at 101°C, at 102°C, and so on. Of course, we have a handy shorthand for summarizing these disjuncts; we just say that any temperature "above 100°C" will produce boiling water. But if this strategy suffices to render the law about water nondisjunctive, why can't we introduce the letter α to represent the disjunction "A_1 or A_2 or . . . or A_n" and β to represent the disjunction "B_1 or B_2 or . . . or B_n"? It may be replied that the different disjuncts in the law about water all bring about boiling by the same type of physical process, whereas the different physical realizations A_i that the higher-level property P might have are heterogeneous in the way they bring about the B_i's that are realizations of Q.[12] The point is correct, but it remains unclear why this shows that laws cannot be disjunctive.

Disjunctiveness makes sense when it is understood as a *syntactic* feature of sentences. However, what does it mean for a proposition to be disjunctive, given that the same proposition can be expressed by different sentences? The problem may be illustrated by way of a familiar example. Suppose that the sentence "every emerald is green" and the sentence "every emerald is grue and the time is before the year 2000, or every emerald is bleen and the time is after the year 2000" are equivalent by virtue of the definitions of the terms "grue" and "bleen" (Goodman 1965). If laws are language-independent propositions of a certain type, and if logically equivalent sentences pick out the same proposition, then both sentences express laws, or neither does. Nothing changes if green is a natural kind whereas grue and bleen are not.

Although Fodor (1975) does not mention grue and bleen, it is fairly clear that his thinking about natural kinds—and his horror of disjunctions—both trace back to that issue.[13] Goodman (1965) held that law-

12. Fodor (1998, 16) says that a disjunction may occur in a bridge law if and only if the disjunction is "independently certified," meaning that "it also occurs in laws at its own level." The disjunction in the law about boiling presumably passes this test.

13. See, for example, Davidson's (1966) discussion of "all emeroses are gred" and also Davidson 1970.

like generalizations are confirmed by their positive instances, whereas accidental generalizations are not. The statement "all emeralds are green" is supposed to be lawlike, and hence instance confirmable, in virtue of the fact that "emerald" and "green" name natural kinds (or are "projectible"); "all emeralds are grue," on the other hand, is supposed to be non-lawlike, and so not confirmable by its instances, because it uses the weird predicate "grue." However, subsequent work on the confirmation relation has thrown considerable doubt on the idea that all and only the lawlike statements are instance confirmable (see, e.g., Sober 1988).

If P and (A_1 or A_2 or . . . or A_n) are known to be nomologically equivalent, then any probabilistic model of confirmation that takes that knowledge into account will treat them as *confirmationally* equivalent. For example, if a body of evidence confirms the hypothesis that a given individual has P, then that evidence also confirms the hypothesis that the individual has (A_1 or A_2 or . . . or A_n). This is a feature, for example, of Bayesian theories of confirmation (on which, e.g., see Howson and Urbach 1989 and Earman 1992). Disjunctiveness has no special meaning within that framework.

Fodor (1975, 21) concedes that the claim that laws must be nondisjunctive is "not strictly mandatory," but then points out that "one denies it at a price." The price is that one loses the connection between a sentence's expressing a law and the sentence's containing kind predicates. "One thus inherits the need for an alternative construal of the notion of a kind"; I am with Fodor when he says that he does not "know what that alternative would be like" (22). Fodor is right here, but his argument is prudential, not evidential. Like Pascal, Fodor is pointing out the disutility of denying a certain proposition, but this is not to show that the proposition is true.

The multiple realizability argument against the reducibility of laws is sometimes formulated by saying that the disjunctions that enumerate the possible realizations of P and Q are "open-ended." This would defeat the derivation described above—the third and fourth premises would be false—but it is important to see that the rules of the game now have changed. The mere fact that P and Q are multiply realizable would no longer be doing the work. And if the point about "open-endedness" is merely epistemological (we now do not *know* all of the physical realizations that P and Q have), it is irrelevant to the claim that higher-level sciences are reducible *in principle*.[14]

14. Moreover, the multiple realizability argument is not needed to show that the thesis of *reducibility in practice* is false; one can simply inspect present-day science to see this.

5. Probabilistic Explanations. The multiple realizability argument is usually developed by considering deterministic laws. However, laws in many sciences are probabilistic. How would the argument be affected by assuming that P and Q are probabilistically related, and that the A_i and the B_i are too?

Suppose that A_1 and A_2 are the only two possible realizations that P can have, and that B_1 and B_2 are the only two realizations that Q can have (the points I'll make also hold for $n > 2$). Suppose further that the probabilistic law connecting P to Q has the form

$$\Pr(Q \mid P) = p.$$

Then it follows that

$$p = \Pr(Q \mid P) = \Pr(Q \mid A_1)\Pr(A_1 \mid P) + \Pr(Q \mid A_2)\Pr(A_2 \mid P).$$

If we substitute $p_1 = \Pr(Q \mid A_1)$ and $p_2 = \Pr(Q \mid A_2)$ into this expression, we obtain

$$p = (p_1)\Pr(A_1 \mid P) + (p_2)\Pr(A_2 \mid P).$$

The probability (p) described in the higher-level law is a *weighted average* of the two probabilities p_1 and p_2; the weighting is determined simply by how often systems with P happen to deploy one micro-realization rather than the other.

It is not inevitable that $p = p_1 = p_2$. For example, suppose that smoking (P) makes lung cancer (Q) highly probable and that cigarette smoke always contains one of two carcinogenic ingredients (A_1 or A_2), which are found only in cigarette smoke. It can easily turn out that one of these ingredients is more carcinogenic than the other.[15] This means that there can be an important difference between higher-level and lower-level explanations of the same event—they may differ in terms of the probabilities that *explanans* confers on *explanandum*. To see why, let us add one more detail to the example. Suppose that lung cancer can be realized by one of two types of tumor (B_1 or B_2) growing in the lungs. Given this, consider an individual who has lung cancer. How are we to explain why this person has that disease? One possible reply is to say that the person smoked cigarettes. A second possibility is to say that the cancer occurred because the person inhaled ingredient A_1. Putnam's multiple realizability argument entails that the second suggestion is either no explanation at all, or is a "terrible" explanation. I suggest, however, that it should be clear to the unjaundiced eye that the second explanation may have its virtues. Perhaps A_1 confers on lung

15. If laws must be time-translationally invariant, then it is doubtful that "$\Pr(Q \mid P) = p$" expresses a law, if P is multiply realizable (Sober 1999).

cancer a different probability from the one entailed by A_2 ($p_1 \neq p_2$), and so the first account entails a different probability of cancer than the second ($p \neq p_1$). Furthermore, perhaps A_1 and A_2 confer different probabilities on the two tumors B_1 and B_2 and these tumors respond differently to different treatments. The additional details provided by the micro-explanation are not stupid and irrelevant. They make a difference—to the probability of the *explanandum*, and to much else.[16] Perhaps it is a good thing for cancer research that the multiple realizability argument has not won the hearts of oncologists.

6. Inference to the Best Explanation. I suspect that the multiple realizability argument has exerted so much influence because of a widespread misunderstanding concerning how *inference to the best explanation* works. The rough idea behind this mode of inference is that one should accept or reject hypotheses by deciding whether they are needed to explain observed phenomena. This inferential procedure seems to bear on the issue of reductionism as follows: We *now* need statements formulated in higher-level sciences because present day physics is not able to tell us how to understand societies, minds, and living things. However, if reductionism is correct, then these higher-level statements will not be needed once we have an ideally complete physics, and so they *then* should be rejected. But surely an ideally complete physics would not make it reasonable to reject all statements in higher-level sciences. This means that those statements must be needed to explain something that statements in an ideal physics could not explain. The multiple realizability argument presents itself as a diagnosis of why this is so.

This line of argument rests on a misunderstanding of inference to the best explanation. If you think that A_1 is one of the micro-realizations that P has, then you should not view "P causes Q" and "A_1 causes Q" as competing hypotheses (Sober 1999). The evidence you have may justify accepting both. Inference to the best explanation is a procedure that belongs to the context of justification. Once you have used that technique to accept a variety of different hypotheses, it is perfectly possible that your set of beliefs will furnish several explanations of a given phenomenon, each perfectly compatible with the others. Some of those explanations will provide more details while others will provide fewer. Some may cite proximal causes while others will cite causes that

16. This argument would not be affected by demanding that a probabilistic explanation must cite the positive and negative causal factors that raise and lower the probability of the *explanandum* (see, e.g., Salmon 1984). Cigarette smoke may raise the probability of lung cancer to a different extent than inhaling A_1 does, and so the two explanations will differ in important ways.

are more distal. The mistake comes when one applies the principle of inference to the best explanation a *second* time—to the set of hypotheses one *already* believes, and rejects hypotheses that one does not "need" for purposes of explanation. Inference to the best explanation is a rule for deciding what to believe; it is not a principle for retaining or eliminating beliefs that one already has perfectly good evidence for accepting. If hypotheses in higher-level sciences can be accepted on the basis of evidence, they will not be cast into the outer darkness simply because physics expands.

It is worth bearing in mind that the phrase "inference to 'the' best explanation" can be misleading. The hypothesis singled out in such inferences is not the best of all explanations (past, present, and future) that could be proposed; it is merely the best of the competing hypotheses under evaluation. Hypothesis testing is essentially a contrastive activity; a given hypothesis is tested by testing it *against* one or more alternatives (Sober 1994). When psychological hypotheses compete against each other, inference to the best explanation will select the best of the competitors; of necessity, the winner in this competition will be a psychological hypothesis, because all the competitors are. Likewise, when physicalistic explanations of a behavior compete against each other, the resulting selection will, of course, be a physicalistic explanation. It is perfectly consistent with these procedures that a given phenomenon should have a psychological *and* a physicalistic explanation. Both reductionists and antireductionists go wrong if they think that the methods of science force one to choose among hypotheses that, in fact, are not in competition at all.[17]

17. This point bears on an argument that Fodor (1998) presents to supplement his (1975) argument against reductionism. I am grateful to Fodor for helping me to understand this new argument. Fodor compares two hypotheses (which I state in the notation I have been using): (i) "if (A_1 or A_2 or . . . or A_n), then Q" and (ii) "if (A_1 or A_2 or . . . or A_n) then P (because the A_i's are possible realizations of P), and if P then Q." Fodor points out that the latter generalization is logically stronger (19); he then claims that it is sound inductive practice to "prefer the strongest claim compatible with the evidence, all else being equal" (20). Since we should accept the stronger claim instead of the weaker one, Fodor concludes that reductionism is false.

I have three objections to this argument. First, I do not think that the two generalizations are in competition with each other. If one thinks that the first conditional is true, and wants to know whether, in addition, it is true that the A_i's are realizations of P, then the proper competitor for this conjecture is that at least one of the A_i's is *not* a realization of P. Second, even if the two hypotheses were competitors, Fodor's Popperian maxim is subject to the well-known "tacking problem"—that irrelevant claims can be conjoined to a well-confirmed hypothesis to make it logically stronger. Fodor, of course, recognizes that $H\&I$ is not always preferable to H, ceteris paribus; however, he thinks that a suitably clarified version of the maxim he describes is plausible and that it will have the consequence he says it has for the example at hand. I have my

7. Two Other Criticisms of the Multiple Realizability Argument. The multiple realizability argument, when it focuses on the explanation of singular occurrences, has three premises:

> Higher-level sciences describe properties that are multiply realizable and that provide good explanations.
>
> If a property described in a higher-level science is multiply realizable at a lower level, then the lower-level science will not be able to explain, or will explain only feebly, the phenomena that the higher-level science explains well.
>
> If higher-level sciences provide good explanations of phenomena that lower-level sciences cannot explain, or explain only feebly, then reductionism is false.
>
> ---
>
> Reductionism is false.

I have criticized the second premise, but the first and third have not escaped critical scrutiny (see, e.g., Lewis 1969, Churchland 1982, Enç 1983, and Kim 1989; Bickle 1998 provides a useful discussion). I will consider these other objections separately.

Philosophers with eliminitivist leanings have criticized the first premise. They have suggested that if "pain," for example, is multiply realizable, then it probably does not have much explanatory power. Explanations that cite the presence of "pain" will be decidedly inferior to those that cite more narrow-gauged properties, such as "human pain," or "pain with thus-and-such a neural realization." Philosophers who advance this criticism evidently value explanations for being deep, but

---

doubts. It is illuminating, I think, to compare this inference problem to a structurally similar problem concerning intervening variables. If the A_i's are known to cause Q, should one postulate a variable (P) that the A_i's cause, and which causes Q? I do not think that valid inductive principles tell one to prefer the intervening variable model over one that is silent on the question of whether the intervening variable exists, when both models fit the data equally well (see Sober 1998 for further discussion). Third, even if the stronger hypothesis should be accepted in preference to the weaker one, I do not see that this refutes reductionism (though it does refute "eliminativist reductionism"). After all, the reductionist can still maintain that "if P then Q" is explained by theories at the lower level.

Notice that Fodor's argument does not depend on whether the A_i's listed are some or all of the possible realizations that P can have; it also does not matter whether the modality involved is metaphysical or nomological. Notice, finally, that this argument concerns inductive inference (the "context of justification," mentioned earlier), not explanation, which is why it differs from the argument of Fodor 1975.

not for being general. I disagree with this one-dimensional view, just as I disagree with the multiple realizability argument's single-minded valuation of generality at the expense of depth. Higher-level explanations often provide fewer explanatory details, but this does not show that they are inferior *tout court.*

It might interest philosophers of mind who have these worries about multiply realized psychological properties to consider the multiply realized properties discussed in evolutionary biology. In cognitive science, it is difficult to point to many present-day models that are well-confirmed and that are articulated by describing multiply realizable properties; this is mostly a hoped-for result of scientific advance. However, in evolutionary biology, such models are extremely common. Models of the evolution of altruism (Sober and Wilson 1998), for example, use the concept of fitness and it is quite clear that fitness is multiply realizable. These models have a useful generality that descriptions of the different physical bases of altruism and selfishness would not possess.

The third premise in the multiple realizablity argument also has come in for criticism. Perhaps *pain* is multiply realizable, but *human pain* may not be. And if *human pain* is multiply realizable, then some even more circumscribed type of pain will not be. What gets reduced is not pain in general, but specific physical types of pain (Nagel 1965). The multiple realizability argument is said to err when it assumes that reductionism requires *global* reduction; *local* reduction is all that reductionism demands. To this objection, a defender of the multiple realizability argument might reply that there are many questions about reduction, not just one. If human pain gets reduced to a neurophysiological state, but pain in general does not, then reductionism is a correct claim about the former, but not about the latter. If psychology provides explanations in which pain—and not just *human* pain—is an *explanans*, then reductionism fails as a claim about *all* of psychology.

Scientists mean a thousand different things by the term "reductionism." Philosophers have usually been unwilling to tolerate this semantic pluralism, and have tried to say what reductionism "really" is. This quest for univocity can be harmless as long as philosophers remember that what they call the "real" problem is to some degree stipulative. However, philosophers go too far when they insist that reductionism requires local reductions but not global reductions. There are many reductionisms—focusing on one should not lead us to deny that others need to be addressed.

8. A Different Argument Against a Different Reductionism. Although the multiple realizability argument against reductionism began with the arguments by Putnam and Fodor that I have reviewed, more recent

appeals to multiple realizability sometimes take a rather different form. The claim is advanced that higher-level sciences "capture patterns" that would be invisible from the point of view of lower-level science. Here the virtue attributed to the higher-level predicate "P" is not that it *explains* something that the lower-level predicate "A_i" cannot explain, but that the former *describes* something that the latter does not. The predicate "P" describes what the various realizations of the property P have in common. The disjunctive lower-level predicate "A_1 or A_2 or . . . or A_n" does not do this in any meaningful sense. If I ask you what pineapples and prime numbers have in common and you reply that they both fall under the disjunctive predicate "pineapple or prime number," your remark is simply a joke. As a result, "if P then Q" is said to describe a regularity that "if (A_1 or A_2 or . . . or A_n) then (B_1 or B_2 or . . . or B_n)" fails to capture.

Whether or not this claim about the descriptive powers of higher- and lower-level sciences is right, it involves a drastic change in subject. Putnam and Fodor were discussing what higher- and lower-level sciences are able to *explain*. The present argument concerns whether a lower-level science is able to *describe* what higher-level sciences *describe*. I suspect that this newer formulation of the multiple realizability argument has seemed to be an elaboration, rather than a replacement, of the old arguments in part because "capturing a pattern" (or a generalization) has seemed to be more or less equivalent with "explaining a pattern" (or a generalization). However, there is a world of difference between describing a fact and explaining the fact so described. This new argument does not touch the reductionist claim that physics can explain everything that higher-level sciences can explain.

9. Concluding Comments. Higher-level sciences often provide more general explanations than the ones provided by lower-level sciences of the same phenomena. This is the kernel of truth in the multiple realizability argument—higher-level sciences "abstract away" from the physical details that make for differences among the micro-realizations that a given higher-level property possesses. However, this does not make higher-level explanations "better" in any absolute sense. Generality is one virtue that an explanation can have, but a distinct—and competing—virtue is depth, and it is on this dimension that lower-level explanations often score better than higher-level explanations. The reductionist claim that lower-level explanations are *always* better and the antireductionist claim that they are *always* worse are both mistaken.

Instead of claiming that lower-level explanations are always better than higher-level explanations of the same phenomenon, reductionists might want to demure on this question of better and worse, and try to

build on the bare proposition that physics in principle can explain any singular occurrence that a higher-level science is able to explain. The level of detail in such physical explanations may be more than many would want to hear, but a genuine explanation is provided nonetheless, and it has a property that the multiple realizability argument has overlooked. For reductionists, the interesting feature of physical explanations of social, psychological, and biological phenomena is that they use the same basic theoretical machinery that is used to explain phenomena that are nonsocial, nonpsychological, and nonbiological. This is why reductionism is a thesis about the *unity* of science. The special sciences unify by abstracting away from physical details; reductionism asserts that physics unifies because everything can be explained, and explained *completely,* by adverting to physical details. It is ironic that "unification" is now a buzz word for antireductionists, when not so long ago it was the *cri de coeur* of their opponents.

To say that physics is capable in principle of providing a complete explanation does not mean that physical explanations will mention everything that might strike one as illuminating. As noted above, the explanations formulated by higher-level sciences can be illuminating, and physics will not mention *them.* Illumination is to some degree in the eye of the beholder; however, the sense in which physics can provide complete explanations is supposed to be perfectly objective. If we focus on *causal* explanation, then an objective notion of explanatory completeness is provided by the concept of *causal completeness*:

$$\begin{aligned}&\Pr(\text{higher-level properties at } t_2 \mid \\ &\text{physical properties at } t_1 \ \&\ \text{higher-level properties at } t_1) = \\ &\Pr(\text{higher-level properties at } t_2 \mid \text{physical properties at } t_1).\end{aligned}$$

To say that physics is causally complete means that (a complete description of) the physical facts at t_1 *determines* the probabilities that obtain at t_1 of later events; adding information about the higher-level properties instantiated at t_1 makes no difference.[18] In contrast, multiple

18. Let M = all the higher-level properties a system has at time t_1. Let P = all the physical properties that the system has at t_1. And let B = some property that the system might have at the later time t_2. We want to show that

$$\Pr(M \mid P) = 1.0$$

entails

$$\Pr(B \mid P) = \Pr(B \mid P \ \&\ M).$$

First note that $\Pr(B \mid P)$ can be expanded as follows:

$$\begin{aligned}\Pr(B \mid P) &= \Pr(B \ \&\ P)/\Pr(P) \\ &= [\Pr(B \ \&\ P \ \&\ M) + \Pr(B \ \&\ P \ \&\ \text{not-}M)]/\Pr(P) \\ &= [\Pr(B \mid P \ \&\ M)\Pr(P \ \&\ M) + \Pr(B \ \&\ \text{not-}M \mid P)\Pr(P)]/\Pr(P) \\ &= \Pr(B \mid P \ \&\ M)\Pr(M \mid P) + \Pr(B \ \&\ \text{not-}M \mid P)\end{aligned}$$

realizability all but guarantees that higher-level sciences are causally incomplete:

$$\Pr(\text{higher-level properties at } t_2 \mid \text{physical properties at } t_1 \,\&\, \text{higher-level properties at } t_1) \neq \Pr(\text{higher-level properties at } t_2 \mid \text{higher-level properties at } t_1).$$

If A_1 and A_2 are the two possible realizations of P, then one should not expect that $\Pr(Q \mid P \,\&\, A_1) = \Pr(Q \mid P \,\&\, A_2) = \Pr(Q \mid P)$ (Sober 1999).

Is physics causally complete in the sense defined? It happens that causal completeness follows from the thesis of simultaneous determination described earlier (Sober 1999). This fact does not settle whether physics *is* causally complete, but merely pushes the question back one step. Why think that the physical facts that obtain at a given time determine all the nonphysical facts that obtain at that time? This is a question I will not try to answer here. However, it is worth recalling that defenders of the multiple realizability argument usually assume that the lower-level physical properties present at a time determine the higher-level properties that are present at that same time. This commits them to the thesis of the causal completeness of physics. If singular occurrences can be explained by citing their causes, then the causal completeness of physics insures that physics has a variety of explanatory completeness that other sciences do not possess. This is reductionism of a sort, though not the sort that the multiple realizability argument aims to refute.

REFERENCES

Berlin, Isaiah (1953), *The Hedgehog and the Fox*. New York: Simon and Shuster.
Bickle, John (1998), *Psychoneural Reduction: The New Wave*. Cambridge, MA: MIT Press.
Churchland, Paul (1982), "Is 'Thinker' a Natural Kind?", *Dialogue* 21: 223–238.
Davidson, Donald (1966), "Emeroses by Other Names", *Journal of Philosophy* 63: 778–780. Reprinted in *Essays on Actions and Events*. Oxford: Oxford University Press, 1980, 225–227.
———. (1970), "Mental Events", in L. Foster and J. Swanson (eds.), *Experience and Theory*. London: Duckworth. Reprinted in *Essays on Actions and Events*. Oxford: Oxford University Press, 1980, 207–225.
Earman, John (1992), *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.
Enç, Berent (1983), "In Defense of the Identity Theory", *Journal of Philosophy* 80: 279–298.
Fodor, Jerry (1968), *Psychological Explanation*. Cambridge, MA: MIT Press.
———. (1975), *The Language of Thought*. New York: Thomas Crowell.
———. (1998), "Special Sciences—Still Autonomous After All These Years", in *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT Press.
Forster, Malcolm and Elliott Sober (1994), "How to Tell When Simpler, More Unified, or

From this last equation, it is clear that if $\Pr(M \mid P) = 1.0$, then $\Pr(B \mid P) = \Pr(B \mid P \,\&\, M)$.

Less *Ad Hoc* Theories Will Provide More Accurate Predictions", *British Journal for the Philosophy of Science* 45: 1–35.
Goodman, Nelson (1965), *Fact, Fiction, and Forecast*. Indianapolis: Bobbs-Merrill.
Hanson, N. Russell (1958), *Patterns of Discovery*. Cambridge: Cambridge University Press.
Hempel, Carl (1965), "Aspects of Scientific Explanation", in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
Howson, Colin and Peter Urbach (1989), *Scientific Reasoning: The Bayesian Approach*. La Salle: Open Court.
Kim, Jaegwon (1989), "The Myth of Nonreductive Materialism", *Proceedings and Addresses of the American Philosophical Association* 63: 31–47. Reprinted in *Supervenience and Mind*. Cambridge: Cambridge University Press, 1993.
Kim, Sungsu (unpublished), "Physicalism, Supervenience, and Causation—a Probabilistic Approach".
Kitcher, Philip (1984), "1953 and All That: A Tale of Two Sciences", *Philosophical Review* 93: 335–373. Reprinted in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 1994, 379–399.
Lewis, David (1969), "Review of *Art, Mind, and Religion*", *Journal of Philosophy* 66: 22–27. Reprinted in N. Block (ed.), *Readings in Philosophy of Psychology*, vol. 1. Cambridge, MA: Harvard University Press, 1983, 232–233.
———. (1973a), "Causation", *Journal of Philosophy* 70: 556–567. Reprinted with a "Postscript" in D. Lewis, *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 1986, 159–213.
———. (1973b), *Counterfactuals*. Cambridge, MA: Harvard University Press. Revised edition 1986.
———. (1986), "Causal Explanation", in D. Lewis, *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 214–240.
Nagel, Ernest (1961), *The Structure of Science*. New York: Harcourt Brace.
Nagel, Thomas (1965), "Physicalism", *Philosophical Review* 74: 339–356.
Oppenheim, Paul, and Hilary Putnam (1958), "Unity of Science as a Working Hypothesis", in H. Feigl, G. Maxwell, and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, Minneapolis: University of Minnesota Press, 3–36.
Putnam, Hilary (1967), "Psychological Predicates", in W. Capitan and D. Merrill (eds.), *Art, Mind, and Religion*. Pittsburgh: University of Pittsburgh Press, 37–48. Reprinted as "The Nature of Mental States" in *Mind, Language, and Reality*. Cambridge: Cambridge University Press, 1975, 429–440.
———. (1975), "Philosophy and our Mental Life", in *Mind, Language, and Reality*. Cambridge: Cambridge University Press, 291–303.
Rosenberg, Alexander (1978), "The Supervenience of Biological Concepts", *Philosophy of Science* 45: 368–386.
———. (1985), *The Structure of Biological Science*. Cambridge: Cambridge University Press.
Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
Sober, Elliott (1983), "Equilibrium Explanation", *Philosophical Studies* 43: 201–210.
———. (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Cambridge, MA: MIT Press. 2nd edition, University of Chicago Press, 1994.
———. (1988), "Confirmation and Lawlikeness", *Philosophical Review* 97: 93–98.
———. (1994), "Contrastive Empiricism", in *From a Biological Point of View*. Cambridge: Cambridge University Press, 114–135.
———. (1998), "Black Box Inference: When Should an Intervening Variable be Postulated?", *British Journal for the Philosophy of Science* 49: 469–498.
———. (1999), "Physicalism from a Probabilistic Point of View", *Philosophical Studies* 95: 135–174.
Sober, Elliott and David S. Wilson (1998), *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Stalnaker, Robert (1968), "A Theory of Conditionals", in N. Rescher (ed.), *Studies in Logical Theory.* Oxford: Blackwell, 98–112.

Waters, Kenneth (1990), "Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics", *PSA 1990.* E. Lansing, MI: Philosophy of Science Association, 125–139. Reprinted in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA: MIT Press, 1994, 402–417.

Leo Kadanoff

Statement

and

Readings

Making a Splash; Breaking a Neck: The Development of Complexity in Physical Systems

Leo P. Kadanoff

The fundamental laws of physics are very simple. They can be written on the top half of an ordinary piece of paper. The world about us is very complex. Whole libraries hardly serve to describe it. Indeed, any living organism exhibits a degree of complexity quite beyond the capacity of our libraries. This complexity has led some thinkers to suggest that living things are not the outcome of physical law but instead the creation of a (super)-intelligent designer.

In this talk, we examine the development of complexity in fluid flow. Examples include splashing water, necking of fluids, swirls in heated gases, and jets thrown up from beds of sand. We watch complexity develop in front of our eyes. Mostly, we are able to understand and explain what we are seeing. We do our work by following a succession of very specific situations. In following these specific problems, we soon get to broader issues: predictability and chaos, mechanisms for the generation of complexity and of simple laws, and finally the question of whether there is a natural tendency toward the formation of complex ‘machines’.

UNIVERSITY OF CHICAGO MAGAZINE

DECEMBER 2002, Volume 95, Issue 2
Written by Sharla Stewart
>> Back to feature

The Complexity Complex

More and more scientists are studying complex systems. Is a new field of study arising, or is science simply getting more complicated?

There is a funny dance that some Chicago physicists and biologists do when the topic of complexity theory comes up. They squirm, skirt the issue, dodge the question, bow out altogether unless they're allowed alternative terms. They twirl the conversation toward their own specific research projects: yes, the projects involve systems that are "rich" and "complicated" and not explained by natural laws of physics; yes, systems that develop universal structures which appear in other, completely different systems on a range of scales; yes, systems that begin with simple ingredients and develop outcomes that are—there's no other word for it—complex.

"The problem with *complexity* is that it has become a buzzword," says Heinrich Jaeger, professor in physics and the James Franck Institute. "What do we mean when we say *complex*? That something is more complicated than its simple components imply? Why not say *complicated*?" he asks. "*Complexity* is a newer word, has a better ring. The word itself has taken on an aura; it's become a label for a lot of things to a lot of people." And that, he says, is a good reason to avoid it: buzzwords are hard to pin down and therefore inherently dangerous.

Jaeger's wariness is well founded. A lazy woman's Lexis-Nexis search for *complexity theory* attests to the term's recent popularity. The search brings up lots of business articles on topics ranging from Southwest Airlines' air-cargo system (modeled on the complex swarming tendencies of ants) to fluctuations in stock prices. An equally large number of popular-science stories come up, typically painting an image of a theory to unlock all mysteries. More often than not these articles cite researchers at the interdisciplinary, 18-year-old Santa Fe Institute, whose single-minded insistence that laws of complexity can explain nearly any phenomenon rankles many an academic.

Yet academics aren't immune to the fever. In April 1999 *Science* magazine ran a special issue in which distinguished researchers reflected on how "complexity" has influenced their fields. As evidence of academe's move into the realm of the complex, one article cited an academic building boom in multidisciplinary science centers—including Chicago's nascent Interdisciplinary Research Building. This fall's release of computer-science wunderkind Steven Wolfram's tome *A New Kind of Science* resulted in a flurry of articles about whether simple, fundamental laws, as Wolfram argues, can explain all things complex—including evolution and free will. (Physics Nobel laureate Steven Weinberg argued in the *New York Review of Books* that they most emphatically cannot.)

Judging from the press, complexity theory is on the verge of, if not already, changing the world.

The buzz is enough to drive many researchers away from the term. But one Chicago physicist is more than willing to use the term *complexity*. Two years ago he gave the University's annual Nora and Edward Ryerson Lecture on the topic, titled "Making a Splash, Breaking a Neck: The

Development of Complexity in Physical Systems," and he contributed "Some Lessons from Complexity" to *Science*'s special complexity issue. He will name names when asked who else at Chicago might provide insight into the rise in studies of complex systems. ("I know Leo said this is what I study, but..." is a common conversation starter.)

The "Leo" in question is bright-eyed, white-bearded Leo P. Kadanoff, the John D. MacArthur distinguished service professor in physics and mathematics, the James Franck and Enrico Fermi Institutes, and the College. A National Medal of Science winner and a founder of the soft condensed-matter field in physics, Kadanoff has mused over complexity theory for the past three decades. What he will tell you is that, despite what the press says, there is no theory—no set of laws—of complexity. Only lessons and "homilies."

A definition of the nature of complexity, Kadanoff says, "has been somewhat elusive." But if one were to try, "what we see is a world in which there seems to be organization built up in some rich and interesting fashion—from huge mountain ranges, to the delicate ridge on the surface of a sand dune, to the salt spray coming off a wave, to the interdependencies of financial markets, to the true ecologies formed by living things. For each kind of organization, we want to understand how it arose and whether it has any general rules associated with it."

The "metachallenge" in seeking these rules, he says, is, "What can you learn from one complex system that you can apply to another? Even though there are not any laws of complexity, there are experiences that you can have with one complex system that will help you study another. Even in systems which are very complex, there are aspects of their behavior which might be simple and predictable."

The job as Kadanoff and others at the University see it, regardless of their willingness to label the systems they study complex, is "to reach into these systems to try to distinguish between the things that are predictable and not predictable; in the ones that are predictable to try to pick out the universal features, and then to do something to characterize the unpredictable parts."

That's the driving force behind complexity studies: to characterize what has for so long eluded characterization. "Complexity," reflects physics professor Tom Witten, "is where a system is more ordered than random because it can be described in a nutshell. The nutshell might be big, but you can describe what's going on. And the more you discover, the more payoff you get, because you have simplified [what's being described] below what it was at the outset. The good thing is that you'll never reach the task's end, and you're often rewarded by finding more.

"I never think about whether something is complex," Witten reiterates. "I think about things because they are intriguing—and wouldn't it be terrible if I ever got a complete nutshell?"

What counts as a complex—complicated, rich, interestingly organized, needing-a-big-nutshell—system depends on the eye of the beholder.

Chicago researchers study what might seem mundane: how grains arrange themselves, for example, or how a drop of liquid breaks apart, or how a surface crumples. And they study what seems almost overwhelmingly complicated: for example, how an entire star manages to explode, or how the genomic architecture of *E. coli* is naturally programmed to lead the single-celled bacterium to form colonies and communicate as a multicellular, highly evolved organism.

The last is the work of biochemistry & molecular biology professor James Shapiro. Complexity watchers might have seen Shapiro last year in the *New York Times* and the *Economist* during a minor media flurry over the founding of the Institute for Complex Adaptive Matter (ICAM), an

independent unit of the Los Alamos National Laboratory and the University of California, Berkeley. He presented a highly quotable lecture on migration patterns created by colonies of the bacteria *Proteus mirabilis*. The *Economist* lauded his research for "illuminating problems as diverse as disease, water treatment, corrosion, and the formation of certain metal ores."

Biological systems, even physicists agree, are as complex as a system comes. Although Shapiro, like most of his colleagues, is wary of the term *complexity theory*, he believes the ideas that arise from studying complexity are opening new realms of study for biologists.

"For the physicist, the properties of a complex system emerge out of the individual interactions that compose it," says Shapiro in his second-floor Cummings Life Science Center office—a block away from the Research Institutes where Jaeger, Kadanoff, Witten, and colleagues muse over complex physical systems. "They look at complexity as systems with many interacting components and then somehow these systems develop interesting properties." He's right. When Kadanoff refers to the complexity apparent in the delicate ridge of a sand dune or the salt spray coming off a wave, he is conjuring the complicated outcomes that can arise when simple grains of sand or water molecules are placed under certain conditions: high winds, for instance.

A growing number of biologists, Shapiro argues, approach complexity from an entirely different angle. "Biological systems are very complicated and very complex, but they have clear functionalities. For organisms things have to be done and done right." What interests biologists, he says, is how the organism uses complexity to adapt. "While the physicist asks, How does complexity generate something that is describable with pattern to it, the biologist asks, How does the organism use complexity to achieve its objectives?"

And where physicists are interested in characterizing the unpredictable outcomes of a complex system—the magnitude and direction of a sand-dune avalanche, or the trajectories and sizes of sea-spray droplets—there is a notable absence of chaos in the systems biologists study. That fact alone is intriguing. "Why is it that biological systems are so unbelievably complex but work so reliably?" asks Shapiro. "Why don't they undergo chaotic transitions? What allows biological systems to utilize complexity but not to be overwhelmed by it?"

Finding the answers, he believes, depends upon understanding two things: how the large numbers of components in biological systems interact to create precise functional behavior, and how basic principles of regulation and control operate at all levels in living organisms. Applying those concepts to genetics requires a shift toward what Shapiro has called "a 21st-century view of evolution."

The past 50 years of genetic research, he argues, have provided clear evidence to contradict the prevailing theory that organisms evolve in a "random walk" from adaptation to adaptation. Rather, evolution is the result of "natural genetic engineering"—a highly refined and efficient problem-solving and genetic-reorganization process carried out by a genomic architecture that is, he notes, remarkably similar to a computational system. The information processing occurs in an organism's cells via molecular interactions, and the data on which the processing runs is stored in the DNA.

Contrary to popular belief, "the character of an organism is not determined solely by its genome," Shapiro maintains. "By itself, DNA is inert." Instead, survival and reproduction are the result of how cells' information-processing systems evaluate multiple internal and environmental signals and draw on the data stored in DNA to adapt quickly and reliably. "Cells have to deal with literally millions of biochemical reactions during each cell cycle and also with innumerable unpredictable

contingencies," Shapiro noted at the 2001 International Conference on Biological Physics in Kyoto, Japan. The constantly looming unpredictability doesn't overwhelm the system because, as Shapiro explains in a 1997 *Boston Review* article, "all cells from bacteria to man possess a truly astonishing array of repair systems which serve to remove accidental and stochastic sources of mutation. Multiple levels of proofreading mechanisms recognize and remove errors that inevitably occur during DNA replication."

In fact, cells protect themselves against "precisely the kinds of accidental genetic change that, according to conventional theory, are the sources of evolutionary variability."

If accidents don't cause evolution, what does? The primary perpetrators of evolutionary change, Shapiro says, are mobile genetic elements—DNA structures found in all genomes that can shuttle from one position to another in the genome, cutting and splicing like a Monsanto engineer. Thanks to these mobile little guys, he notes in the *Review*, "genetic change can be specific (these activities can recognize particular sequence motifs) and need not be limited to one genetic locus (the same activity can operate at multiple sites in the genome). In other words, genetic change can be massive and nonrandom."

Shapiro's contribution to the new view of evolution is to demonstrate that the elements in the computational genome are universal beyond people, plants, and animals. Bacterial genomes, his work demonstrates, also operate and evolve via natural genetic engineering. The process is not random; it's in- fluenced by the bacteria's experience. Moreover, bacteria experience their environment not as individual cells oblivious to others in the colony, but as a multicellular organism. This is evident in the patterns they create.

"That the patterns exist tells us that the bacteria are highly organized, highly differentiated, and highly communicative," he explains. "In biology when you see regularity and pattern and control working, you say, Well, what is it functionally related to, what's the adaptive utility for the organism?"

On his Macintosh PowerBook Shapiro points his browser to his Web site, where he's posted movies of bacteria colonies growing and migrating. Running in black and white, the QuickTime films have the scratchy monochromatics of the silent era. One depicts five *E. coli* cells scattered on agar. The squirmy, haloed cells begin growing and dividing, and then the daughter cells grow and divide, and soon there are five little colonies surrounded by halos. "The daughter cells are clearly interacting," says Shapiro. "What I am interested in is, are they interacting because they're communicating or simply because each cell is internally programmed independently of the other cell? The way to tell is by looking at what happens when the scattered colonies encounter one other."

The five colonies seem to seek each other out, growing first toward each other, meeting and merging, then spreading outward en masse. "The very least you can say from this observation is that *E. coli* cells maximize cell-to-cell contact." How the cells communicate with each other—whether they sense a chemical signal, perhaps in the halo, or a physical signal from the other bacteria—has yet to be determined. "But that they interact," says Shapiro, "is quite clear."

Another film depicts an *E. coli* colony advancing across a petri dish on which a glass fiber lies diagonally. The edge of the colony moves along until, *boop*!, it hits the fiber's top end. Suddenly the bugs at the colony's own top edge are released. They use their flagella to swim around the fiber, nosing into it and wiggling vigorously. "According to conventional wisdom and how they were grown," says Shapiro, "those individual cells shouldn't have been motile." Meanwhile, the

lower edge of the colony has not yet met the fiber; its slow advance continues. Shapiro points out that the cells around the fiber swim and divide but do not spread over the agar; only the older, organized colony expands over the surface. After two hours the colony's lower edge meets the fiber's lower diagonal. The colony spends some time on the fiber, filling in its mass, before eventually spreading past it and continuing to advance. Yet, rather than being swept up and carried along like picnic crumbs on the backs of ants, the fiber remains in place. "That tells you that the whole colony is not expanding; just the region at the edge is moving outwards," explains Shapiro. "There's a small zone of active movement, and then everything stays in place." The colony expands over the agar not simply by cells dividing and spilling over; rather, an organized structure is at work.

Shapiro's movies of *Proteus mirabilis* reveal an even more organized growth and migration structure. In *Proteus* specialized cells called swarmers are responsible for colony spreading. After a period of eating and dividing, the colony releases swarmers outward; the expanded colony pauses, eats and divides, and eventually sends more swarmers out. The resulting pattern is a series of rings similar to a tree's. Swarmer cells, Shapiro notes, move only in groups—isolated, they go nowhere—and they do not divide. Short, fat cells are responsible for cell multiplication. In a way that may suggest a supercomputer coordinating the activity of large numbers of interconnected processors, the expanding *Proteus* colony coordinates the movement of large numbers of swarmer cells.

Without the focus on the issues of complexity, Shapiro believes, biologists would be at a loss to explain the behavior he's caught on film. "In biological systems, at least, trying to understand how the components of these complicated, complex systems interact and do something adaptive is central to understanding them at a deeper level and probably," he adds, "to understanding all of nature."

Back the lens out several hundred thousand light years and expand the frame exponentially. A neutron star, its surface roiling in flame and gas—this time in full, glorious color—explodes.

Talk about complex. Now imagine reenacting it.

That's what a long row of academic posters in the fourth-floor hallway of the Research Institutes Building on Ellis Avenue is dedicated to: simulations of the complex interactions that contribute to a supernova and other exploding stars. This gallery of fantastic images and nearly incomprehensible astrophysical explanations is the work of the federally sponsored Accelerated Strategic Computing Initiative's (ASCI) five-year-old Center for Astrophysical Thermonuclear Flashes.

"Part of the challenge—what's fun—is to take apart a complicated pattern. There's an art to this. It's not cut and dried; there's no recipe for a supernova as yet," says Robert Rosner, the center's associate director and the William E. Wrather distinguished service professor in astronomy & astrophysics, physics, the Enrico Fermi Institute, and the College. (Until his October appointment as chief scientist at Argonne National Laboratory, Rosner was the center's director.) "We have to figure out how to take it apart into simpler pieces. Sometimes we get something that is, as yet, impossible to understand. The trick is to get pieces that we can explain and to reassemble these understood pieces into a whole which we can comprehend as an explanation of how an evolved star explodes."

Where the computation metaphor allows Shapiro to consider bacteria as highly organized, problem-solving organisms, the nutshells that Rosner's group wraps around supernovae are

equations that, when crunched, create simulations. The orgy of brilliant, curving, flaming gases depicted in "Helium Detonations on Neutron Stars" is one of the largest nutshells the center has obtained to date. The image (on pages 38–39) is the result of an integrated calculation, one that involves many subsidiary calculations conveying all the smaller complicated interactions and chaos-producing dynamics. Together they create a massive burst of exploding helium on the surface of a hypothetical collapsed star that's dense with closely packed neutrons. Before the group could simulate the burst—much less a supernova—it first had to find the correct equations to describe a detonation, regardless of whether it occurs in a star or a laboratory.

"We ask first, can we understand these events in isolation, separate from their environment and other events? An exploding star, whether a detonation on the surface of a neutron star or an explosion within a white dwarf, leading to a supernova, involves not just detonation, but flames—deflagration—and instability. What happens, for example, when we put a heavy fluid on top of a light fluid?" Rosner asks. "If we can answer those questions, we move up from there."

A heavy fluid (cold, dense fuel) sinking into a light fluid (hot ashes) during a nuclear burn—the so-called Raleigh-Taylor instability, which the center's research scientist Alan Calder has modeled—creates turbulence, or chaotic flow in a fluid, which is physicist Kadanoff's speciality. The center's simulation of the Raleigh-Taylor instability, an abstract pitching wave of reds, yellows, and oranges (on pages 44–45), confirms what Rosner, Kadanoff, and other physicists already know: that the more minute the detail they try to define in the fluid's resulting structure of swirling plumes, the more it eludes them. "The deeper you look at this thing, it never settles down," says Kadanoff.

Turbulence, Rosner explains, is a "real-life exhaustive problem" that presently lies beyond researchers' predictive abilities. It mystifies them not only in simulations of stellar bodies but also in understanding how coffee and cream move when stirred. "The challenge for experimentalists," he says, "is to measure at every point the fluid's temperature, its flow velocity, and density." Turbulence lies within the realm of complexity that at best, as Kadanoff put it, researchers "do something to characterize."

The ability to simulate an experiment and remain faithful to what actually happens, Rosner says, to look at "fully turbulent" systems, like those in a neutron star rather than in a coffee cup, and know exactly what is happening, will "bring simulations to another realm of experimental science."

The turbulence problem underscores a larger point in studies of complex systems. Physical experiments and equation-crunching simulations must for the foreseeable future at least maintain a symbiotic relationship. Given the level of unpredictability in complex systems, using one without the other is like going blind in one eye: you lose depth perception.

In his office Heinrich Jaeger has a poster of a rail yard filled with open boxcars. The cars are piled high with grain, sloping against a cornflower-blue sky. Perched on one boxcar's top edge is a man reading a newspaper. As the photographer no doubt intended, the man snags the viewer's eye, and the grain piles recede into the background.

But not for Jaeger. What Jaeger sees are universal elements and unpredictability in those grain piles. How they pile, what triggers an avalanche, how most flowed through the chute that shot them into the boxcars, how some jammed: these are the complex behaviors Jaeger wishes to describe. While Rosner and Kadanoff think in equations and at computer screens, Jaeger and his colleagues in the Materials Research Science and Engineering Center (MRSEC) work with actual matter: grains, fluids, and various surfaces. Theirs are the experiments that feed simulations,

and the experiments they conduct aim to reduce a complex system to its simplest, easiest-to-observe components.

"Much of modern scientific work in pattern formation and pattern recognition," Sidney Nagel, professor in physics and the James Franck Institute, reflects in a 2001 *Critical Inquiry* essay, "is an attempt to put what the eye naturally sees and comprehends into mathematical form so that it can be made quantitative." Where most viewers might skim over the monotonous grain piles, Jaeger observes their patterns and behavior, composing equations to describe their dynamics; Nagel's patterns of choice, meanwhile, are the elongated necks of dripping drops of fluid. "I am seduced by the shape of objects on a small scale," Nagel's essay continues. "The forces that govern their forms are the same as those that are responsible for structures at ever increasing sizes; yet on the smaller scale those forms have a simplicity and elegance that is not always apparent elsewhere."

For condensed-matter physicists such as experimentalists Nagel and Jaeger and theorist Witten, even when objects are reduced to their simplest forms, there is always an element of wide-eyed wandering. "Serendipity is OK. A bubbly atmosphere is extremely powerful. The key is that when you find something good, you need to realize it," says Jaeger. "There is no clear goal. But when you're awake while wandering around, when you're paying attention, and something comes along that's exciting, you can pick it up. It's a high-risk, high-payoff approach—and the preferred approach if you're charting territory no one's been in before."

It's the only approach a researcher can take with complex systems. Jaeger's granular materials, from the nanoscale to the scale of marbles, fall in the realm of complexity (though he prefers *complicated*) because they often defy what's already known in condensed-matter physics. Taken together, "large conglomerations of discrete particles," he and Nagel propose in a 1996 *Physics Today* article, "behave differently from any of the other standard and familiar forms of matter: solid, liquids, and gases, and [granular material] should therefore be considered an additional state of matter in its own right." Nagel's stretched fluid necks, similarly, are nonlinear: too many phenomena are involved to be accounted for in linear equations. Down the hall Witten studies the nonlinear behavior of distorted matter: the crumpling of silver Mylar paper, and on a more minute level, of polymers packed into a small space.

Once the physicists are set free from linear reasoning, they can set about seeking the complex forms' universals and characterizing their unpredictables. What Nagel has discovered is that all drops breaking apart, regardless of their size, experience a "finite time singularity"—their necks grow infinitely thinner and the forces acting on them infinitely larger until the infinite becomes finite, and the neck breaks. The break-up is a universal element repeated in all drops, of any size and any fluid. Witten, meanwhile, sees analogous singularities in the peaks and ridges of a thin crumpled sheet. These singularities do not develop at a moment in time like Nagel's. Instead, one approaches the singular shapes by making the sheets thinner and thinner. By examining the limiting behavior of these sheets, he finds universal shapes on scales ranging from cell walls to mountain ranges. The next nutshell to wrap around the phenomenon is why and where different-sized ridges buckle during crumpling; his guess is that the distribution of various sized ridges and peaks is also universal from material to material.

These MRSEC experiments and others like them are the building blocks for simulations created by Rosner's and Kadanoff's groups. "Simulators," Rosner notes, "solve equations. We must ask, first, Are we solving the right equations, and second, Are the equations correctly solved? Experimentalists tell us whether we're solving the right equations. Can our calculations produce

what the experimentalists can measure in the lab? The next question is, Can we solve problems that are produced in nature? Somewhat. But that answer will change in the coming decade." (He estimates that in three years turbulence simulations will be cracked.)

Just as Rosner is still unable to precisely simulate full turbulence, graduate students in Kadanoff's group have been unable to fully simulate all the details observed in an experiment by Nagel's graduate students. Nagel's team uses strobe photography to capture what happens in the lab: a fluid placed in a strongly charged electric field rises in a mound toward an electrode. The fluid comes to a point, and some motion occurs between fluid and electrode, resolving itself in a form strikingly similar to a lightning bolt. After the bolt flashes, in the space between the fluid and the electrode a spray of fine water droplets—something like rain—appears. Kadanoff's group has been able to simulate only as far as the mound rising to a point; the outcome, lightning and rain, is still too complex for equations.

"There is a lesson from this," Kadanoff noted in his 2000 Ryerson lecture. "Complex systems sometimes show qualitative changes in their behavior. Here a bump has turned into lightning and rain. Unexpected behavior is possible, even likely."

To study complexity, as Kadanoff has remarked, is to attempt to say something about the "interesting" organization of the world around us, to quantify what seems simple yet defies quantification. It is a search for metaphors that, like Shapiro's use of a computation framework, open new ways of thinking. "When a system transitions from simple to complex is something that I wouldn't know how to define," Shapiro reflects. "And I think that's actually a great problem: how do we distinguish between what we call simple and what we call complex? Even things that seem simple, when you look at them in enough detail, they inevitably become more complex."

But the idea, Witten observes, "is that there's a magic way to say what's happening, where once you say one further thing, the rest is simple. Is that what *complexity* means, or is that just the purpose of science?"

phone: 773/702-2163 • fax: 773/702-0495 • uchicago-magazine@uchicago.edu

VIEWPOINT

Simple Lessons from Complexity

Nigel Goldenfeld[1] and Leo P. Kadanoff[2]

The complexity of the world is contrasted with the simplicity of the basic laws of physics. In recent years, considerable study has been devoted to systems that exhibit complex outcomes. This experience has not given us any new laws of physics, but has instead given us a set of lessons about appropriate ways of approaching complex systems.

One of the most striking aspects of physics is the simplicity of its laws. Maxwell's equations, Schrödinger's equation, and Hamiltonian mechanics can each be expressed in a few lines. The ideas that form the foundation of our worldview are also very simple indeed: The world is lawful, and the same basic laws hold everywhere. Everything is simple, neat, and expressible in terms of everyday mathematics, either partial differential or ordinary differential equations.

Everything is simple and neat—except, of course, the world.

Every place we look—outside the physics classroom—we see a world of amazing complexity. The world contains many examples of complex "ecologies" at all levels: huge mountain ranges, the delicate ridge on the surface of a sand dune, the salt spray coming off a wave, the interdependencies of financial markets, and the true ecologies formed by living things. Each situation is highly organized and distinctive, with biological systems forming a limiting case of exceptional complexity. So why, if the laws are so simple, is the world so complicated? Here, we try to give a partial answer to this question and summarize general lessons that can be drawn from recent work on complexity in physical systems.

To us, complexity means that we have structure with variations. Thus, a living organism is complex because it has many different working parts, each formed by variations in the working out of the same genetic coding. One look at ocean or sky gives the conviction that there is some natural tendency toward the formation of structure in the physical world. Chaos is also found very frequently. Chaos is the sensitive dependence of a final result upon the initial conditions that bring it about. In a chaotic world, it is hard to predict which variation will arise in a given place and time. Indeed, errors and uncertainties often grow exponentially with time.

A complex world is interesting because it is highly structured. A chaotic world is interesting because we do not know what is coming next. But the world contains regularities as well. For example, climate is very complex, but winter follows summer in a predictable pattern. Our world is both complex and chaotic. From this, an elementary lesson follows:

Nature can produce complex structures even in simple situations, and can obey simple laws even in complex situations.

Creating complexity. Fluids frequently produce complex behavior, which can be either highly organized (think of a tornado) or chaotic (like a highly turbulent flow). What is seen often depends on the size of the observer. A fly caught in a tornado would be surprised to learn that it is participating in a highly structured flow.

The equations that describe how the fluid velocity at one point in space affects the velocity at other points in space are derived from three basic ideas:

Locality. A fluid contains many particles in motion. A particle is influenced only by other particles in its immediate neighborhood.

Conservation. Some things are never lost, only moved around, such as particles and momentum.

Symmetry. A fluid is isotropic and rotationally invariant.

To make a computer fluid, construct (*1*) a kind of square dance in which particles move around, obeying the three basic ideas. In the simplest case, the dance is done on a regular hexagonal lattice (Fig. 1, upper panel). Each particle is characterized by a lattice position and by one of six directions of motion. These arrows are momentum vectors. The square dance starts when the caller says "Promenade"; this call instructs each dancer to proceed one step in the direction of its arrow (Fig. 1, middle panel). And then the caller says "Swing your partner." This is an instruction to rotate all the arrows on a given site through 60°, if they happen to add up to zero total momentum (Fig. 1, lower panel). Notice that both particle number and momentum are conserved in each step. Take thousands of particles and thousands of steps, average a bit to smooth out the data, and thereby find a pattern of motion identical to fluid motion. The square dance behaves like a fluid simply because its steps obey the three fundamental laws of fluid motion (*2*).

Gradually, through examples like this, it has dawned on us that very simple ingredients can produce very beautiful, rich, and patterned outputs. Thus, our square dancers, through their simple hops and swings, produce the entire beautiful world of fluids in motion. For simple elementary actors to produce patterned and complex output, we require many events. Our example included many events because it had many actors and much time.

For physicists it is delightful, but not surprising, that the computer generates realistic fluid behavior, regardless of the precise details of how we do the coding. If this were not the case, then we would have extreme sensitivity to the microscopic modeling—what one might loosely call "model chaos"—and physics as a science could not exist: In order to model a bulldozer, we would need to be careful to model its constituent quarks! Nature has been kind enough to have provided us with a convenient separation of length, energy, and time scales, allowing us to excavate physical laws from well-defined strata, even though the consequences of these laws are very complex. But we might not be so lucky with complexity in biological or economic situations.

Understanding complexity. To extract physical knowledge from a complex system, one must focus on the right level of description. There are three modes of investigation of systems like this: experimental, computational, and theoretical. Experiment is best for exploration, because experimental techniques (combined with the human eye) can scan large ranges of data very efficiently.

Computer simulations are often used to check our understanding of a particular physical process or situation. In our fluid dynamics example, the large-scale structure is independent of detailed description of the motion on the small scales. We can exploit this kind of "universality" by designing the most convenient "minimal model." For example, most fluid flow programs should not be modeled by molecular dynamics simulations. These simulations are so slow that they may not be able to reach a regime that will enable us to safely extrapolate to large systems. So we are likely to get the wrong answer. Instead, we should model at the macro level, using large time steps and large systems. For example, some computational biologists try to simulate protein dynamics by following each and every small part of the molecule. The result? Most of the computer cycles are spent watch-

[1]Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801, USA. E-mail: nigel@uiuc.edu [2]Departments of Physics and Mathematics, James Franck Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60615, USA. E-mail: leoP@uchicago.edu

ing little CH groups wiggling back and forth. Nothing biologically significant occurs in the time they can afford.

Use the right level of description to catch the phenomena of interest. Don't model bulldozers with quarks.

This lesson applies with equal strength to theoretical work aimed at understanding complex systems. Modeling complex systems by tractable closure schemes or complicated free-field theories in disguise does not work. These may yield a successful description of the small-scale structure, but this description is likely to be irrelevant for the large-scale features. To get these gross features, one should most often use a more phenomenological and aggregated description, aimed specifically at the higher level. Thus, financial markets should not be modeled by simple geometric Brownian motion–based models, all of which form the basis for modern treatments of derivative markets. These models were created to be analytically tractable and derive from very crude phenomenological modeling. They cannot reproduce the observed strongly non-Gaussian probability distributions in many markets, which exhibit a feature so generic that it even has a whimsical name, fat tails. Instead, the modeling should be driven by asking "What are the simplest nonlinearities or nonlocalities that should be present?"—that is, by trying to separate universal scaling features from market-specific features. The inclusion of too many processes and parameters will obscure the desired qualitative understanding.

Every good model starts from a question. The modeler should always choose the correct level of detail to answer the question.

Complexity and statistics. As a fluid moves around, it may carry with it some "passive" elements that do not themselves influence the flow. Both energy and the density of impurities undergo this kind of motion, in which they convect (go with the flow) and diffuse (move randomly). The convective motion tends to move initially distant regions of the fluid close to one another, thereby producing enhanced gradients. The diffusion tends to smooth out the gradients.

In many situations, these "passive scalars" are carried along by a rapid and turbulent flow, so that the convective mixing tends to dominate the diffusion. Computer simulations and experiments show that the density of the scalar soon develops a profile in which there are many flat regions surrounded by abrupt jumps. The flat regions are produced by the combined effects of convection and diffusion in well-mixed regions of the sample. However, because the density must generally follow the initial gradient, mixed regions must be separated by jumps.

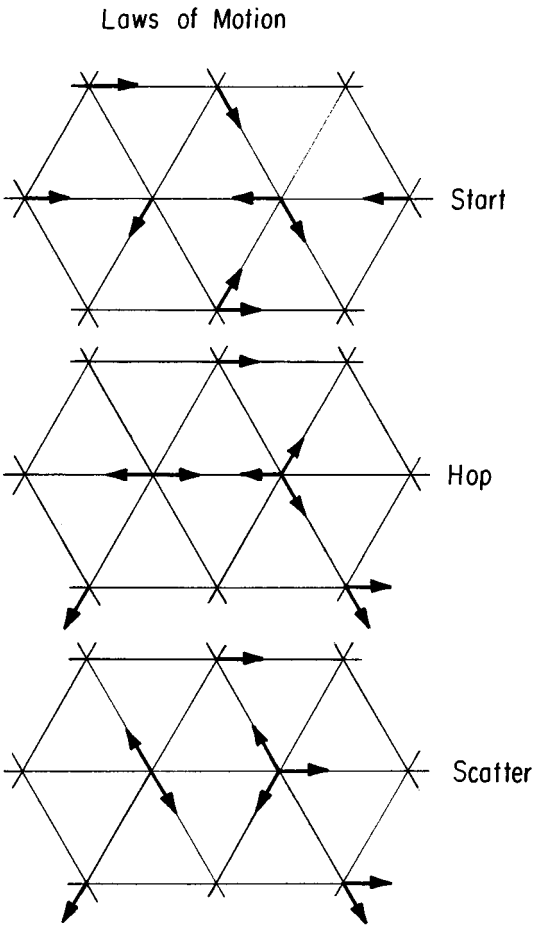


Fig. 1. Three stages in the update algorithm of a lattice gas. Between the upper and middle panels, each particle moves in the direction of its arrow to arrive at a nearest neighboring site. Next, particles "collide" whenever the total momentum on a site is zero; these collisions occur between the middle and lower panels.

This behavior, in which the system is dominated by really big events, is called intermittency. Intermittency seems to be a ubiquitous feature of dynamical systems. The weather turns stormy suddenly. There are ice ages. The stock market crashes. A plague takes hold. An airplane runs into turbulence. In every case, there is a big jump in the behavior of a dynamical system, and that big jump can have big human consequences.

These ubiquitous jumps come in all sizes, with the big jumps being less likely. Empirically, the size of the jumps is often given by a probability distribution, which for large jumps takes the form

$$P(\text{jump}) = \frac{1}{2\sigma}\exp\left(-\frac{|\text{jump}|}{\sigma}\right) \qquad (1)$$

(*3*), where σ is the standard deviation. Contrast this with the usual Gaussian form

$$P(\text{jump}) = \frac{1}{(2\pi\sigma)^{1/2}}\exp\left[-\frac{(\text{jump})^2}{(2\sigma^2)}\right] \qquad (2)$$

which has been the usual guess in statistical problems since the time of Galton. Chaotic and turbulent systems often show exponential behaviors, like Eq. 1. Improbable (very bad) events are much more likely with the exponential form than with the Gaussian form (Eq. 2). For example, a 6σ event has a chance of 10^{-9} of occurring in the Gaussian case, whereas with the exponential form the chance is 0.0025. Estimates, particularly Gaussian estimates, formed by short time series will give an entirely incorrect picture of large-scale fluctuations. These considerations have important consequences in, for example, financial markets, as emphasized recently by Mandelbrot (*4*). Thus, we come to another lesson:

Complex systems form structures, and these structures vary widely in size and duration. Their probability distributions are rarely normal, so that exceptional events are not that rare.

The development of complexity in physics. Long ago, Katchalsky (*5*) and Prigogine (*6*) described the formation of complex structures in nonequilibrium systems. Their "dissipative structures" could have a degree of complication that could grow rapidly in time. It is believed that comparably complex structures do not exist in equilibrium. Turing (*7*) described a mechanism, involving reaction diffusion equations, for the development of organization in living things. As we have seen from the examples quoted here and many others, in nonequilibrium situations many-particle systems can get very complicated indeed (*8*).

It is likely that this tendency is the basis of life. A restricted version of this idea is given

in Bak, Tang, and Wiesenfeld's "self-organized criticality" (*9*). In an essay entitled "More Is Different," Anderson (*10*) described how features of organization may arise as an "emergent" property of systems. An example of this point of view is given by work on complexity "phase transitions" and accompanying speculations that various aspects of biological systems sit on a critical point between order and complexity (*11*).

The next few years are likely to lead to an increasing study of complexity in the context of statistical dynamics, with a view to better understanding physical, economic, social, and especially biological systems. It will be an exciting time. As science turns to complexity, one must realize that complexity demands attitudes quite different from those heretofore common in physics. Up to now, physicists looked for fundamental laws true for all times and all places. But each complex system is different; apparently there are no general laws for complexity. Instead, one must reach for "lessons" that might, with insight and understanding, be learned in one system and applied to another. Maybe physics studies will become more like human experience.

References and Notes

1. U. Frisch, B. Hasslacher, Y. Pomeau, *Phys. Rev. Lett.* **56**, 1505 (1986); J. Hardy, O. de Pazzis, U. Frisch, *J. Math. Phys.* **14**, 1746 (1973); *Phys. Rev. A* **13**, 1949 (1976).
2. Early work on the derivation of hydrodynamics from conservation laws can be found in S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases* (Cambridge Univ. Press, Cambridge, ed. 3, 1970).
3. A. R. Kerstein, *J. Fluid Mech.* **291**, 261 (1997); S. Wunsch, thesis, University of Chicago (1998). For experiments, see, for example, B. Castaing *et al.*, *J. Fluid Mech.* **204**, 1 (1989). For theory, see E. Siggia and B. Shraiman, *Phys. Rev. E* **49**, 2912 (1994).
4. B. Mandelbrot, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk* (Springer-Verlag, New York, 1997).
5. A. Katchalsky and P. F. Curan, *Nonequilibrium Processes in Biophysics* (Harvard Univ. Press, Cambridge, MA, 1967).
6. G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
7. A. Turing, *Philos. Trans. R. Soc. London Ser. B* **327**, 37 (1952).
8. For example, L. Kadanoff, A. Libchaber, E. Moses, and G. Zocchi [*Recherche* **22**, 629 (1991)] discussed the development of interlinked structures in a Rayleigh-Benard flow.
9. P. Bak, C. Tang, K. Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987); J. M. Carlson, J. T. Chayes, E. R. Grannan, G. H. Swindle, *ibid.* **65**, 2547 (1990).
10. P. W. Anderson, *Science* **177**, 393 (1972).
11. S. A. Kauffman, *The Origin of Order* (Oxford Univ. Press, Oxford, 1993); *At Home in the Universe* (Oxford Univ. Press, Oxford, 1995).

12. Supported in part by NSF grant NSF-DMR-93-14938 (N.G.) and by the ASCI Flash Center at the University of Chicago under U.S. Department of Energy contract B341495 (L.P.K.).

VIEWPOINT

Complexity in Chemistry

George M. Whitesides* and Rustem F. Ismagilov

"Complexity" is a subject that is beginning to be important in chemistry. Historically, chemistry has emphasized the approximation of complex nonlinear processes by simpler linear ones. Complexity is becoming a profitable approach to a wide range of problems, especially the understanding of life.

"Complexity" is a word rich with ambiguity and highly dependent on context (*1*). Chemistry has its own understandings of this word. In one characterization, a complex system is one whose evolution is very sensitive to initial conditions or to small perturbations, one in which the number of independent interacting components is large, or one in which there are multiple pathways by which the system can evolve. Analytical descriptions of such systems typically require nonlinear differential equations. A second characterization is more informal; that is, the system is "complicated" by some subjective judgment and is not amenable to exact description, analytical or otherwise.

In chemistry, almost everything of interest is complex by one or both definitions. Consider the design and synthesis of a simple organic substance ($<10^2$ covalently bonded, first-row atoms) as a candidate drug—a representative activity for organic, medicinal, and biological chemists. A single step in the multistep synthesis of such a substance might involve 10^{22} molecules of several types (each comprising as many as 10^2 anharmonically oscillating bonds) and several times this number of interacting nuclei and electrons, all immersed in 10^{24} molecules of solvent. The synthesis itself might proceed by perhaps 10 different strategies (that is, sequences of reactions) for making and breaking bonds and for generating the intermediate compounds that ultimately result in the final compound; each strategy might have many thousands of possible variants differing in synthetic detail. The design of a molecule that has the right properties (shape, surface properties, and associated electrostatic fields) to interact specifically with one part of the surface of a target protein molecule presents yet another set of complicated challenges (Fig. 1) (*2*).

Faced with the impossibility of handling any such real system exactly, chemistry has evolved a series of approaches to the treatment of complex systems, which range from reasoning by analogy, through averaging, linearization, drastic approximation, and pure empiricism, to detailed analytical solution. The study of complexity in systems of reactions (or of processes or of properties) that can be described by nonlinear equations has been limited to the few that are both complex enough to be interesting and simple enough to be tractable. The emphasis in thinking about complicated systems has been to find methods that are predictive, even if they are nonanalytical. Philosophically, chemistry is a branch of science that attempts to predict and control rather than simply to observe and analyze: A large industrial reactor that produces heat in unpredictable bursts is more immediately terrifying than interesting. The optimization of combustion for the production of work, the understanding of mechanisms of drug action, and the development of strategies for organic synthesis are all problems of great complexity. They are also problems of sufficient urgency, which must be solved as best as possible, even if analytical solutions for them are not practical.

Chemistry is now evolving away from the manipulation of sets of individual molecules and toward the description and manipulation of systems of molecules, that is, living cells and materials. This evolution toward complexity is, perhaps counterintuitively, generating new types of problems that are sufficiently simple in some aspects for "complexity" in its analytical sense to provide a valuable way of thinking about them. These problems are often at the border between chemistry and other fields such as physics, biology, biophysics, and materials science. They may represent efforts to describe properties (for example, flux through a catalytic pathway in metabolism, distribution of greenhouse gases in the atmosphere, and fracture toughness of a polymer) that strongly depend on time, space, and conditions and in which the granularity of the description that is de-

Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: gwhitesides@gmwgroup.harvard.edu

Ricardo B. R. Azevedo

Statement

and

Readings

Emergence and Reductionism in Biology

Ricardo B. R. Azevedo

Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001

Most scientists confront the tension between reductionism and emergence in their research. Biologists are no exception. Indeed, they are especially sensitive to this tension because biological systems cover such a vast range of levels of complexity, in entities, patterns and processes: from macromolecules to ecosystems, from metabolism to animal communication, from organismal development to evolution. Biology defies reduction to physics because many of its central concepts, such as sexual reproduction and natural selection, are meaningless in physics. However, reductionism has served the biological sciences well. For example, major advances in developmental and evolutionary biology have relied on dramatic simplifications, namely, a naive view of gene action and what Mayr contemptuously called "beanbag genetics". But, while it is easy to criticize a reductionist approach from the "emergence" perspective, turning such criticisms into fruitful, alternative research programs has proved more challenging (e.g., developmental systems theory). Here I consider these issues in the context of my own research into the ways in which development influences the evolutionary process.

10 June 1977, Volume 196, Number 4295 SCIENCE

Evolution and Tinkering

François Jacob

Some of the 16th-century books devoted to zoology and botany are illustrated by superb drawings of the various animals that populate the earth. Certain contain detailed descriptions of such creatures as dogs with fish heads, men with chicken legs, or even women without heads. The notion of monsters that blend the characteristics of different species is not itself surprising: everyone has imagined or sketched such hybrids. What is disconcerting today is that in the 16th century these creatures belonged, not to the world of fantasies, but to the real world. Many people had seen them and described them in detail. The monsters walked alongside the familiar animals of everyday life. They were within the limits of the possible.

When looking at present-day science fiction books, one is struck by the same phenomenon: the abominable animals that hunt the poor astronaut lost on a distant planet are products of recombinations between the organisms living on the earth. The creatures coming from outer space to explore the earth are depicted in the likeness of man. You can watch them emerging from their unidentified flying objects (UFO's); they are vertebrates, mammals without any doubt, walking erect. The only variations concern body size and the number of eyes. Generally these creatures have larger skulls than humans, to suggest bigger brains, and sometimes one or two radioantennae on the head, to suggest very sophisticated sense organs. The surprising point here again is what is considered possible. It is the idea, more than a hundred years after Darwin, that, if life occurs anywhere, it is bound to produce animals not too different from the terrestrial ones; and above all to evolve something like man.

The interest in these monsters is that they show how a culture handles the possible and marks its limits. It is a requirement of the human brain to put order in the universe. It seems fair to say that all cultures have more or less succeeded in providing their members with a unified and coherent view of the world and of the forces that run it. One may disagree with the explanatory systems offered by myths or magic, but one cannot deny them unity and coherence. In fact, they are often charged with too much unity and coherence because of their capacity to explain anything by the same simple argument. Actually, despite their differences, whether mythic, magic, or scientific, all explanatory systems operate on a common principle. In the words of the physicist Jean Perrin, the heart of the problem is always "to explain the complicated visible by some simple invisible" (*1*). A thunderstorm can be viewed as a consequence of Zeus' anger or of a difference of potential between the clouds and the earth. A disease can be seen as the result of a spell cast on the patient or of an infection by a virus. In all cases, however, one watches the visible effect of some hidden cause related to the whole set of invisible forces that are supposed to run the world.

The World View of Science

Whether mythic or scientific, the view of the world that man constructs is always largely a product of imagination. For the scientific process does not consist simply in observing, in collecting data, and in deducing from them a theory. One can watch an object for years and never produce any observation of scientific interest. To produce a valuable observation, one has first to have an idea of what to observe, a preconception of what is possible. Scientific advances often come from uncovering a hitherto unseen aspect of things as a result, not so much of using some new instrument, but rather of looking at objects from a different angle. This look is necessarily guided by a certain idea of what the so-called reality might be. It always involves a certain conception about the unknown, that is, about what lies beyond that which one has logical or experimental reasons to believe. In the words of Peter Medawar, scientific investigation begins by the "invention of a possible world or of a tiny fraction of that world" (*2*). So also begins mythical thought. But it stops there. Having constructed what it considers as the only possible world, it easily fits reality into its scheme. For scientific thought, instead, imagination is only a part of the game. At every step, it has to meet with experimentation and criticism. The best world is the one that exists and has proven to work already for a long time. Science attempts to confront the possible with the actual.

The price to be paid for this outlook, however, turned out to be high. It was, and is perhaps more than ever, renouncing a unified world view. This results from the very way science proceeds. Most other systems of explanation—mythic, magic, or religious—generally encompass everything. They apply to every domain. They answer any possible question. They account for the origin, the present, and the end of the universe. Science proceeds differently. It operates by detailed experimentation with nature and thus appears less ambitious, at least at first glance. It does not aim at reaching at once a complete and definitive explanation of the whole universe, its beginning, and its present form. Instead, it looks for partial and provisional answers about those phenomena that can be isolated and well defined. Actually, the beginning of modern science can be dated from the time when such general questions as, "How was the universe created?

The author is a professor of cell genetics at the Institut Pasteur, 28 Rue du Dockteur Roux, 75015, Paris, France. This article is the text of a lecture delivered at the University of California, Berkeley, in March 1977.

What is matter made of? What is the essence of life?" were replaced by such limited questions as "How does a stone fall? How does water flow in a tube? How does blood circulate in vessels?" This substitution had an amazing result. While asking general questions led to limited answers, asking limited questions turned out to provide more and more general answers.

At the same time, however, this scientific method could hardly avoid a parceling out of the world view. Each branch of science investigates a particular domain that is not necessarily connected with the neighboring ones. Scientific knowledge thus appears to consist of isolated islands. In the history of sciences, important advances often come from bridging the gaps. They result from the recognition that two hitherto separate observations can be viewed from a new angle and seen to represent nothing but different facets of one phenomenon. Thus, terrestrial and celestial mechanisms became a single science with Newton's laws. Thermodynamics and mechanics were unified through statistical mechanics, as were optics and electromagnetism through Maxwell's theory of magnetic field, or chemistry and atomic physics through quantum mechanics. Similarly different combinations of the same atoms, obeying the same laws, were shown by biochemists to compose both the inanimate and the living worlds.

The Hierarchy of Objects

Despite such generalizations, however, large gaps remain, some of which probably will not be bridged for a long time, if ever. Today, there exists a series of sciences that differ, not only by the nature of the objects that are studied, but also by the concepts and the language that are used. These sciences can be arranged in a certain order—physics, chemistry, biology, psychosociology—an order that corresponds to the hierarchy of complexity found in the objects of these sciences. Following the line from physics to sociology, one goes from the simpler to the more complex objects and also, for obvious reasons, from the older to the younger science, from the poorer to the richer empirical content, as well as from the harder to the softer system of hypotheses and experimentation. In order to obtain a unified world view through science, the question has repeatedly been raised as to the possibility of making bridges between adjacent disciplines. Because of the hierarchy of objects, the problem is always to explain the more complex in terms and concepts applying to the simpler. This is the old problem of reduction, emergence, whole and parts, and so forth. Is it possible to reduce chemistry to physics, biology to physics plus chemistry, and so forth? Clearly an understanding of the simple is necessary to understand the more complex, but whether it is sufficient is questionable.

This type of question has resulted in endless arguments. Obviously, the two critical events of evolution—first the appearance of life and later that of thought and language—led to phenomena that previously did not exist on the earth. To describe and to interpret these phenomena, new concepts, meaningless at the previous level, are required. What can the notions of sexuality, of predator, or of pain represent in physics or chemistry? Or the ideas of justice, of increase in value or of democratic power in biology? At the limit, total reductionism results in absurdity. For the pretention that every level can be completely reduced to a simpler one would result, for example, in explaining democracy in terms of the structure and properties of elementary particles; and this is clearly nonsense.

This problem can be considered in a different way. One can look at the series of objects, moving from the simpler to the more complex. Molecules are made of atoms. They therefore obey the laws that determine the behavior of atoms. But, in addition, two statements can be made about molecules. First, they can exhibit new properties, such as isomerization, racemization, and so forth. Second, the subject matter of chemistry, the molecules found in nature or produced in the laboratory, represents only a small fraction of all the possible interactions between atoms. Chemistry constitutes, therefore, a special case of physics. This is even more so with biology that deals with a complex hierarchy of objects ranging from cells to populations and ecosystems. The objects which exist at each level constitute a limitation of the total possibilities offered by the simpler level. For instance, the set of molecules found in living organisms represents a very restricted range of chemical objects. At the next level, the number of animal species amounts to several millions; however, this is small relative to the number that could exist. All vertebrates are composed of a very limited number of cellular types, at most 200, such as muscle cells, skin cells, and nerve cells. The great diversity of vertebrates results from differences in the arrangement, in the number, and in the proportion of these 200 types. Similarly, the human societies with which ethnology and sociology deal represent only a restricted group of all possible interactions between human beings.

Constraints and History

Nature functions by integration. Whatever the level, the objects analyzed by natural sciences are always organizations, or systems. Each system at a given level uses as ingredients some systems of the simpler level, but some only. The hierarchy in the complexity of objects is thus accompanied by a series of restrictions and limitations. At each level, new properties may appear which impose new constraints on the system. But these are merely additional constraints. Those that operate at any given level are still valid at all more complex levels. Every proposition that is true for physics is also true for chemistry, biology, or sociology. Similarly every proposition that is valid for biology holds true in sociology. But as a general rule, the statements of greatest importance at one level are of no interest at the more complex ones. The law of perfect gases is no less true for the objects of biology or sociology than for those of physics. It is simply irrelevant in the context of the problems with which biologists, and even more so sociologists, are concerned.

This hierarchy of successive integrations, characterized by restrictions and by the appearance of new properties at each level, has several consequences. The first is the necessity of analyzing complex objects at all levels. If molecular biology, which presents a strong reductionist attitude, yielded such a successful analysis of heredity, it was mainly because, at every step, the analysis was carried out simultaneously at the level of the molecules and at the level of the black box, the bacterial cell. This applies also to recent developments in immunology. And it seems likely that such a convergence of analysis will play an important role in the study of human beings and their societies.

The second point concerns predictability. Is it possible to make predictions at one level on the basis of what is known at a simpler one? Only to a very limited extent. The properties of a system can be explained by the properties of its components. They cannot be deduced from them. Starting from fundamental laws of physics, there is no way of reconstructing the universe. This means that a

particular system, say a cell, has only a certain probability of appearing. All predictions about its existence can only be statistical. Molecular biology has shown that ultimately the characteristics of a cell rest on the structure of its molecular components. But the appearance of life on the earth was not the necessary consequence of the presence of certain molecular structures in prebiotic times. In fact, there is absolutely no way of estimating what was the probability for life appearing on earth. It may very well have appeared only once.

The third point concerns the nature of the restrictions and limitations found at every step of increasing complexity. Can one explain why, among all the possible interactions at one level, only certain are actually observed at the more complex one? How is it that only some types of molecular structures are present, for instance, in living organisms? Or only some interactions in human societies? There is no general answer to such questions, and it seems doubtful that there will ever be a specific answer for any one particular level of complexity. Complex objects are produced by evolutionary processes in which two factors are paramount: the constraints that at every level control the systems involved, and the historical circumstances that control the actual interactions between the systems. The combination of constraints and history exists at every level, although in different proportions. Simpler objects are more dependent on constraints than on history. As complexity increases, history plays a greater part. But history has always to be introduced into the picture, even in physics. According to present theories, heavier nuclei are composed of lighter ones and ultimately of hydrogen nuclei and neutrons. The transformation of heavy hydrogen into helium occurs during the fusion process, which is the main source of energy in the sun as well as in hydrogen bombs. Helium and all the heavier elements are thus the result of a cosmological evolution. According to present views, the heavier elements are considered as products of supernovae explosions. They seem to be very rare and not to exceed 1 or 2 percent by mass of all matter, while helium represents one-fifth and hydrogen four-fifths of all matter. The earth and the other planets of the solar system have thus been made of very rare material under conditions that seem to be rarely encountered in the cosmos. The source of hydrogen itself is left to theories and speculations concerning the origin of the universe.

Natural Selection

The constraints to which systems are subjected vary with the level of complexity. There are always some constraints imposed by stability and thermodynamics. But as complexity increases, additional constraints appear—such as reproduction for living systems, or economic requirements for social systems. Consequently, there cannot be any general law of evolution, any recipe that accounts for increasing complexity at all levels. Since Darwin, biologists have progressively elaborated a reasonable, although still incomplete, picture of the mechanism that operates in the evolution of the living world, namely, natural selection. For many, it has been tempting to invoke a similar mechanism of selection to describe any possible evolution, whether cosmological, chemical, cultural, ideological, or social. But this seems condemned to fail from the outset. The rules of the game differ at each level. New principles have, therefore, to be worked out at each level.

Natural selection is the result of two constraints imposed on every living organism: (i) the requirement for reproduction, which is fulfilled through genetic mechanisms carefully adjusted by special devices such as mutation, recombination, and sex to produce organisms similar, but not identical, to their parents; and (ii) the requirement for a permanent interaction with the environment because living beings are what thermodynamicists call open systems and persist only by a constant flux of matter, energy, and information. The first of these factors generates random variations and produces populations in which all individuals are different. The interplay of the two factors results in differential reproduction and consequently in populations that evolve progressively as a function of environmental circumstances, of behavior, and of new ecological niches. But natural selection does not act merely as a sieve eliminating detrimental mutations and favoring reproductions of beneficial ones as is often suggested. In the long run, it integrates mutations, and it orders them into adaptatively coherent patterns adjusted over millions of years, and over millions of generations as a response to environmental challenges. It is natural selection that gives direction to changes, orients chance, and slowly, progressively produces more complex structures, new organs, and new species. Novelties come from previously unseen association of old material. To create is to recombine.

Engineer and Tinkerer

The action of natural selection has often been compared to that of an engineer. This, however, does not seem to be a suitable comparison. First, because in contrast to what occurs in evolution, the engineer works according to a preconceived plan in that he foresees the product of his efforts. Second, because of the way the engineer works: to make a new product, he has at his disposal both material specially prepared to that end and machines designed solely for that task. Finally, because the objects produced by the engineer, at least by the good engineer, approach the level of perfection made possible by the technology of the time. In contrast, evolution is far from perfection. This is a point which was repeatedly stressed by Darwin who had to fight against the argument of perfect creation. In the *Origin of Species*, Darwin emphasizes over and over again the structural or functional imperfections of the living world. For instance, when he discusses natural selection (*3*, p. 472):

> Nor ought we to marvel if all the contrivances in nature be not, as far as we can judge, absolutely perfect. We need not marvel at the sting of the bee causing the bee's own death; at drones being produced in such vast numbers for one single act, and being then slaughtered by their sterile sisters; at the astonishing waste of pollen by our firtrees; at the instinctive hatred of the queen bee for her own fertile daughters; at ichneumonidae feeding within the live bodies of caterpillars; and at other such cases. The wonder indeed is, on the theory of natural selection, that more cases of the want of absolute perfection have not been observed.

There are innumerable statements of this type in the *Origin of Species*. In fact, one of the best arguments against perfection comes from extinct species. While the number of living species in the animal kingdom can be estimated to be around a few million, the number of extinct ones since life existed on earth has been estimated by Simpson (*4*) at around five hundred million.

Natural selection has no analogy with any aspect of human behavior. However, if one wanted to play with a comparison, one would have to say that natural selection does not work as an engineer works. It works like a tinkerer—a tinkerer who does not know exactly what he is going to produce but uses whatever he finds around him whether it be pieces of string, fragments of wood, or old cardboards; in short it works like a tinkerer who uses everything at his disposal to produce some kind of workable object. For the engineer, the realization

of his task depends on his having the raw materials and the tools that exactly fit his project. The tinkerer, in contrast, always manages with odds and ends. What he ultimately produces is generally related to no special project, and it results from a series of contingent events, of all the opportunities he had to enrich his stock with leftovers. As was discussed by Levi-Strauss (*5*), none of the materials at the tinkerer's disposal has a precise and definite function. Each can be used in a number of different ways. In contrast with the engineer's tools, those of the tinkerer cannot be defined by a project. What these objects have in common is "it might well be of some use." For what? That depends on the opportunities.

Evolution as Tinkering

This mode of operation has several aspects in common with the process of evolution. Often, without any well-defined long-term project, the tinkerer gives his materials unexpected functions to produce a new object. From an old bicycle wheel, he makes a roulette; from a broken chair the cabinet of a radio. Similarly evolution makes a wing from a leg or a part of an ear from a piece of jaw. Naturally, this takes a long time. Evolution behaves like a tinkerer who, during eons upon eons, would slowly modify his work, unceasingly retouching it, cutting here, lengthening there, seizing the opportunities to adapt it progressively to its new use. For instance, the lung of terrestrial vertebrates was, according to Mayr (*6*), formed in the following way. Its development started in certain freshwater fishes living in stagnant pools with insufficient oxygen. They adopted the habit of swallowing air and absorbing oxygen through the walls of the esophagus. Under these conditions, enlargement of the surface area of the esophagus provided a selective advantage. Diverticula of the esophagus appeared and, under continuous selective pressure, enlarged into lungs. Further evolution of the lung was merely an elaboration of this theme—enlarging the surface for oxygen uptake and vascularization. To make a lung with a piece of esophagus sounds very much like tinkering.

Unlike engineers, tinkerers who tackle the same problem are likely to end up with different solutions. This also applies to evolution, as exemplified by the variety of eyes found in the living world [see (*7*)]. It is obviously a great advantage under many conditions to possess light receptors, and the variety of photoreceptors in the living world is amazing. The most sophisticated are the image-forming eyes that provide information, not only on the intensity of incoming light, but also on the objects light comes from, on their shape, color, position, motion, speed, distance, and the like. Such sophisticated structures are necessarily complex. They can develop only in organisms already complex themselves. One might suppose, therefore, that there is just one way of producing such a structure. This is not the case. Eyes appeared a great many times in the course of evolution, based on at least three principles—pinhole, lens, and multiple tubes. Lens eyes, like ours, appeared both in mollusks and vertebrates. Nothing looks so much like our eye as the octopus eye. Both work in almost exactly the same way. Yet they did not evolve in the same way. Whereas in vertebrates photoreceptor cells of the retina point away from light, in mollusks they point toward light. Among all solutions found to the problem of photoreceptors, these two are similar but not identical. In each case, natural selection did what it could with the materials at its disposal.

Evolution does not produce novelties from scratch. It works on what already exists, either transforming a system to give it new functions or combining several systems to produce a more elaborate one. This happened, for instance, during one of the main events of cellular evolution: namely, the passage from unicellular to multicellular forms. This was a particularly important transition because it carried an enormous potential for a specialization of the parts. Such a transition, which probably occurred several times, did not require the creation of new chemical species, for there are no major differences between molecular types of uni- and multicellular organisms. It was mainly a reorganization of what already existed.

Molecular Tinkering

It is at the molecular level that the tinkering aspect of natural selection is perhaps most apparent. What characterizes the living world is both its diversity and its underlying unity. The living world contains bacteria and whales, viruses and elephants, organisms living at −20°C in polar areas and others at 70°C in hot springs. All these objects, however, exhibit a remarkable unity of chemical structures and functions. Similar polymers, nucleic acids or proteins, always made of the same basic elements, the four bases and the 20 amino acids, play similar roles. The genetic code is the same and the translating machineries are very nearly so. The same coenzymes mediate similar reactions. Many metabolic steps remain essentially the same, from bacteria to man. Obviously, for life to emerge, a number of new molecular types had first to be formed. During chemical evolution in prebiotic times and at the beginning of biological evolution, all those molecules of which every living being is built had to appear. But once life had started in the form of some primitive self-reproducing organism, further evolution had to proceed mainly through alterations of already existing compounds. New functions developed as new proteins appeared. But these were merely variations on previous themes. A sequence of a thousand nucleotides codes for a medium-sized protein. The probability that a functional protein would appear de novo by random association of amino acids is practically zero. In organisms as complex and integrated as those that were already living a long time ago, creation of entirely new nucleotide sequences could not be of any importance in the production of new information.

The appearance of new molecular structures during much of biological evolution must, therefore, have rested on alteration of preexisting ones. This is exemplified by the finding that large segments of genetic information, that is, of DNA, turn out to be homologous, not only in the same organism, but also among different organisms, even among those that are phylogenetically distant. Similarly, as more is known about amino acid sequences in proteins, it appears not only that proteins fulfilling similar functions in different organisms have frequently similar sequences, but also that proteins with different functions often exhibit rather large segments in common. The hypothesis most generally envisaged to account for these similarities was proposed by Horowitz (*8*), by Ingram (*9*), and by Ohno (*10*). A segment of DNA, corresponding to one or several genes, is assumed to be duplicated by some genetic mechanism. When a gene exists in more than one copy in a cell or a gamete, it is released from the contraints imposed on functions by natural selection. Mutations can then accumulate more or less freely and result in modified protein structures, some of which can eventually fulfill new functions. Since natural selection exerts a continual pressure on organisms, an alteration in a protein can be further improved by other, later changes. It can also lead to a perturbation in the interactions with other proteins and eventually favor modifications of these proteins. A large fraction

of the genome of complex organisms might actually derive from a few ancestral genes.

Biochemical changes do not seem, therefore, to be a main driving force in the diversification of living organisms. The really creative part in biochemistry must have occurred very early. For the biochemical unity that underlies the living world makes sense only if most of the important molecular types found in organisms, that is, most of the metabolic pathways involved in the production of energy and in biosynthesis or degradation of the essential building blocks already existed in very primitive organisms such as bacteria. Once this stage passed, biochemical evolution continued as more complex organisms appeared. But it is not biochemical novelties that generated diversification of organisms. In all likelihood, it worked the other way around. It is the selective pressure resulting from changes in behavior or in ecological niches that led to biochemical adjustments and changes in molecular types. What distinguishes a butterfly from a lion, a hen from a fly, or a worm from a whale is much less a difference in chemical constituents than in the organization and the distribution of these constituents. The few big steps of evolution required acquisition of new information. But specialization and diversification occurred by using differently the same structural information. Among neighboring groups, vertebrates for instance, chemistry is the same. What makes one vertebrate different from another is a change in the time of expression and in the relative amounts of gene products rather than the small differences observed in the structure of these products. It is a matter of regulation rather than of structure [see (*11*)].

After egg fertilization, embryonic development occurs in a fixed order and according to a precise schedule set by the genetic program contained in the chromosomes. This program determines when and where lines of differentiated cells will emerge, when and where different proteins will be made and in what amounts. Both the quality and quantity of the different proteins vary in time and space during development. Thus in the adult, the various types of cells or tissues contain different repertoires of molecular types in agreement with their functions. The genetic program is executed through complex regulatory circuits that switch the different biochemical activities of the organism on or off. Very little is known as yet about the regulatory circuits that operate in the development of complex organisms. It is known, however, that, among related organisms such as mammals, the first steps of embryonic development are remarkably similar, with divergences showing up only progressively as development proceeds. These divergences concern much less the actual structure of cellular or molecular types than their number and position. It seems likely that divergence and specialization of mammals, for instance, resulted from mutations altering regulatory circuits rather than chemical structures. Small changes modifying the distribution in time and space of the same structures are sufficient to affect deeply the form, the functioning, and the behavior of the final product—the adult animal. It is always a matter of using the same elements, of adjusting them, of altering here or there, of arranging various combinations to produce new objects of increasing complexity. It is always a matter of tinkering.

Consequences of Tinkering

Marks of this tinkering are thus found at every level thoughout the living world. Of course, they can be found in human beings as shown by the following few examples. In humans, as in many mammals, there exist very complex processes responsible for such functions as blood coagulation, inflammatory reactions against foreign bodies, and the immunological defenses mediated by the so-called complement system. These three processes have been independently analyzed in some detail during recent years. Each one exhibits an unexpected complexity. Each involves about ten proteins, none of which initially has enzymatic activity. Conversion of the first protein into a catalytically active form triggers a cascade of reactions. The first protein cleaves the second one at a specific point; a product of this reaction cleaves the third protein, and so on. In this series of reactions, the individual proteins are thus split in sequence and the released fragments serve as activators, or inhibitors, in other reactions of the chain. Furthermore, these three chains of reactions are not wholly independent. A product of cleavage in one chain can suddenly become an active element in another chain or even play a role in a completely different process. These products may serve as signals to connect chemically unrelated, but physiologically dependent, systems. It is as though some protein molecules, which happened to be formed, were used here or there as a source of smaller but active peptides as new functions were taking shape. Recently, a number of peptides of different sizes have been found to participate in a variety of physiological processes. Some of them, such as hormone peptides or brain peptides, are known not to be chemically transformed in the reaction they activate or inhibit. They appear just to bind to some protein to favor an allosteric transition, thus acting as simple chemical signals. For the biologist, it is thus generally impossible to make a prediction, or even an inspired guess, about the nature of such molecules and their structural relations with other constituents. All he can do is to detect them, purify them, and analyze them. Later, as the structures of more proteins become known, there will perhaps be a chance to define the functional interrelations and evolutionary relationship among such molecules.

Another example of tinkering can be found in early human embryonic development. Embryonic development is a tremendously complicated process of which little is known at present. Studies of the past 10 or 20 years have revealed an amazing phenomenon. In various human populations, 50 percent of all conceptions are estimated to result in spontaneous abortion [see (*12*)]. A large fraction of these abortions occur during the first 3 weeks of pregnancy and generally pass unnoticed. Thus, in half of the total conceptions, something is wrong to begin with. Many of these spontaneous abortions appear to be due to an odd number of chromosomes; instead of having one set of chromosomes derived from its mother and one from its father, the embryo lacks a chromosome, or has an extra one, or even has three sets instead of two. As a result, some functions necessary to embryonic development are not performed correctly. The fetus dies and is expelled. Thus many potentially malformed fetuses disappear; not all, unfortunately, since some of them still come to term. This reveals the imperfections of a mechanism that is at the very core of any living system and that has been refined over millions of years.

A third example of tinkering which is very intriguing when one thinks about it is the association between reproduction and what is generally called pleasure. Sex is one of the most efficient inventions of evolution. In lower organisms which apparently reproduce asexually by fission, the genetic program is scrupulously recopied at every generation. Within a population, it always remains the same, except for rare mutations. Division of the organism is an automatic process resulting from growth. When something resembling sexuality exists, as in bacteria, it is a luxury. In such pop-

ulations, adaptation necessarily involves the selection of rare mutants under environmental conditions. In contrast, sexual reproduction, which probably occurred early in evolution, compels reassortment of genetic programs in interbreeding populations. As a result, every genetic program (that is, every individual) is different from the others. This permanent reshuffling of genetic elements provides tremendous potentialities of adaptation. But once sexuality had become a necessary condition for reproduction, it required special mechanisms: one, allowing individuals of opposite sexes to recognize and meet each other and a second, driving them to unite. The first of these requirements has been fulfilled by a variety of specific signaling systems—visual, auditory, or olfactory—of amazing precision and efficiency. The second has been met through the development of genetically determined and very rigid programs of behavior. For instance, in birds, at the proper season, the view of an individual of the opposite sex initiates a whole process of rituals, courtship, and parade leading almost automatically to copulation, nidation, and progeny care. The course of evolution, however, is characterized by a trend to greater flexibility in the execution of the genetic program. As this program became more open, so to speak, the behavior became less rigidly determined by the genes. Reactions to sexual signals were no longer completely automatic. In order to drive the individuals toward reproduction, sexuality had therefore to be associated with some other devices. Among these was pleasure. In the Oxford dictionary, pleasure is defined as "the opposite of pain," obviously, but also as "the condition of consciousness induced by the enjoyment of what is felt or viewed as good or desirable." It seems likely that feelings of discomfort and pleasure must already have existed for a long time in complex animals. An animal is more likely to have progeny if a feeling of discomfort dissuades it from entering harmful situations. It is clear that the existence of nervous centers, connected with sense organs and able to correlate what is felt as pleasant or unpleasant with what is actually good or bad for survival, is of great selective value. In fact, such centers are now known to exist. Some 20 years ago, neurobiologists detected in the brain, first in the rat and later in many vertebrates, the presence of two remarkable centers—one called the center of aversion and the other called the center of autostimulation. Fitted with correctly implanted electrodes and given the means of activating at will the latter center, a rat gives himself pleasure until it collapses from sheer exhaustion. Experiments performed during brain surgery and descriptions of feelings by the patients leave very little doubt as to the existence of such centers in man and to its association with sexual activity. Thus pleasure appears as a mere expedient to push individuals to indulge in sex and therefore to reproduce. A rather successful expedient indeed, as judged by the state of the world population.

A Final Example of Tinkering: The Human Brain

Although our brain represents the main adaptive feature of our species, what it is adapted to is not clear at all. What is clear, however, is that, like the rest of our body, our brain is a product of natural selection, that is, of differential reproductions accumulated over millions of years under the pressure of various environmental conditions. Our brain has therefore evolved at our gonad's service, as already emphasized by Freud many years ago. But curiously enough, brain development in mammals was not as integrated a process as, for instance, the transformation of a leg into a wing. The human brain was formed by superposition of new structures on old ones. To the old rhinencephalon of lower mammals a neocortex was added that rapidly, perhaps too rapidly, took a most important role in the evolutionary sequence leading to man. For some neurobiologists, especially McLean (*13*), these two types of structures correspond to two types of functions but have not been completely coordinated or hierarchized. The recent one, the neocortex, controls intellectual, cognitive activity. The old one, derived from the rhinencephalon, controls emotional and visceral activities. In contrast to the former, the latter does not seem to possess any power of specific discrimination, or any capacity for symbolization, language, or self-consciousness. The old structure which, in lower mammals, was in total command has been relegated to the department of emotions. In man, it constitutes what McLean calls "the visceral brain." Perhaps because development is so prolonged and maturity so delayed in man, these centers maintain strong connections with lower autonomic centers and continue to coordinate such fundamental drives as obtaining food, hunting for a sexual partner, or reacting to an enemy. This evolutionary procedure—the formation of a dominating neocortex coupled with the persistence of a nervous and hormonal system partially, but not totally under the rule of the neocortex—strongly resembles the tinkerer's procedure. It is somewhat like adding a jet engine to an old horse cart. It is not surprising, in either case, that accidents, difficulties, and conflicts can occur.

It is hard to realize that the living world as we know it is just one among many possibilities; that its actual structure results from the history of the earth. Yet living organisms are historical structures: literally creations of history. They represent, not a perfect product of engineering, but a patchwork of odd sets pieced together when and where opportunities arose. For the opportunism of natural selection is not simply a matter of indifference to the structure and operation of its products. It reflects the very nature of a historical process full of contingency.

As Simpson (*4*) pointed out, the interplay of local opportunities—physical, ecological, and constitutional—produces a net historical opportunity which in turn determines how genetic opportunities will be exploited. It is this net historical opportunity that mainly controls the direction and pace of adaptive evolution. This is why the probability is practically zero that living systems, which might well exist elsewhere in the cosmos, would have evolved into something looking like human beings. Even if life in outer space uses the same material as on the earth, even if the environment is not too different from ours, even if the nature of life and of its chemistry strongly limits the way to fulfill certain functions, the sequence of historical opportunities there could not be the same as here. A different play had to be performed by different actors. Despite science fiction, Martians cannot look like us. And we might as well have looked like one of those 16th-century monsters.

References and Notes

1. J. Perrin, *Les Atomes* (Alcan, Paris, 1914).
2. P. B. Medawar, *The Hope of Progress* (Doubleday, New York, 1973).
3. C. Darwin, *On the Origin of Species* (London, 1859).
4. G. G. Simpson, *Evolution* **6**, 342 (1952).
5. C. Levi-Strauss, *La Pensée Sauvage* (Plon, Paris, 1962).
6. E. Mayr. *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **23**, 1231 (1964).
7. G. G. Simpson, *The Meaning of Evolution* (Yale Univ. Press, New Haven, Conn., 1967).
8. N. Horowitz, *Adv. Genet.* **3**, 33 (1950).
9. V. M. Ingram, *Hemoglobins in Genetics and Evolution* (Columbia Univ. Press, New York, 1963).
10. S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, New York, 1970).
11. M. C. King and A. C. Wilson, *Science* **188**, 107 (1975).
12. A. Boue and J. G. Boue, in *Physiology and Genetics of Reproduction*, E. Coutinho and F. Fuchs, Eds. (Plenum, New York, 1975), vol. 4b, p. 317.
13. P. McLean, *Psychosom. Med.* **11**, 338 (1949).

"Less Is More" and the Art of Modeling Complex Phenomena

Simplification May But Need Not Be the Key to Handle Large Networks

PETER SCHUSTER

Peter Schuster,
Editor in Chief of Complexity,
is at the Institut für Theoretische Chemie der Universität Wien,
A-1090 Wien, Austria;
E-mail: pks@tbi.univie.ac.at

In the October 21, 2005 issue of the magazine *Science* a perspectives article by Stefan Bornholdt addresses the problem of the proper level of details in the description of complex systems [1]. "Less is more" is used in this article to encourage the usage of highly simplified dynamical elements for modeling large and complex nonlinear systems. The special case considered is modeling of large genetic networks, but the problem is much more general, arises often in science and beyond and, perhaps, deserves broader attention. Some older examples of "less is more" that have already reached a certain degree of maturity and common acceptance may be useful with respect to the recent revival of this paradigm. I shall present here two different problems from the interface of physics and chemistry, which were heavily debated in the past and for which consensus has been achieved by now, before returning to the burning biological questions.

Some older examples of "less is more" that have already reached a certain degree of maturity and common acceptance may be useful with respect to the recent revival of this paradigm.

The first example starts with Paul Dirac's famous comment [2], "… 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. …" The reactions of the scientific public were extremely ambiguous: Quantum chemists used Dirac's statement as the figure-head for their many decades long search for better and better approximations to the Schrödinger equations, whereas the majority of experimental chemists were truly upset. The reason for the uneasiness also shared by other nonphysicists was certainly not only the overstatement of a then 27-year-old and somewhat arrogant physicist but also an intuitive feeling that quantum mechanics provides the tool to reduce and integrate chemistry into physics, thereby sacrificing chemistry's autonomy [3, 4]. Putting aside the philosophical questions, we are left with a pragmatic problem: Is quantum mechanics appropriate to describe molecules and chemical bonds for the chemist at the workbench? There are two reasons, among others, that suggest the application of a different, preferentially less sophisticated level of description: (i)

 Vol. 11, No. 2
DOI 10.1002/cplx.20107

Despite the spectacular progress of computational quantum mechanics that allows for incredibly accurate computation of structures and properties of small molecules [5, 6], calculations of large molecules are still far away from being satisfactory and, even more important, the predictive power of the full-blown quantum chemical approaches is rather week. In other words, we can calculate but we don't understand unless we crank up the highly sophisticated and costly computational machinery derived from the Schrödinger equation. (ii) Empirical chemical knowledge unlike the usage of the stationary Schrödinger equation unconsciously involves a time span of observation, and this matters when we discuss what we mean by chemical compounds. To give a naïve but illustrative example, dimethyl ether and ethanol, CH_3OCH_3 and CH_3CH_2OH, respectively, are two distinct chemical compounds, although they have the same Hamiltonian and are described by the same Schrödinger equation. Clearly, computing the energy landscape in the Born-Oppenheimer approximation and estimating the lifetime of the two isomers of C_2H_6O will undoubtedly reveal that the time scale for an interconversion of the two molecules is extremely long, and therefore we are not in danger that the ether is converted into more stable ethanol during an experiment. There are other pairs of isomers, for example, in nonclassical carbocations [7] that change structures too fast to be observed. Qualitative molecular theory, being a largely simplified and coarse-grained distillation of quantum mechanics for chemists, is often useful and makes successful predictions, although it is lacking the solid anchor in physics. The hybridization concept, for example, allows for many correct predictions of rough molecular geometries and the Walsh rules are likewise successful. The Woodward-Hoffman rules are a valuable tool for predicting the reactivity in certain classes of reactions. There is, however, one important fact to keep in mind: The qualitative picture fails inevitably when quantum mechanics is truly indispensable as it is, for example, in spectroscopy, in photochemistry, and in interactions of electromagnetic radiation with matter, in general. In addition there is no way around large-scale computation, if we are heading for predictions of quantitatively reliable and sufficiently precise results.

Qualitative molecular theory, being a largely simplified and coarse-grained distillation of quantum mechanics for chemists, is often useful and makes successful predictions, although it is lacking the solid anchor in physics.

My second case is another example from the crossroads of chemistry and physics: chemical reaction dynamics [8, 9]. Chemical reactions are commonly described at three main levels of sophistication [9]: (i) the qualitative or substance level, "What produces what under which conditions?"; (ii) the elementary step level, dealing with rate constants and their dependence on parameters like temperature, pressure, and other external conditions; and (iii) the full-blown chemical dynamics level where the interconversion of molecules are resolved to individual reactive collisions. Level (iii) provides marvelous insights into unexpected details of reactive quantum scattering and creates the link to computational quantum chemistry discussed in the previous paragraph. To mention just one illustration of such details: Simple reactive collisions, like $F + H_2 \rightarrow FH + H$, involve several atomic and molecular states and may be calculated now by quantum scattering techniques on multiple energy surfaces [8] and then, the computed results agree with molecular beam experiments. Paul Crutzen, who did the epoch-making studies on atmospheric chemistry and, in particular, ozone destruction by manmade pollutants [10], would have been completely lost if he had attempted to reach his goal on the quantum scattering level. All his success was based on consequent and precise level (ii) kinetic studies on vapor phase reactions. Sometimes even cruder descriptions between the elementary step resolution and the qualitative description are important. An illustrative example is the beautiful work on nonlinear chemical reactions in solution [11]. The famous oscillatory Belousov-Zhabotinskii reaction comprises some 20 or more elementary steps. In the "Oregonator" model developed by Richard Field and Richard Noyes [12, 13] these steps are cast into five overall reactions that allow for a perfect and accurate prediction of the course of the reaction and even of very subtle reaction details. Many examples could be added, which all demonstrate the more or less self-evident but nevertheless often forgotten fact: The proper model description of a complex system depends on both the context of the problem and the question one wants to ask.

Coming back to the initial problem concerning the proper level of description for complex biological networks we recognize a situation that is not very different from the two examples mentioned above. There are, for example, several levels of description for neural networks, I shall mention here only two of them: (i) The single neuron level, which is described in great detail by the famous Hodgkin-Huxley equation relating action potential and electric current in the neuron [14], and (ii) the highly coarse-grained level of neural networks that initiated a whole new area of computation (see, for example, the Hopfield networks [15]). At present both levels are still highly relevant: Level (i), because progress in the molecular biology of the neuron allows for a precise characterization of the molecular players in the Hodgkin-Huxley equation and calls for extensions of the original version to more realistic gating models, and level (ii), because we are still lacking a comprehensive theory for the emergence of collective properties in neural networks, in particular in the brain. With the current computational facilities it is also thinkable to combine both levels and to compute relatively large ensembles of Hodgkin-Huxley neurons. Here as well as above we have to face the "too much detailed" problem at the molecular

level. On the other hand, when we are focusing on the role of individual classes of ion gates and specific neurotransmitters, the molecular level is indispensable.

Genetic and metabolic networks—*genabolic* networks might be a good name for the combination of both—are no exceptions of the rule [16, 17]. There are features for which the description by means of Boolean functions, as advertised in [1], is the most appropriate level to learn generic properties of signaling and regulation. When it comes to other questions, for example, the control of cellular activities by second messengers and hormones, the molecular level will be essential. The entire disciplines of computational systems biology and cell biology are in an exciting and very fast development. Despite impressive progress in the past few years several techniques have yet to be established and large-scale computations on the dynamics of whole cells and organisms will be impossible without specific advances in algorithms and their implementations. I see parallels to the development in computational chemistry, where the scientific questions were indeed completely forgotten for a few decades and people focused almost exclusively on the solution of computational problems. During such periods of technical progress, a reminder like Stefan Bornholdt's perspective that suggests not to forget the ultimate goals and to think about simpler approaches is undoubtedly in place.

The proper model description of a complex system depends on both the context of the problem and the question one wants to ask.

Is less more? The answer to the question, as I wanted to point out here, is subtle. It is "could be" rather than "yes," and whether or not it is true depends on the context and the problem to be investigated. The figure in the article in *Science* [1] distinguishes nicely four levels of description—single gene, small genetic circuits, medium-size, and large-scale genetic networks—and I think each one is justified in its own right. The art of modeling is to choose the proper degree of detail. One take-home lesson from the development of computational quantum chemistry, however, is that decades of methodological development, where everyone in the field focuses on the problem to compute faster and faster, larger and larger systems, may pay at the end, after the technical problems had been solved and the scientific questions come back into the focus of interest.

REFERENCES

1. Bornholdt, S. Less is more in modeling large genetic networks. Science 2005, 310, 449–450.
2. Dirac, P.A.M. Quantum mechanics of many-electron systems. Proc Roy Soc Lond 1929. A123, 714–733.
3. Benfey, T. Reflections in the philosophy of chemistry and a rallying call for our discipline. Found Chem 2000, 2, 195–205.
4. Lombardi, O.; Labarca, M. The ontological autonomy of the chemical world. Found Chem 2005, 7 125–148.
5. Pople, J.A. Quantum chemical models (Nobel lecture). Angew Chem Int Edit 1999, 38, 1894–1902.
6. Kohn, W. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. Rev Mod Phys 1999, 71, 1253–1266.
7. von Ragué Schleyer, P.; Maerker C. Exact structures of carbocations established by combined computational and experimental methods. Pure Appl Chem 1995, 67, 755–760.
8. Althorpe, S.C.; Clary, D.C. Quantum scattering calculations on chemical reactions. Annu Rev Phys Chem 2003, 54, 493–529.
9. Herschbach, D.R. Molecular dynamics of elementary chemical reactions (Nobel lecture). Angew Chem Int Edit 1987, 26, 1221–1243.
10. Crutzen, P.J. My life with O_3, NO_x and other YZO_x compounds (Nobel lecture). Angew Chem Int Edit 1996, 35, 1758–1777.
11. Sagués, F.; Epstein, I.R. Nonlinear chemical dynamics. Dalton Trans 2003, 1201–1217.
12. Field, R.; Körös, E.; Noyes, R.M. Oscillations in chemical systems. II. Thorough analysis of temporal oscillations in the bromate-cerium-malonic acid system. J Am Chem Soc 1972, 94, 8649–8664.
13. Field, R.; Noyes, R.M. Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. J Chem Phys 1972, 60, 1877–1884.
14. Hodgkin, A.L.; Huxley, A.F. A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 1952, 117, 500–544.
15. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 1982, 79, 2554–2558.
16. Ptashne, M.; Gann, A. Genes & signals. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2002.
17. Fell, D. Understanding the control of metabolism. Portland Press Ltd.: London, 1997.

LETTERS

Sexual reproduction selects for robustness and negative epistasis in artificial gene networks

Ricardo B. R. Azevedo[1], Rolf Lohaus[1], Suraj Srinivasan[1], Kristen K. Dang[2] & Christina L. Burch[3]

The mutational deterministic hypothesis for the origin and maintenance of sexual reproduction posits that sex enhances the ability of natural selection to purge deleterious mutations after recombination brings them together into single genomes[1]. This explanation requires negative epistasis, a type of genetic interaction where mutations are more harmful in combination than expected from their separate effects. The conceptual appeal of the mutational deterministic hypothesis has been offset by our inability to identify the mechanistic and evolutionary bases of negative epistasis. Here we show that negative epistasis can evolve as a consequence of sexual reproduction itself. Using an artificial gene network model[2,3], we find that recombination between gene networks imposes selection for genetic robustness, and that negative epistasis evolves as a by-product of this selection. Our results suggest that sexual reproduction selects for conditions that favour its own maintenance, a case of evolution forging its own path.

A century of genetic research has revealed two general properties of spontaneous mutations with detectable effects on fitness: most of them are deleterious, and they frequently interact with each other[4,5]. Many types of interactions are possible, including directional epistasis, in which the average effect of spontaneous mutations changes in the presence of other mutations in the genome[6]. Directional epistasis can be either negative (synergistic) or positive (antagonistic), depending on whether the average effect of mutations becomes more or less harmful, respectively, as the number of other mutations in the genome increases (Fig. 1). Directional epistasis holds particular interest for evolutionary biologists because it is expected to determine the outcome of multiple evolutionary processes, notably the evolution of sex and recombination[1]. Empirical studies on a variety of organisms have reported every conceivable form of directional epistasis: negative[7–9], positive[6,10] and no significant directional epistasis[11,12]. These mixed results have not helped to clarify either the mechanistic or evolutionary causes of directional epistasis[13].

In contrast, evolutionary simulations using computational models of RNA secondary structure[14], viral replication[15] and artificial life[14] have demonstrated that the average strength and direction of epistasis can be shaped by natural selection. One mechanism by which epistasis evolves in these models[13] is through a negative correlation among genotypes between the extent of genetic robustness (or genetic canalization, measured as the insensitivity of a phenotype to mutation) and the direction of epistasis. As a consequence, selection for higher robustness produces a correlated response in the strength of epistasis in all three models, towards either weaker positive or stronger negative epistasis[14,15]. The repeatability of this result in models of different biological systems suggests that the strength and direction of epistasis observed in living organisms depend on their history of selection for genetic robustness.

Theory predicts that traits can evolve to be robust to genetic perturbations (that is, mutation and recombination) under a variety of selective regimes[16–18], as long as the following two conditions are met: genes must interact to determine the trait[17–19], and the population must contain sufficient genetic variation[18]. Whereas the former condition is inherent to particular organisms, the latter condition will depend on population genetic parameters such as the mutation and recombination rates. Experimental tests of these predictions using computational models confirm that high mutation rates, such as those experienced by RNA viruses, favour the evolution of genetic robustness[2,3,18,20]. Sexual reproduction (that is, increased recombination) is also expected to impose stronger selection for genetic robustness than asexual reproduction[21,22], but this hypothesis has never been tested experimentally[21].

To test this hypothesis, and to determine whether the evolution of genetic robustness is accompanied by the evolution of negative epistasis, we return to the computational model of genetic networks used in two previous studies[2,3]. We chose this model primarily because it explicitly incorporates one of the key characteristics required for the evolution of robustness[17–19]—genetic interactions. Furthermore, empirical data from biological systems has consistently suggested that extant gene networks are robust to changes in biochemical rate parameters and levels of gene activity[19,23]. Previous work with this model has shown that genetic robustness (again, measured as robustness to mutation) evolves readily if networks are subjected to selection for the production of a stable gene expression pattern[2,3]. Here we explore the extent to which recombination contributed to the evolution of genetic robustness in this model, and ask whether recombination, through its effect on robustness[2,21,22], can cause the direction of epistasis to evolve.

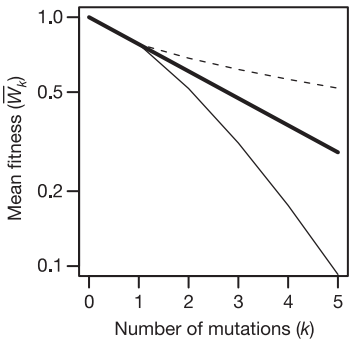


Figure 1 | Types of directional epistasis for deleterious mutations. Three hypothetical relationships between fitness (log scale) and number of deleterious mutations are plotted. All relationships depicted have the same mutational robustness ($\overline{W}_1 = 0.78$) but different directions of epistasis: negative epistasis (plain line, concave downwards; $1 - \beta < 0$), no directional epistasis (bold, straight line; $1 - \beta = 0$) and positive epistasis (dashed line, concave upwards; $1 - \beta > 0$).

[1]Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA. [2]Department of Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7575, USA. [3]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3280, USA.

Briefly, the model[2,3] represents individuals as networks of N interacting transcriptional regulators (Fig. 2a). The genotype of an individual is represented by an $N \times N$ matrix R, whose elements r_{ij} describe the regulatory effect of the product of gene j on the expression of gene i (Fig. 2b). The number of regulatory interactions is determined by a connectivity parameter (c) that specifies the proportion of non-zero matrix elements. This matrix of regulatory relationships acts on gene expression patterns, which are represented by a state vector $\mathbf{S}(t)$, whose elements $s_i(t)$ describe the expression states of genes $i = 1, 2, \ldots, N$ at time t. The expression state of a gene can vary continuously between complete repression, $s_i(t) = -1$, and complete activation, $s_i(t) = 1$. Gene expression states change over time according to the following equation:

$$s_i(t+1) = f\left[\sum_{j=1}^{N} r_{ij} s_i(t)\right] \quad (1)$$

where $f(x)$ is a sigmoidal filter function[2] that determines how the total regulatory input influences gene expression (see Supplementary Fig. 1a). Development is modelled as the progression from an initial gene expression state to an equilibrium gene expression pattern (Fig. 2c; see Methods). In this model, genotypes that achieve any stable, fixed-point equilibrium expression pattern are considered developmentally stable[2], and therefore viable. Genotypes that do not achieve a stable equilibrium (for example, oscillatory gene expression) are considered inviable[2].

The mechanistic underpinnings of this model allow *a priori* predictions about the evolution of genetic robustness. In this model, the reproductive success of a viable genotype (that is, its fitness) is the proportion of its offspring that are also viable. When offspring are produced asexually, and differ from their parents only by mutations, the fitness of a genotype is given by:

$$W_{\text{asex}} = \sum_{k=0}^{cN^2} \phi_k \overline{W}_k \quad (2)$$

where $\phi_k = \mu^k e^{-\mu}/k!$ is the Poisson probability that offspring acquire k mutations when the mutation rate per individual network per generation is μ, and $\overline{W}_k$ is the mean fitness of the genotype after the addition of k mutations. For mutation rates $\mu \leq 0.1$, terms for $k > 1$ can be effectively ignored, so that $W_{\text{asex}} \approx 1 - \mu + \mu\overline{W}_1$. $\overline{W}_1$ is our measure of mutational robustness, the probability that a genotype with one mutation is viable. Therefore, given a sufficiently high mutation rate, asexually reproducing networks are expected to evolve mutational robustness. In contrast, when offspring are produced via sexual reproduction, and differ from their parents primarily as a result of recombination, the fitness of a genotype is given by $W_{\text{sex}} = W_{\text{asex}}(1 - L)$, where L is the probability that a mating with a random individual in the population will result in an inviable offspring, a measure of the recombination load[24]. Therefore, sexual populations should experience selection for two distinct types of genetic robustness: mutational robustness and recombinational robustness.

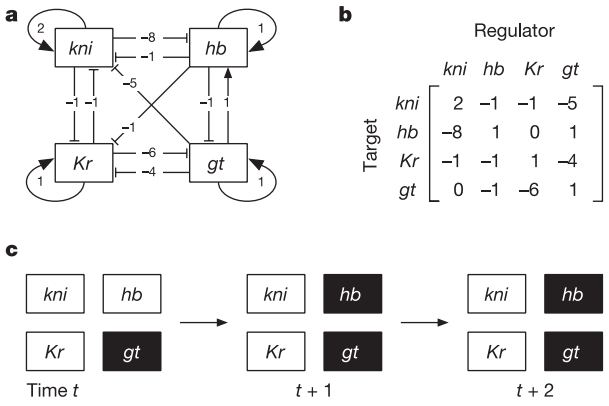


Figure 2 | Application of our network model to the gap gene system of *Drosophila melanogaster*. **a**, Network representation of the regulatory interactions between four gap genes[29] (*gt*, *giant*; *hb*, *hunchback*; *kni*, *knirps*; *Kr*, *Krüppel*). Activations and repressions are denoted by arrows and bars, respectively. Numbers indicate the relative interaction strengths[30]. **b**, Interaction matrix (R) representing the network in **a**. The element in row i and column j (r_{ij}) denotes the regulatory effect of the product of gene j on the expression of gene i. **c**, Graphical representation of the gene expression states of each gap gene over three successive time steps. For the purpose of this illustration we consider gene i to be ON (filled box) if $s_i(t) > 0$, and OFF (open box) if $s_i(t) \leq 0$. The change in gene expression pattern matches events at ~80% anterior–posterior position in the *Drosophila* embryo between early and mid cleavage cycle 14A (ref. 29). Successive iterations beyond the $t + 1$ step do not change the gene expression pattern, the hallmark of a stable equilibrium.

In order to explore the effect of sexual reproduction on the evolution of genetic robustness, we first investigated the behaviour of this model using conditions that are known to produce robustness to mutation[2] ($\mu = 0.1$, $N = 10$ genes, $c = 0.75$; see Methods), varying only the reproductive mode from sexual to asexual. Fifty clonal populations of 500 individuals were founded by different randomly generated, viable genotypes. Each population was subjected to selection for the ability to produce a stable gene expression pattern and allowed to evolve separately via sexual and asexual reproduction. We monitored evolution until an equilibrium level of mutational robustness was achieved. Contrary to earlier claims[2,3], our simulations show that sexual reproduction has a substantive effect on the evolution of mutational robustness (Fig. 3a). Although mutational robustness increased in asexual populations, it reached a significantly lower equilibrium value than in sexual populations (paired t-test: $t = 31.0$, 49 degrees of freedom (d.f.), $P < 0.0001$). An investigation of epistasis in these evolved populations revealed that sexual reproduction also had a qualitative effect on the evolution of directional epistasis ($t = 23.6$, 49 d.f., $P < 0.0001$). The magnitude of epistasis evolved regardless of reproductive mode, but the direction of epistasis only changed when reproduction was sexual. At equilibrium, asexual populations exhibited average positive epistasis of a reduced magnitude, whereas sexual populations exhibited negative epistasis.

Why does sexual reproduction cause the evolution of increased mutational robustness? Sexual reproduction is not expected to increase the strength of selection for mutational robustness directly. However, it is expected to select for recombinational robustness, and this could cause a correlated response in mutational robustness. We devised two experiments to test this hypothesis. In the first experiment, we investigated whether the effect of sexual reproduction on mutational robustness depended on the high mutation rate ($\mu = 0.1$). Theory predicts that mutational robustness will evolve through the direct action of selection only if μ is greater than the reciprocal of the effective population size[7] (that is, $\mu > 0.002$ in our simulations). Thus, we tested our hypothesis by re-running the simulations at a mutation rate of $\mu = 0.002$ (Fig. 3a). At this low mutation rate, mutational robustness failed to evolve in asexual populations within 50,000 generations. However, mutational robustness did increase significantly in sexual populations within 20,000 generations. The inability of asexual populations to respond to selection for mutational robustness confirms that selection acting directly on mutational robustness is ineffective when $\mu = 0.002$. Thus, the mutational robustness that evolved in these sexual populations did not evolve through the direct action of selection. Rather, it must have evolved as a correlated response to selection for recombinational robustness, the only other source of selection in these simulations.

In the second experiment we constructed genetically variable populations (see Supplementary Methods) and allowed them to evolve in the absence of new mutations ($\mu = 0$), that is, in the absence of selection for mutational robustness. In this experiment,

sexual populations showed significant increases in mutational robustness, whereas asexual populations did not (Fig. 3b). These results confirm our hypothesis because sexual and asexual populations differed only by the presence and absence, respectively, of selection for recombinational robustness. By manipulating the amount of genetic variation present in the founder populations, we also showed that the evolutionary response in mutational robustness increased with the strength of selection for recombinational robustness (that is, with the magnitude of the recombination load, L; Supplementary Fig. 2).

Directional epistasis evolved in both of these experiments (Fig. 3) in a similar manner to the initial simulations. Conditions that showed no change in mutational robustness also showed no change in directional epistasis. However, conditions that caused an evolutionary response in mutational robustness also caused the evolution of negative, or less positive, epistasis. Taken together, these results confirm that mutational robustness and negative epistasis both evolved in response to selection for recombinational robustness.

The most likely explanation for the evolution of negative epistasis in these simulations is that epistasis evolved as a correlated response to selection for genetic robustness. The direction of epistasis was negatively correlated with mutational robustness among a random sample of viable gene networks (Supplementary Fig. 3). Similar correlations were found in digital organisms and RNA secondary structure[14], supporting the theoretical prediction[14,25] that it is impossible to change genetic robustness and the direction of epistasis independently.

Although we recognize that our model describes a simplified view of transcriptional regulation, it captures an important feature of real genetic regulatory systems: genetic interactions are abundant, causing mutations to have different effects depending on the genetic background in which they arise. We propose that sexual reproduction will favour the evolution of increased genetic robustness and, therefore, negative epistasis in any system with two key properties: numerous genetic interactions and abundant genetic variation—both known requirements for the evolution of genetic robustness[17–19]. Consistent with this proposal, the parameters that determine the number of genetic interactions (connectivity and gene number) and the amount of genetic variation (mutation rate, population size and the strength of stabilizing selection) all influenced whether negative epistasis evolved in our simulations (Fig. 3; see also Supplementary Figs 4–7). In contrast, the network topology, the shape and variance of the mutational distribution, and environmental stochasticity did not qualitatively affect the outcome (Supplementary Figs 1, 5, 8 and 9). Most notably, negative epistasis failed to evolve in networks that were both small and sparsely connected (Supplementary Figs 5 and 6). However, the requirements for the evolution of negative epistasis were not too restrictive. Sexual reproduction produced negative epistasis even in small networks as long as they were sufficiently connected, and in sparsely connected networks as long as they contained a sufficient number of genes (Supplementary Figs 5 and 6). Stabilizing selection acting on the gene expression pattern also prevented the evolution of negative epistasis, but only when it was exceptionally strong (Supplementary Fig. 7). The wealth of genetic interactions in the transcriptional networks of real organisms[26] and the abundance of genetic variation in natural populations[27] suggest that negative epistasis will evolve in many sexually reproducing organisms.

The evolution of negative epistasis in our simulations is remarkable because it suggests that sexual reproduction selects for conditions that favour its own maintenance. Because negative epistasis enhances the ability of natural selection to purge deleterious mutations in sexual populations, our results could explain the maintenance of sexual reproduction in the face of its numerous

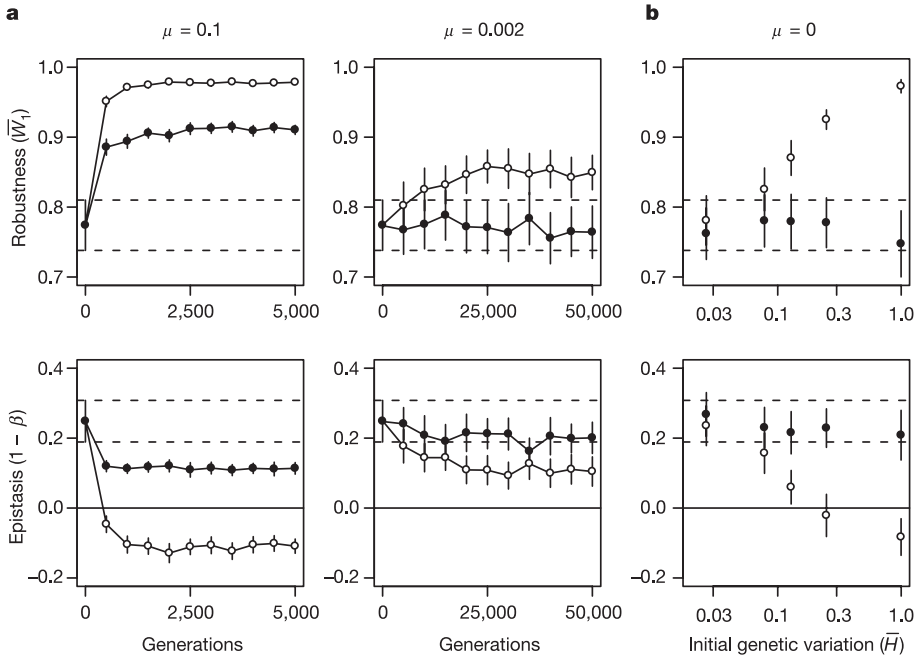


Figure 3 | Sexual reproduction selects for mutational robustness and negative epistasis. a, Selection for the ability to produce any stable gene expression pattern was imposed on 50 replicate populations subjected to high ($\mu = 0.1$) and low ($\mu = 0.002$) mutation rates. Plots show the average evolutionary responses in robustness to mutation and direction of epistasis. **b**, The 50 individuals used to found the homogeneous populations described in **a** were used to found new populations of 500 individuals with 1, 3, 5, 10 or 75 random mutations each. These populations were then allowed to evolve without the occurrence of new mutations ($\mu = 0$) until genetic variation was exhausted. We plot the robustness to mutation and the direction of epistasis at equilibrium (that is, after evolution stops) against the initial genetic variation in each treatment (see Supplementary Methods). Each population in **a** and **b** was evolved under either asexual (filled circles) or sexual (open circles) reproduction. Data are expectations and 95% confidence intervals for the median value among the 500 individuals in each population. Dashed lines indicate the 95% confidence intervals for the mean of the 50 founder networks.

costs[28]. We hypothesize that sexual reproduction enabled evolution of the robustness apparent in the developmental networks of multicellular organisms and that negative epistasis should be associated with robustness in these systems. If these hypotheses are correct, they will help to explain the prevalence of sexual reproduction among living organisms.

METHODS

Network model. Networks are generated by randomly filling the entries of the R matrix (for example, Fig. 2b) with $(1-c)N^2$ zeros and cN^2 standard normal random variates. A corresponding initial gene expression pattern, $\mathbf{S}(0)$, is created for each network by randomly setting each $s_i(0)$ to either -1 or 1. Development begins with the initial gene expression pattern, $\mathbf{S}(0)$, and proceeds through 100 iterations of equation (1). We determined that an equilibrium steady state was achieved when the following criterion was met[2]:

$$\sum_{\theta=t-10}^{t} D[\mathbf{S}(\theta),\overline{\mathbf{S}}(t)] \leq 10^{-3}, \quad \text{where } D[\mathbf{S},\mathbf{S}'] = \frac{1}{4N}\sum_{i=1}^{N}(s_i - s_i')^2$$

is a measure of the difference between the gene expression patterns $\mathbf{S}$ and $\mathbf{S}'$, and $\overline{\mathbf{S}}(t)$ is the average of the gene expression levels over the time interval from $t-10$ to t. The ability of a genotype to reach equilibrium within 100 iterations is termed developmental stability[2].

Evolution. In a typical evolutionary simulation, a single random individual capable of producing a stable gene expression pattern is cloned to generate a population of 500 identical individuals. In an asexually reproducing population, offspring are generated by picking an individual at random from the population and allowing it to produce a clone of itself, such that each non-zero entry in the R interaction matrix mutates (replacement with an independent standard normal random variate) with probability $\mu/(cN^2)$. In our model, mutations should be viewed as acting on the cN^2 *cis*-regulatory elements, not the coding sequences of the N genes themselves; in addition, mutations cannot alter the number of genes, or establish new interactions between genes. Only offspring capable of producing a stable gene expression pattern survive. This process is repeated until 500 developmentally stable individuals are produced, which go on to found the following generation. In a sexual population, offspring are generated by picking two individuals at random from the population, and selecting rows of the R matrices from each parent with equal probability (analogous to free recombination between units formed by each gene and its *cis*-regulatory elements, but with no recombination within regulatory regions), while allowing each non-zero entry to mutate as above. Each selective regime was applied to a fixed panel of 50 replicate populations, each derived from a single independently generated random individual and initial gene expression pattern; simulations were run for as long as was necessary to obtain an equilibrium (that is, no significant change) in the second half of the simulation. In each simulation, all individuals experience the same initial gene expression pattern as the founder individual.

Robustness and epistasis. The mean effects of k mutations on fitness were modelled by the relationship[14,15]: $\log(\overline{W}_k) = -\alpha k^{\beta}$ (Fig. 1, equation (2)). Mutational robustness and directional epistasis were measured by $\overline{W}_1$ and $1-\beta$, respectively. To estimate these parameters for a given genotype, we generated 100 individuals with five successive rounds of random mutations each, and measured the proportion of viable genotypes, $\overline{W}_k$, with $k = 1, 2, \ldots, 5$ mutations. We modelled $\overline{W}_k$ using a generalized linear model with complementary log–log link and a binomial error structure[29]:

$$\log[-\log(\overline{W}_k)] = \log(\alpha) + \beta\log(k).$$

$\overline{W}_1$ was measured directly and β was estimated using maximum likelihood.

Received 1 October; accepted 22 November 2005.

1. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336,** 435–440 (1988).
2. Siegal, M. L. & Bergman, A. Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl Acad. Sci. USA* **99,** 10528–10532 (2002).
3. Wagner, A. Does evolutionary plasticity evolve? *Evolution* **50,** 1008–1023 (1996).
4. Lynch, M. *et al.* Perspective: Spontaneous deleterious mutation. *Evolution* **53,** 645–663 (1999).
5. Whitlock, M. C., Phillips, P. C., Moore, F. B. G. & Tonsor, S. J. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26,** 601–629 (1995).
6. Burch, C. L. & Chao, L. Epistasis and its relationship to canalization in the RNA virus phi 6. *Genetics* **167,** 559–567 (2004).
7. de Visser, J. A. G. M., Hoekstra, R. F. & van den Ende, H. An experimental test for synergistic epistasis and its application in *Chlamydomonas*. *Genetics* **145,** 815–819 (1997).
8. Mukai, T. The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. *Genetics* **61,** 749–761 (1969).
9. Whitlock, M. C. & Bourguet, D. Factors affecting the genetic load in *Drosophila*: synergistic epistasis and correlations among fitness components. *Evolution* **54,** 1654–1660 (2000).
10. Bonhoeffer, S., Chappey, C., Parkin, N. T., Whitcomb, J. M. & Petropoulos, C. J. Evidence for positive epistasis in HIV-1. *Science* **306,** 1547–1550 (2004).
11. de Visser, J. A. G. M., Hoekstra, R. F. & van den Ende, H. Test of interaction between genetic markers that affect fitness in *Aspergillus niger*. *Evolution* **51,** 1499–1505 (1997).
12. Elena, S. F. & Lenski, R. E. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390,** 395–398 (1997).
13. Michalakis, Y. & Roze, D. Evolution. Epistasis in RNA viruses. *Science* **306,** 1492–1493 (2004).
14. Wilke, C. O. & Adami, C. Interaction between directional epistasis and average mutational effects. *Proc. R. Soc. Lond. B* **268,** 1469–1474 (2001).
15. You, L. & Yin, J. Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage T7. *Genetics* **160,** 1273–1281 (2002).
16. Kawecki, T. J. The evolution of genetic canalization under fluctuating selection. *Evolution* **54,** 1–12 (2000).
17. Rice, S. H. The evolution of canalization and the breaking of von Baer's laws: Modeling the evolution of development with epistasis. *Evolution* **52,** 647–656 (1998).
18. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A population genetic theory of canalization. *Evolution* **51,** 329–347 (1997).
19. Nijhout, H. F. The nature of robustness in development. *Bioessays* **24,** 553–563 (2002).
20. Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412,** 331–333 (2001).
21. de Visser, J. A. G. M. *et al.* Perspective: Evolution and detection of genetic robustness. *Evolution* **57,** 1959–1972 (2003).
22. Stearns, S. C. The evolutionary links between fixed and variable traits. *Acta Paleontol. Pol.* **38,** 215–232 (1994).
23. Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J. III & Doyle, J. Robustness of cellular functions. *Cell* **118,** 675–685 (2004).
24. Charlesworth, B. & Barton, N. Recombination load associated with selection for increased recombination. *Genet. Res.* **67,** 27–41 (1996).
25. Wagner, G. P., Laubichler, M. D. & Bagheri-Chaichian, H. Genetic measurement of theory of epistatic effects. *Genetica* **102–103,** 569–580 (1998).
26. Milo, R. *et al.* Network motifs: Simple building blocks of complex networks. *Science* **298,** 824–827 (2002).
27. Barton, N. H. & Keightley, P. D. Understanding quantitative genetic variation. *Nature Rev. Genet.* **3,** 11–21 (2002).
28. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1978).
29. Crawley, M. J. *Statistical Computing* (Wiley, Chichester, 2002).
30. Jaeger, J. *et al.* Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics* **167,** 1721–1737 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Flatt, Y. Fofanov, F. Galis, J. Kingsolver, A. Monteiro, M. Travisano and G. Wagner for discussions. The UH, UNC and NIH (grant to C.L.B.) provided financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B.R.A. (razevedo@uh.edu).

Stuart A. Kauffman

Statement

and

Readings

Complexity in Physics and Life

I will briefly discuss the staggering non-ergodicity of the universe at the level of complex chemistry and above, where the universe has not had time to make all possible molecules of, say, carbon, nitrogen, hydrogen, oxygen, phosophorous, and sulphur, with up to 10,000 atoms per molecule. I will discuss the origin of molecular self reproduction, probably via collectively autocatalytic sets of molecules, the onset of agency in the universe, probably via a union of molecular self reproduction in a confined environment such as a membrane, and thermodynamic work cycles, and the puzzling problem of self propagating organization of process that occurs in life, and perhaps the biotic universe, that vexed Kant over 200 years ago. Here, work is the constrained release of energy into a few degrees of freedom, but it typically takes work to construct those very constraints. Cells propagate the interweaving of work, constraint construction, and constraints as one sense of "information", that propagates until the cell reproduces. We need to lift physics up to biology to understand the living state.

Biol Philos
DOI 10.1007/s10539-007-9066-x

Propagating organization: an enquiry

Stuart Kauffman · Robert K. Logan · Robert Este
Randy Goebel · David Hobill · Ilya Shmulevich

Received: 2 January 2007 / Accepted: 19 February 2007

Abstract Our aim in this article is to attempt to discuss propagating organization of process, a poorly articulated union of matter, energy, work, constraints and that vexed concept, ''information'', which unite in far from equilibrium living physical systems. Our hope is to stimulate discussions by philosophers of biology and biologists to further clarify the concepts we discuss here. We place our discussion in the broad context of a ''general biology'', properties that might well be found in life anywhere in the cosmos, freed from the specific examples of terrestrial life after 3.8 billion years of evolution. By placing the discussion in this wider, if still hypothetical, context, we also try to place in context some of the extant discussion of information as intimately related to DNA, RNA and protein transcription and translation processes. While characteristic of current terrestrial life, there are no compelling grounds to suppose the same mechanisms would be involved in any life form able to evolve by heritable variation and natural selection. In turn, this allows us to discuss at least briefly, the focus of much of the philosophy of biology on population genetics, which, of course, assumes DNA, RNA, proteins, and other features of terrestrial life. Presumably, evolution by natural selection—and perhaps self-organization—could occur on many worlds via different causal mechanisms. Here we seek a non-reductionist explanation for the synthesis, accumulation,

S. Kauffman · R. K. Logan · R. Este
Institute for Biocomplexity and Informatics, University of Calgary, Calgary, AB, Canada

R. K. Logan (✉)
Department of Physics, University of Toronto, Toronto, ON, Canada
e-mail: logan@physics.utoronto.ca

R. Goebel
Department of Computer Science, University of Alberta, Edmonton, AB, Canada

D. Hobill
Department of Physics and Astronomy, University of Calgary, Calgary, ON, Canada

I. Shmulevich
Institute for Systems Biology, Seattle, WA, USA

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | ☐ LE | ☐ TYPESET
MS Code : BIPH55R1 | ☐ CP | ☐ DISK

and propagation of information, work, and constraint, which we hope will provide some insight into both the biotic and abiotic universe, in terms of both molecular self reproduction and the basic work cycle energy, work, work as the constrained release of energy into a few degrees of freedom, yet the typical requirement for work itself to construct those very constraints on the release of energy that then constitute further work. Information creation, we argue, arises in two ways: first information as natural selection assembling the very constraints on the release of energy that then constitutes work and the propagation of organization. Second, information in a more extended sense is ''semiotic'', that is *about* the world or internal state of the organism and requires appropriate response. The idea is to combine ideas from biology, physics, and computer science, to formulate explanatory hypotheses on how information can be captured and rendered in the expected physical manifestation, which can then participate in the propagation of the organization of process in the expected biological work cycles to create the diversity in our observable biosphere.Our conclusions, to date, of this enquiry suggest a foundation which views information as the construction of constraints, which, in their physical manifestation, partially underlie the processes of evolution to dynamically determine the fitness of organisms within the context of a biotic universe.

Keywords ■

An organized being is then not a mere machine, for that has merely moving power, but it possesses in itself formative power of a self-propagating kind which it communicates to its materials though they have it not of themselves; it organizes them, in fact, and this cannot explained by the mere mechanical faculty of motion.
Immanuel Kant – Critique of Judgement

Introduction

Our broad aim is to understand propagating organization as exemplified by the vast organization of the co-evolving biosphere. Our effort is a rather mysterious undertaking, for we entirely lack a theory of organization of process, yet the biosphere, from the inception of life to today manifestly propagates organization of process. Indeed, we believe that the evolving universe as a whole also manifests the propagation of organization. We shall focus most of our efforts on the biotic case, but undertake an initial extension of our analysis to the abiotic case as well.

The role of information in biology, what it ''is,'' how it accumulates, and how it is ''used,'' has been directly addressed by mainstream biologists and philosophers of biology. By and large, the biological concept of information derives from the DNA, RNA, protein processes of ''coding'', transcription, and translation. Yet in the broader sense that we seek to articulate, information in terrestrial life is likely to be one of the key unifying concepts in the emerging field of systems biology. As part of the propagating organization within living cells, the cell operates as an information-processing unit, receiving information from its environment, propagating that information through complex molecular networks, and using the information stored in its DNA and cell-molecular systems to mount the appropriate response. Indeed,

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | □ LE | □ TYPESET
MS Code : BIPH55R1 | □ CP | □ DISK

biology is acquiring many characteristics of an information science (Hood and Galas 2003).

It is sometimes the case that science progresses by finding the concepts and language to "see that which is directly in front of us." Such is the case with the present enquiry. We are persuaded that we are not wholly successful, but hope that we shall have at least started a far broader discussion.

Two predecessors to this article can be found in *Investigations* (Kauffman 2000), and "Emergence, Autonomous Agents, and Organization" (Kauffman and Clayton 2006). At its core, *Investigations* seeks to understand the physical nature of agency itself, and proposes that a molecular autonomous agent, able to act on its own behalf in an environment, is an autocatalytic system carrying out at least one thermodynamic work cycle. Much follows from this tentative definition, which implies that an autonomous agent is an open non-equilibrium chemical system, and finds general biotic importance in the fact that work cycles link spontaneous and non-spontaneous (exergonic and endergonic) processes. This linkage has built up the enormous complexity of the biosphere.

Further analysis reveals this work to be the constrained release of energy into a few degrees of freedom. But if one asks where the constraints themselves come from—as in the example of a cylinder and piston that confine the expansion of the working gas in the head of the cylinder to yield the translational motion of the piston, hence the release of energy into a few degrees of freedom—one finds that it typically takes work to construct the constraints[1]

Thus we arrive at the first surprise—it takes constraints on the release of energy for work to happen, but work for the constraints themselves to come into existence. This circle of work and constraint shall turn out to be part of the theory of propagating organization that we shall discuss.

Most importantly, contemporary cells are both collectively autocatalytic and do work cycles, in part to construct constraints on the release of energy. When released, this energy constitutes further work that drives non-spontaneous processes that builds structures, drives processes, and also builds further constraints on the release of energy, which when released can build still more such constraints. In short, cells carry out propagating work linking spontaneous processes, constraints, work, and non-spontaneous processes, and more broadly as we shall see, the propagating organization of process. In doing so, the cell carries out a set of interlocked tasks that achieve a closure of tasks whereby the cell literally builds a rough copy of itself. We know this, yet we have no clear way to say what we know. This closure of work, constraints, tasks, and information, as we shall see below, is a new state of matter, energy, information, and organization that constitutes the living state.

The new insight that we explore in this article is that the constraints that allow autonomous agents to channel free energy into work are connected to information: in fact, simply put, the constraints *are* the information, are partially causal in the diversity of what occurs in cells, and are part of the organization that is propagated.

In "Emergence, Autonomous Agents, and Organization" (Kauffman and Clayton 2006), the tentative definition of autonomous agent is extended to include construction of boundaries enclosing the agent, discrimination of "yuck" (meaning

[1] Here we use the word "constraint" in a very general sense that includes "global constraints" (e.g. conservation of energy, symmetry conditions etc.,) and "local constraints" or boundary conditions (e.g. initial conditions, reflection or absorption at a spatial location).

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

poison) or ''yum'' (meaning food), and at least one choice of action: flee (or not), approach (or not). Our language is teleological. We believe that autonomous agents constitute the minimal physical system to which teleological language rightly applies.

It is important that our definition of a molecular autonomous agent applies to terrestrial life, but is, in principle far broader. The concepts identify a new class of far from equilibrium chemical thermodynamic systems, and we suspect, could form the basis of life in a variety of molecular instantiations. For example, already Sievers and von Kiedrowski (1994) and Lee et al. (1997a, b), have made collectively self-reproducing DNA and peptide systems. Quite directly, Lee et al. (ibid.) have shown that self-reproduction does not depend upon the double helix structure of DNA and RNA. Thus self-reproduction on a basis other than template replication, transcription and translation has been achieved. Further, work on the origin of life based on self reproducing liposomes (Mavelli and Luisi 1996), the theory of the probable emergence of collectively autocatalytic sets of molecules (Kauffman 1993), and autocatalysis in organic reaction mixtures (Smith and Morowitz 2004), begins to suggest a broad physical basis for life in the cosmos. Molecular autonomous agents have yet to be created, but work cycles and molecular motors are accomplished experimental facts.

If we succeed in creating, or finding life which is radically different from contemporary earth life, the way will open up for a general biology, and a new union of physics, chemistry, biology, and the information sciences. Core to this, we feel, will be an understanding of propagating organization of process.

We comment that it is unlikely that very early life was based on DNA, RNA and proteins via transcription and translation, given the huge complexity of the molecular apparatus to achieve these events, including encoded enzymes that charge transfer RNA with the correct amino acids to achieve translation. We emphasize this, because we wish to place a discussion of information, and its relation to work, constraint, and propagating organization in a wider context that the contemporary debate among philosophers of biology and biologists about the information status of the DNA $\rightarrow$ RNA $\rightarrow$ protein chain.

In turn, a general biology will necessarily confront us with a discussion of evolution by heritable variation and natural selection, perhaps typically without the familiar concepts of DNA, gene frequencies, alterations of gene frequencies as the microevents of microevolution and the diverse philosophic opinions that have ranged over these issues. We will have to explore the general conditions allowing for evolution and the emergence of biospheres.

This article is organized as follows:

In the ''Darwinian adaptations and preadaptations'' section we discuss Darwinian adaptations and preadaptations, argue that the first implies that biology cannot be reduced to physics, while the second, stunningly, implies that the future evolution of the biosphere cannot be finitely prestated. Much follows from these surprising conclusions.

In the ''Shannon information'' section we discuss Shannon information and argue that it does not apply to the evolution of the biosphere. One reason is that due to Darwinian preadaptations, the ensemble of possibilities and their entropy cannot be calculated.

In the ''Schrödinger's aperiodic crystal...'' section we begin with Schrödinger's famous statement that a periodic crystal cannot ''say'' a lot, while an aperiodic

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

crystal can say a lot, and will contain a microcode. We shall argue that the proper and deep understanding of Schrödinger's intuition is that an aperiodic crystal contains a very large number of diverse constraints that are partially causal in guiding the huge diversity of specific events and processes which occur physically in cells. From this we shall arrive at a new formulation: constraints are information and information is constraints. The first part of this twosome, constraints are information is, we believe, secure. The second part, information is constraints, may be more problematic.

In ''The relativity of information'' section we discuss the relativity of the concept of information.

In the ''Semiosis as a special case...'' section we shall place our definition of biotic information in the larger context in which information is ''about'' something, arguing that when an autonomous agent discriminates yuck or yum, the molecular signatures of yuck or yum are about yuck or yum, hence the rudiment of semiotics. We shall locate biotic (but not linguistic) semiosis, as a subcase of information as constraints.

In the ''Heritable variation and natural...'' section we shall stress that constraints as information, and, derivatively, semiotic information, must have causal consequences for the autonomous agent. These consequences increase its fitness such that the information is assembled by natural selection into the ongoing evolution of the biosphere. Without this coupling to fitness, the information and its effects would not come to exist in the universe. Therefore we shall argue that natural selection constitutes the assembly machinery, when coupled with heritable variation, that literally assembles the propagating organization of matter, energy, constraint, work, and information. This constitutes the propagating organization in autonomous agents, whose co-evolution drives the biosphere's progressive exploration of what we call the Adjacent Possible. This discussion is reminiscent of some aspects of Maynard Smith's argument(2000a, b) that selection confers on genes a specific informational character, and Sterelny and Griffiths (1999) broadened concept that selection confers on many features of a cell or organism the features of information.

In ''The evolution of the abiotic...'' section we attempt to extend our analysis to the abiotic universe. We find that our analysis that considers information as constraints is equivalent to the statement that information consists in boundary conditions and in global constraints. But, in classical and quantum physics, boundary conditions—like the cylinder and piston—are only partially causal for what occurs. Physicists often ''put in by hand'' the boundary conditions of a problem, such as the behavior of the cylinder, piston, and working gas system. But in the unfolding of the biosphere or universe since the Big Bang, the very coming into existence of new boundary conditions—information we argue—is itself part of the full dynamics of the total system. We thus assume a context with information understood as boundary conditions on the release of energy that makes diverse processes happen. So we argue that in the proper union of matter, energy and information it is precisely the union of matter, energy, and boundary conditions that, in an expanding and cooling universe, progressively break symmetries, invade the Adjacent Possible, and cause an increasing diversity of events, processes and structures to come into existence. The evolution of the biosphere is but one case of this general process.

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

In the ''Population genetics and evolution...'' section we briefly discuss the general context of successful evolution by heritable variation and natural selection in a general biology. Here the ''neighborhood'' relations between different autonomous agents is an issue. More essentially, without propagating organization of process there would be nothing upon which selection could act. Thus, we suggest, the habit of population genetics of ignoring the root physical basis of life may first of all constrain our understanding of evolution unnecessarily to contemporary earth life, and misses entirely what we shall describe as the evolution of perhaps any biosphere into its ''Adjacent Possible'', a fundamental feature of life that underlies the specifics of evolution by altering gene frequencies.

Darwinian adaptations and preadaptations

Were one to have asked Darwin what the function of the heart is, he would presumably have responded that the function of the heart is to pump blood. But the heart has a wealth of other causal consequences, such as heart sounds. Heart sounds are not the function of the heart. That is, the causal consequence of the heart that matters, the virtue for which it was selected, was the pumping of blood. So the function of a part (or organ) of an organism is typically, if not always, a subset of its causal consequences. This has major implications. Among these, the function of a part (or organ) of an organism cannot be analyzed except in the context of the whole organism in its selective environment. But further, this fact is just one of the reasons that biology cannot be reduced to physics. In Kauffman and Clayton (2006), it is argued that, if we grant the physicist a theory of everything, say string theory to cite one example, and the capacity to deduce upwards to all that occurs in the universe—an impossibility given throws of the quantum dice—the physicist could deduce all the causal features of the heart, *but would have no way to pick out the pumping of blood as the relevant causal property which is the function of the heart* and which is the property that gave rise to the evolutionary emergence of this organ.

To do so, the physicist would have to discuss whole organisms as causal agents in their own right, evolving under natural selection in changing environments. That is, the physicist would have to become a biologist and talk biology talk. Thus, biology cannot be reduced to physics, rather physics has to be lifted up to biology.

A second reason we feel biology is emergent with respect to physics is that Darwin's natural selection is utterly neutral with respect to the physical basis of heritable variation and hence natural selection. Life might be based on DNA, RNA and proteins, or might be based on autocatalytic organic chemical reactions systems, and/or polymer systems that create a bounding membrane. This implies that a physicist armed with a theory of everything might, (actually could not) deduce that a specific molecular autonomous agent would have offspring of differential reproductive success, the physicist cannot deduce Darwin's natural selection itself which transcends any specific realization of it. Indeed, for small changes in the constants of nature, life might still be possible, hence Darwin's ''higher order'' or emergent law cannot even be reduced to the physics of this universe.

In short, for these and other reasons, we wish to join forces with those who argue for a limitation of reductionism, and the reality of emergence with respect to the furniture of the universe (Silberstein 2003).

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | LE | TYPESET
MS Code : BIPH55R1 | CP | DISK

Darwin had many brilliant insights. Among these is what is now called a Darwinian preadaptation. Here the central concept is that a causal property of a part of an organism that is not of selective significance in the normal environment might become useful in a different environment, and hence become subject to selection. It is critical to point out, first that Darwinian preadaptations have occurred repeatedly in evolution, and second, that such an evolutionary step results in the emergence in the biosphere of a novel function. For example, lungs evolved from the swim bladders of certain early fish. The swim bladders, partially filled with water and partially with air, adjusted the height in the water column to establish neutral buoyancy of the fish. But the swim bladder, with air in it, was preadapted for use as a lung, and air breathing was a novel functionality with its own causal consequences that allowed life to conquer land thereby changing the universe.

We now raise a central question discussed in *Investigations*. Is it possible to say ahead of time what all possible Darwinian preadaptations are for human beings, or for the whole biota of the contemporary biosphere for that matter? The answer appears to be ''No.'' We cannot finitely prestate all possible Darwinian preadaptations. Part of the difficulty, or impossibility, in doing so is that we cannot even begin the task of prestating what all possible selective environments will be. That is, there appears to be no finitely stateable procedure which would allow us to enumerate all possible selective environments.

Part of the challenge is that the concept of such environments is systematically vague. It is not even clear how to begin on the project of listing all possible environments for all actual, let alone possible, organisms. While we do not know how to prove our claim, we believe it to be true and shall assume that it is.

We point out that the property or causal consequence which becomes the subject of a Darwinian preadaptation need not be a mutant property. It might be a normal feature of the organism, but normally of no selective significance until the new environment is encountered. Therefore, an attempt to enumerate the possible preadaptations by trying to count the number of mutations possible to a genome is irrelevant. Darwinian preadaptations cannot, in general, be prestated.

Much follows from the claim that we cannot finitely prestate all possible Darwinian preadaptations of all contemporary organisms. First, it means in a radical sense that we cannot predict the future evolution of the biosphere. We literally have no idea of what such preadaptations may be. Second, it means that a frequency interpretation of probability statements does not apply to possible probability statements about the evolution of the biosphere. In the normal frequency interpretation of probability, say that a fair coin will be heads about 5,000 times out of 10,000 coin flips, one can finitely prestate all possible outcomes. This is not possible for the evolving biosphere. Third, and dramatically, the incapacity to say ahead of time what the relevant preadaptations will be means that we cannot write down a stateable set of variables in equations whose dynamics captures the evolution of the biosphere. But all our mathematical techniques in physics begin with a prestatement of the full set of variables and the configuration space of the system. This is true in Newtonian dynamics, statistical mechanics, general relativity and in quantum mechanics if one does not believe in hidden variables. If one believes in hidden variables then because they are hidden they cannot be prestated hence the caveat for quantum mechanics.

But we cannot prestate the configuration space of the biosphere. Now a classical physist might argue that, if we take the solar system, it is just a large classical 6*N*

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

dimensional system where N is the number of particles in the solar system and the current biosphere is, with the rest of the solar system, a point in that vast space. Let us grant the move. Then, we rejoin, the physicist has no way to pick out the collective variables, the lungs and hearts and wings, and features of the environment that are the relevant causal variables for the ongoing evolution of the biosphere. Thus, again we see that we cannot write down causal laws with a prestated set of (collective) variables for the evolution of the biosphere.

We shall not discuss it further here, but the same incapacity to prestate the evolution of the economy and its technology also arises, as does the incapacity to prestate the evolution of human culture. But all this has the deepest implications. Reductionist science is powerful, but is limited. This sets us free in astonishing ways, for organisms live their lives forward, they do not deduce them. We appear to live in a universe in which our reductionistic world view is inadequate: there is the emergence of life, and value as we discuss below. Human language and culture also represent propagating organization (Logan 2006, 2007). Moreover we live in and partially co-create a ceaselessly ''creative'' biosphere, economy, and human culture. This glimmers a new scientific world view, beyond reductionism with broad potential societal ramifications (Kauffman 2006).

Shannon information

Shannon (1948) information theory has been a brilliant mathematical construct. At its core, Shannon envisioned a Source with a set of messages, symbol strings, over which a well defined probability distribution might be attributed. Then he envisioned a (perhaps noisy) channel over which information is transmitted. He then envisioned a receiver and, importantly, a decoder. Shannon's move was to calculate the entropy of the set of messages at the Source. The information that propagated down the channel and was received at the receiver removed uncertainty with respect to the entropy of the Source. This reduction of uncertainty, hence the lowering of the entropy of the Source, constitutes the amount of information transmitted. One interpretation, not given by Shannon himself who abjured to say what information ''is,'' is that information is just the reduction in uncertainty at the receiver. This definition leaves open exactly what the claim might mean. It might be the reduction of uncertainty in a human receiver's mind, for example.

Importantly, and widely recognized, is the fact that Shannon information considers the amount of information, nominally in bits, but is devoid of semantics. There is no sense of what information is ''about'' in Shannon information.

Now we ask whether Shannon information applies to the evolution of the biosphere. We answer that it does not. In particular, Shannon information requires that a prestated probability distribution (frequency interpreted) be well stated concerning the message ensemble, from which its entropy can be computed. But if Darwinian preadaptations cannot be prestated, then the entropy calculation cannot be carried out ahead of time with respect to the distribution of features of organisms in the biosphere. This, we believe, is a sufficient condition to state that Shannon information does not describe the information content in the evolution of the biosphere.

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

There are further difficulties with Shannon information and the evolving biosphere. What might constitute the ''Source''? Start at the origin of life, or the last common ancestor. What is the source of something like ''messages'' that are being transmitted in the process of evolution from that Source? The answer is entirely unclear. Further, what is the transmission channel? Contemporary terrestrial life is based on DNA, RNA, and proteins via the genetic code. It is insufficient to state that the channel is the transmission of DNA from one generation to the next. Instead, one would have to say that the actual ''channel'' involves successive life cycles of whole organisms. For sexual organisms this involves the generation of the zygote, the development of the adult from that zygote, the pairing of that adult with a mate, and a further life cycle. Hence, part of one answer to what the ''channel'' might be is that the fertilized egg is a channel with the Shannon information to yield the subsequent adult. But it has turned out that even if all orientations of all molecules in the zygote were utilized, there is not enough information capacity to store the information to yield the adult. This move was countered by noting that, if anything, development is rather more like an algorithm than an information channel (Apter and Wolpert 1965). In short, a channel to transmit Shannon information along life cycles does not exist, so again, Shannon information does not seem to apply to the biosphere.

It seems central to point out that the evolution of the biosphere is not the transmission of information down some channel from some source, but rather the persistent, ongoing, co-construction, via propagating organization, heritable variation, and natural selection, of the collective biosphere. Propagating organization requires work. It is important to note that Shannon ignored the work requirements to transmit ''abstract'' information, although it might be argued that the concept of constraints is implicit in the restrictions on the messages at the Source. While we mention this, we have no clear understanding physically of what such constraints are.

One might be tempted to argue that a Shannon-like information theory could be applied to the vast set of selective events that have led to the specific DNA sequences that are in contemporary organisms. But does this move work? Can we specify a finite ensemble of possible DNA sequences out of which the present DNA sequences have been derived? If we consider all DNA sequences longer than, say 1,000 nucleotides, it would take vastly large repetitions of the history of the universe for the universe to construct one copy of each possibility. This cannot physically constitute the ensemble. Is the ensemble the set of DNA sequences that have been explored in the actual evolution of the biosphere, some accepted, most rejected? This approach initially seems promising, but has the obvious difficulty that we cannot specify the ensemble explored in 3.8 billion years, hence do not and cannot know the Shannon information content of the biosphere. A further difficulty with this approach is that it measures the information content of the biosphere as a function of the number of DNA sequences ''tried'' in evolution. But very different numbers of attempted mutations might have led to the same biosphere, hence quantitating the information of the biosphere by the number of attempted DNA mutations is not in direct correspondence to any specific biosphere.

We conclude that a Shannon Information content analysis of the information content of the evolving biosphere is not legitimate.

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

Schrödinger's aperiodic crystal: "instructional" information as constraint or boundary condition

In *What is Life*, Schrödinger (1992) is concerned with the order in organisms and hence the physical basis of the gene. He argues, based on X-ray mutation induction frequency, that the gene must have a few hundred to a few thousand atoms, and points out that statistical mechanical equilibria cannot account for the stability of the organism over generations. He then posits that quantum mechanics in the form of chemical bonds is the answer. Then he brilliantly points out that the order of life cannot be based on a periodic crystal, for such a crystal cannot say a lot, or carry much information. He places his bet on aperiodic crystals which can, in strong contrast, say a lot, or carry much information, even a microcode which will somehow specify the adult.

He was brilliantly right, and presaged DNA and the genetic code. Now we come to the critical issue. In just what sense can an aperiodic crystal "say a lot?" Schrödinger does not himself say more than suggesting that the aperiodic crystal can contain a microcode.

We believe Schrödinger was deeply correct, and that the proper and deep understanding of his intuition is precisely that an aperiodic solid crystal can contain a wide variety of microconstraints, or micro boundary conditions, that help cause a wide variety of different specific events to happen in the cell or organism. Therefore we starkly identify information, which we here call "instructional information" or "biotic information," not with Shannon, but with constraints or boundary conditions. The amount of information will be related to the diversity of constraints and the diversity of processes that they can partially cause to occur. By taking this step, we embed the concept of information in the ongoing processes of the biosphere, for they are causally relevant to that which happens in the unfolding of the biosphere.

We therefore conclude that constraints are information and, as we argue below, information is constraints which we term as instructional or biotic information to distinguish it from Shannon information. We use the term "instructional information" because of the instructional function this information performs and we sometimes call it "biotic information" because this is the domain it acts in, as opposed to human telecommunication or computer information systems where Shannon information operates. This step, identifying information as constraint or boundary condition, is perhaps the central step in our analysis. We believe it applies in the unfolding biosphere and the evolving universe, expanding and cooling and breaking symmetries, that we will discuss below.

Is this interpretation right? It certainly seems right. Precisely what the DNA molecule, an aperiodic solid, does, is to "specify" via the heterogeneity of its structural constraints on the behavior of RNA polymerase, the transcription of DNA into messenger RNA. Importantly, this constitutes the copying or propagating of information. Also, importantly, typically, the information contained in aperiodic solids requires complex solids, i.e., molecules, whose construction requires the linking of spontaneous and non-spontaneous, exergonic and endergonic, processes. These linkages are part of the work cycles that cells carry out as they propagate organization.

It is essential to note that the set of constraints in a contemporary cell is not merely the DNA and RNA, but lies also in the specific stereochemistry of a vast

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | LE | TYPESET
MS Code : BIPH55R1 | CP | DISK

horde of specific molecular species. So, when an enzyme binds two substrates and holds them in proximity, lowering the potential energy barrier to their joining, the enzyme is acting as a constraint on the motion of the two substrates, hence as a catalyst. The working of a cell is, in part, a complex web of constraints, or boundary conditions, which partially direct or cause the events which happen. Importantly, the propagating organization in the cell is the structural union of constraints as instructional information, the constrained release of energy as work, the use of work in the construction of copies of information, the use of work in the construction of other structures, and the construction of further constraints as instructional information. This instructional information further constrains the further release of energy *in diverse specific ways*, all of which propagates organization of process that completes a closure of tasks whereby the cell reproduces.

Our discussion here has some of the flavor of Sterelny and Griffiths (1999) in their discussion of an extended concept of information beyond DNA, RNA and protein sequences. On the other hand, none of those who have written on the concept of information in biology have taken up the struggle to relate it to constraints, work, and propagating organization of process such as that in reproducing cells.

The relativity of information

In the ''Shannon information'' section we have argued that the Shannon conception of information are not directly suited to describe the information of autonomous agents that propagate their organization. In the ''Schrödinger's aperiodic crystal...'' section we have defined a new form of information, instructional or biotic information as the constraints that direct the flow of free energy to do work.

The reader may legitimately ask the question ''isn't information just information?'', i.e., an invariant like the speed of light. Our response to this question is *no*, and to then clarify what seems arbitrary about the definition of information. Instructional or biotic information is a useful definition for biotic systems just as Shannon information was useful for telecommunication channel engineering, and Kolmogorov (Shiryayev 1993) information was useful for the study of information compression with respect to Turing machines.

The definition of information is relative and depends on the context in which it is to be considered. There appears to be no such thing as absolute information that is an invariant that applies to all circumstances. Just as Shannon defined information in such a way as to understand the engineering of telecommunication channels, our definition of instructional or biotic information best describes the interaction and evolution of biological systems and the propagation of organization. Information is a tool and as such it comes in different forms. We therefore would like to suggest that information is not an invariant but rather a quantity that is relative to the environment in which it operates. It is also the case that the information in a system or structure is not an intrinsic property of that system or structure; rather it is sensitive to history and environment. To drive home this point we will now examine the historic context in which Shannon (1948) information emerged.

Before delving into the origin of Shannon information we will first examine the relationship of information and materiality. Information is about material things and furthermore is instantiated in material things but is not material itself. Information is

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

an abstraction we use to describe the behavior of material things and often is thought as something that controls, in the cybernetic sense, material things. So what do we mean when we say the constraints are information and information is constraints as we did in the ''Schrödinger's aperiodic crystal...'' section.

''The constraints are information'' is a way to describe the limits on the behavior of an autonomous agent who acts on its own behalf but is nevertheless constrained by the internal logic that allows it to propagate its organization. This is consistent with Hayle's (1999, p. 72) description of the way information is regarded by information science: ''It constructs information as the site of mastery and control over the material world.'' She claims, and we concur, that information science treats information as separate from the material base in which it is instantiated. This suggests that there is nothing intrinsic about information but rather it is merely a description of or a metaphor for the complex patterns of behavior of material things. In fact, the key is to what degree information is a completely vivid description of the objects in question.

This understanding of the nature of information arises from Shannon's (1948) original formulation of information, dating back to his original paper:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one that will actually be chosen since this is unknown at the time of design. If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely.

A number of problems for biology emerge from this view of information. The first is that the number of possible messages is not finite because we are not able to prestate all possible preadaptations from which a particular message can be selected and therefore the Shannon measure breaks down. Another problem is that for Shannon the semantics or meaning of the message does not matter, whereas in biology the opposite is true. Biotic agents have purpose and hence meaning.

The third problem is that Shannon information is defined independent of the medium of its instantiation. This independence of the medium is at the heart of a strong AI approach in which it is claimed that human intelligence does not require a wet computer, the brain, to operate but can be instantiated onto a silicon-based computer. In the biosphere, however, one cannot separate the information from the material in which it is instantiated. The DNA is not a sign for something else it is the actual thing in itself, which regulates other genes, generates messenger RNA, which in turn control the production of proteins. Information on a computer or a telecommunication device can slide from one computer or device to another and then via a printer to paper and not really change, McLuhan's ''the medium is the message'' aside. This is not true of living things. The same genotype does not always produce the same phenotype.

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

According to the Shannon definition of information, a structured set of numbers like the set of even numbers has less information than a set of random numbers because one can predict the sequence of even numbers. By this argument, a random soup of organic chemicals has more information that a structured biotic agent. The biotic agent has more meaning than the soup, however. The living organism with more structure and more organization has less Shannon information. This is counterintuitive to a biologist's understanding of a living organism. We therefore conclude that the use of Shannon information to describe a biotic system would not be valid. Shannon information for a biotic system is simply a category error.

A living organism has meaning because it is an autonomous agent acting on its own behalf. A random soup of organic chemicals has no meaning and no organization. We may therefore conclude that a central feature of life is organization—organization that propagates.

Semiosis as a special case of constraint as information

We wish next to consider the minimal physical conditions for semiosis. We shall not concern ourselves with fully human linguistic symbols, but with the semiosis of our minimal molecular autonomous agent. Consider an agent that is confronted by molecules in its environment, which constitute ''yuck'' or ''yum.'' To respond to these environmental features, the agent, assumed to be bounded (Kauffman and Clayton 2006), must also have yuck and yum receptors, capable in the simplest case of ''recognizing'' molecules of yuck or yum, and responding appropriately by avoiding yuck and eating yum. Assume such molecular machinery exists in the agent. They of course exist in prokaryotic and eukaryotic cells. We wish to say that the agent confronting yuck or yum receives information ''about'' yuck or yum. This appears to constitute the minimal physical system to which semiotic information might apply. And it is worth noting that the ''meaning,'' or semiotic content of the yuck and yum molecules is built into the propagating organization of the cell. The cell, we want to say, has embodied knowledge and know-how with respect to the proper responses to yuck and yum, which was assembled for the agent and its descendants by heritable variation and natural selection.

The existence of yuck and yum as semiotic signs is a subcase of constraint as information. How does the agent detect yuck? A concrete case would be that a yuck molecule binds a yuck receptor, constraining the receptor's motions, which in turn acts as a constraint in unleashing a cell signaling cascade leading to motion away from yuck. Further, if yuck is present below a detection threshold, it will not be detected by the agent. Hence that threshold, and the receptor itself, act as a constraints partially determining the behavior of the agent in fleeing or not fleeing.

One can construct an underlying set theoretical interpretation for yuck and yum semantics in two equivalent ways: The first posits a set of instances, and a set of properties to which each instance is assigned. The second posits a set of instances and detectors, or classifying operators, that classify ''properties'' of instances. Note that in the second case, those properties need not themselves be discussed because the detectors do the job. If the second stance is taken, then detectors, ''yuck'' and ''not yuck,'' suffice and no extension beyond instructional information is required. If the second stance suffices, we want to say not only that constraints are information but also that information is constraints. We recognize that this second step is arguable and do not analyze this issue further here.

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

Semiotic information can not itself embody ''agentness,'' for it has no agency; but identified agents can be observed to respect the semiotic interpretation like yuck and yum. This inspectable behavior provides the opportunity to attribute constraint-directed behavior to the agent organism.

Another important point in this attempt to understand propagating organization is that the semiotic behavior can identify a source of free energy, yum in this case, from which work can be extracted and propagate in the cell. This behavior is part of a theory that unifies matter, energy, information and propagating organization.

We end this section with the description of a final interesting feature of the yum receptor. A wide variety of molecules might bind to the yum receptor with modest affinity, hence mimic true yum molecules. So the yum receptor can be ''fooled.'' This might allow another agent to emit a poison that mimics the yum molecule, fools the receptor, and leads to the death of the agent. So evolves the biosphere. Now ask, can a Shannon channel be ''fooled?'' Clearly noise can be present in the channel. Due to noise a 1 value can replace a 0 value in the constrained sense of 1 and 0 as subsets of the physical carriers of 1 and 0. But the Shannon channel cannot be fooled: ''fooling'' is a semantic property of detectors, hence not present in a Shannon channel. Therefore, while one might be tempted to measure the amount of semiotic information using a Shannon-like approach, the fact that semiosis in an organism can be fooled suggests that a symbol based Shannon move is inappropriate.

We conclude that semiotic information in molecular agents such as organisms is a special case of information as constraint. For semiotic information to be ''about'' something, and to be extracted, it appears that a constraint must be present in one or more variables that are themselves causally derived from that which the information is about.

Like the threshold level of yum needed for detection, to use the information, the extracted semoitic information must do work on some system. That work might copy the information, for example into a record, or might construct constraints on the release of energy which is further work. Here, semiotic information becomes part of propagating organization.

We comment that in standard semiotic analyses with human agents and language, there are three elements to semiotic information, namely,

1. The subject of the information or the agent being informed;
2. The object of the information or what the information is about; and
3. The possibly arbitrary, sign or symbol referring to the object.
4. With Monod (1971) in Chance and Necessity we add that allosteric chemistry allows arbitrary molecules to cause events. If we wish to call such molecules ''symbols'' that ''refer to'' ''yum,'' the standard semiotic analysis just noted applies to molecular autonomous agents. Note that Monod's example is broader than DNA, RNA and proteins. It is the general arbitrariness of allosteric chemistry that allows arbitrary molecules to cause events. Information is thus broader than coding.

Heritable variation and natural selection as assembly processes

We have now grounded biotic information as ''instructional information'' or constraint, or boundary condition, that partially causes subsequent events in the

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

unfolding of the biosphere. In this view information is not an abstraction, but is causally efficacious in the biosphere and we argue below in the unfolding of the abiotic universe. And we have grounded semiotic information as information detected about external (to the agent) features of the environment about which it learns. These semiotic cases are also cases of constraints, or boundary conditions, detecting and categorizing inputs and partially causing subsequent events. We note again that we remain neutral for the moment about whether information needs to be extended beyond instructional information for a set theory analysis of the categorization of objects.

At the level of complex molecules, as noted above, the universe has not had time to create all possible versions. For example, the universe has not had time to create all proteins length 200, by about 10 to the 67th power repetitions of the history of the universe.

Consider a simple set of organic molecules and all the reactions they can collectively undergo. Call the initial set of molecules the Actual. Now among the reactions that might happen, some may lead to molecular species that are not present in the initial Actual. Call these new molecular species the Adjacent Possible. They are the molecular species that are reachable in a single reaction step from the current actual. It is of fundamental importance that the biosphere has been evolving into the Adjacent Possible for 3.8 billion years, from an initial diversity of perhaps 1,000 organic molecules to trillions. The biotic world advances into the adjacent possible in terms of molecules, morphologies, species, behaviors, and technologically from pressure flaked stones; it lurks in everything from the global economy to the computer, and the millions of products in the current global economy.

Once at a level of complexity sufficiently above the atom, the universe, the biosphere, and the technosphere can never exhaust the diversity of things and events that can happen. The evolving universe and biosphere advance persistently into the adjacent possible. This means that what comes to exist at these levels of complexity is typically unique in the universe.

Now consider a heritable variation which gives rise to a new constraint, physical biotic information, that helps cause a sequence of events in a molecular agent. If that heritable variation is to the selective benefit of the agent, the new constraint, the new biotic information, will be grafted into the organism, its progeny, and the ongoing evolution of the biosphere.

It is essential to note that in the absence of heritable variation, an increase in fitness, and natural selection, this new functionality would not come to exist in the universe: but lungs and flight have come to exist. The mechanisms of heritable variation and natural selection comprise an assembly process by which propagating organization is modified in normal Darwinian adaptations and preadaptations where new functionalities arise, and these modifications are built into the ongoing evolution of the biosphere.

It is clear then, that heritable variation and natural selection are sufficient mechanisms in the biosphere to build an expanding mesh of functionalities as the biosphere invades the adjacent possible. We will ask next whether similar processes can happen in the abiotic universe.

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

The evolution of the abiotic expanding universe: propagating organization diversifying sources of constraint, free energy, and coupling of spontaneous and non-spontaneous processes

We here ask whether we can find generalizations of the above analysis of information, matter, energy, constraint, work, in the biosphere, in the abiotic expanding universe.

For some time, scholars have struggled to find the union of matter, energy, and information. Cases such as Maxwell's demon, the Bekenstein bound on the entropy of a black hole, and the holographic principle, all seem to be places in physics where matter, energy, and information come together. These cases merit attention, but we leave them unanalyzed, except for this comment.

For information to be united with matter and energy, information must be part of the physical unfolding of the universe. Thus, consider Maxwell's demon. It has been shown that the demon cannot "win" with respect to the Second Law of Thermodynamics for a closed equilibrium system (Kauffman 2000). However, in a non-equilibrium setting, the demon can win by making measurements that reduce the entropy of the measured system, with respect to the demon, *faster* than the most compressed record of the measured system grows, on average, in length. Now physicists usually end their argument with a claim rather like, "Then, in principle, work could be extracted." Such a statement is inadequate for a theory that unites matter, energy, and information. What is required is that, in the non-equilibrium setting, a displacement from equilibrium that is a source of free energy must be detected by at least one measurement; a physical system able to couple to that source of free energy must have come to exist and must actually extract free energy, and must release that energy in a constrained way to carry out actual work. Thereafter, this work may propagate.

If we conceive of an abiotic physical system able to carry out these processes of measurement and work extraction in the abiotic universe, it will have to be an abiotically derived system able to perform such measurements, recording the results, and employ the record of the measurements to extract actual work. Such a system will be a case of propagating organization with boundary conditions as constraints, including measurements in the record as constraints on the behavior of the system conditional on the recorded measurements, and the constrained release of energy in work. Whether the coming into existence in the universe of such a system is plausible abiotically is certainly open to question but may be worthy of consideration. Biotically, of course, such systems abound: sources of free energy from sunlight to prey are detected and coupled to work extraction. Records of sources of free energy in the form of food are seen in ant phermone trails. The measurement of a source of free energy and extracting that free energy typically involves thresholds and other constraints or boundary conditions. For example, ants will not follow a phermone trail if it is below a detection threshold, and the boundaries of the trail are boundary conditions on the ants' motions.

These considerations suggest that we take information to be constraint or its physical equivalent, boundary conditions that partially cause events, where the coming into existence of the constraint is itself part of propagating organization. If we do so, the issue starts to clarify in a simple way. It is fully familiar in physics that one must specify the laws, particles, the initial and boundary conditions, then

Journal : 10539 Dispatch : 26-2-2007 Pages : 19
Article No. : 9066 LE TYPESET
MS Code : BIPH55R1 CP DISK

calculate the behavior of the system in a defined state space. Now it is common, as noted, in physics, to ''put in by hand'' the boundary conditions, as in the cylinder and piston case. But in the evolving biosphere, itself part of the evolving universe, and in the evolving universe as a whole, new boundary conditions come into existence and partially determine the future unfolding of the biosphere or the universe. These evolving boundary conditions and constraints are part of the propagating organization of the universe.

We consider a single, but complex case in cosmic evolution. It is well known that molecular grains are found in interstellar space. These grains aggregate up to the scale of planetessimals. Now it is also well known that the grains have surfaces with complex molecular features on which complex chemistry appears to be occurring. The grains themselves act as constraints, or boundary conditions, that confine reacting substrates, hence may catalyze reactions, some of which may be endergonic, requiring, for example, photons. In some cases, the product molecules presumably are bound to the growing grain, thereby modifying the boundary conditions afforded by the grain, which in turn modifies the chemical reactions that can occur. Furthermore, the product molecules can be novel substrates—hence novel sources of free energy—which again allow novel chemical reactions to occur. In short, the grains appear to behave as constraints that can partially guide spontaneous or non-spontaneous processes, can, in addition, link spontaneous and non-spontaneous processes, can create new constraints enabling such processes and linked processes, and can create novel sources of free energy in the form of novel substrates able to enter into new chemical reactions.

Assume the above account is roughly correct. Then the growing grains appear to be cases in which matter, energy, and continuously evolving boundary conditions and novel sources of free energy *emerge*, and condition the future evolution of the grains. The grains are at levels of complexity sufficiently above atoms so that what occurs is typically unique in the universe. It seems virtually sure that no two modest size grains are molecularly identical. Here we confront a union of matter, energy, and evolving and diversifying boundary conditions linking, for example, spontaneous and non-spontaneous processes, and providing diversifying sources of free energy, which alter the ever diversifying structures that come to exist in the evolving expanding universe.

If this approach has merit, it appears to afford a direct union of matter, energy, and information as constraint or boundary condition.

Population genetics and evolution in any biosphere

Philosophy of biology has largely grown up in the constrained environment of current terrestrial life. Its analysis of heritable variation and selection has largely ignored the physical basis of the propagating organization and closure of tasks that achieve the living state and underlie heritable variation. Moreover, we have discussed above the fact that at levels of complexity above atoms, the universe is on a unique trajectory into the Adjacent Possible. These physical facts are utterly requisite to descent with heritable variation and natural selection. But these aspects are simply assumed, without deeper analysis, as available to evolution. Life would have a hard time evolving at the level of complexity of quarks, gluons, and atoms.

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | LE | TYPESET
MS Code : BIPH55R1 | CP | DISK

The diversity is insufficient at least. While we do not now know the implications of the broadened view of a general biology and the evolution of biospheres in a general biology, we suspect that these issues are worth careful consideration. We will make or find life anew in the next century almost certainly. Adaptation, preadaptation, the relation between the specific physical basis of each form of life and the capacity for heritable variations will become the subjects of intense study. And meanwhile, the possibility of general laws remains open to investigation. For example, it has long been hypothesized that cells are dynamically critical, poised between order and chaos. Recent evidence begins to support this possibility (Shmulevich and Kauffman 2004; Ramo et al. 2006; Serra et al. 2004; Aldana et al. 2005; M. Nykter et al. in preparation). Since critical networks are rare in the space of dynamical systems, if cells are critical it is precisely a marriage of self organization affording such critical behavior, and the selective usefulness of criticality, that would account for the putative results noted above. Perhaps molecular autonomous agents in any biosphere are dynamically critical. Perhaps the hinted fourth law of thermodynamics discussed in Investigations is true of all biospheres. We simply do not know. But that does not imply that we should not search for such laws—laws that are emergent with respect to physics and part of the emergent, endlessly ''creative'' universe in which we appear to live.

We would end by inviting philosophers of biology, physics, and others, to help think through the potential implications of a new scientific world view that goes beyond the reductionism of the past three and a half centuries to emergence and a creative evolution in biology and the human economic and cultural realms that cannot be prestated. We believe that such a change in scientific worldview, if merited, will bring with it large societal changes.

Summary

We have traveled a new path in which we have discussed Darwinian adaptations and the non-reducibility of biology to physics, the mysterious Darwinian preadaptations which seem to preclude finite prestatement and lead to evolution where the state space cannot be prestated. This brings us to serious doubts about whether Shannon information directly apply to the evolution of the biosphere, and lead to Schrödinger's aperiodic crystal and the hypothesis that information is constraints and boundary conditions, to semiotic information and records, and to the realization that, in the biosphere, it is heritable variation and natural selection that build the intricate web of propagating organization. This provides the basis for considering a new union of matter, energy, information-constraint, and work in cells. This leads to questions about the abiotic universe, where information as boundary conditions affords a simple means to unite matter energy and information.

We have been led to doubt that Shannon information is physically instantiated, whereas the evolving universe and biosphere are.

We seek a new theory of propagating organization, the unfolding of Kant's statement at the outset of this article. We further seek a theory of the diversifying sources of free energy and constraints that are used to couple spontaneous and non-spontaneous processes into an ever expanding diversity of processes in the biosphere and universe. We do not believe our analysis is fully adequate, but believe it is a start.

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | LE | TYPESET
MS Code : BIPH55R1 | CP | DISK

References

Aldana M, Shmulevich I, Kauffman SA (2005) Eukaryotic cells are dynamically ordered or critical but not chaotic. PNAS, September 2005

Apter MJ, Wolpert L (1965) Cybernetics and development. I. Information theory. J Theor Biol 8:244–257

Hayle K (1999) The condition of virtuality. In: Peter L (eds) The digital dialectic. MIT Press, Cambridge, MA

Hood L, Galas D (2003) The digital code of DNA. Nature 421:444–448

Kauffman S (1993) The origins of order: self-organization and selection in evolution. Oxford University Press, New York

Kauffman S (2000) Investigations. Oxford University Press, Oxford

Kauffman S (2006) The third culture beyond reductionism: reinventing the sacred. Edge.org, November 20, 2006

Kauffman S, Clayton P (2006) On emergence, agency, and organization. Biol Philos 21:501–521

Lee DH, Severin K, Reza Ghadiri M (1997a) Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. Curr Opin Chem Biol 1:491–496

Lee DH, Severin K, Yokobayashi Y, Reza Ghadiri M (1997b) Emergence of symbiosis in peptide self-replication through a hypercyclic network. Nature 390:591–594

Logan RK (2006) The extended mind model of the origin of language and culture. In Nathalie G, Jean P, Van B, Diederik A (eds) Evolutionary epistemology, language and culture. Springer, Dordrecht

Logan RK (2007) The extended mind: the origin of language and culture. University of Toronto Press, Toronto

Mavelli F, Luisi PL (1996) Autopoietic self-reproducing vesicles: a simplified kinetic model. J Phys Chem 100:16600–16607

Maynard SJ (2000a) The concept of information in biology. Philos Sci 67:177–194

Maynard SJ (2000b) Reply to commentaries. Philos Sci 67:214–218

Monod J (1971) Chance and necessity. Knopf, New York

Ramo P, Kessel J, Yli O (2006) Perturbation avalanches and criticality in gene regulatory networks. J Theor Biol 242(1):164–170

Schrödinger E (1992) What is life? Cambridge University Press, Cambridge

Serra R, Kessell J, Semeria A (2004) Genetic network models and statistical properties of gene expression data in knock-out experiments. J Theor Biol 227(1):149–157

Shannon CE (1948) A mathematical theory of communication. Bell Syst Technical J 27:379–423, 623–656

Shiryayev AN (ed) (1993) Selected works of A.N. Kolmogorov, vol III, Information theory and the theory of algorithms (mathematics and its applications). Kluwer Academic Publishing, New York

Shmulevich I, Kauffman SA (2004) Activities and sensitivities in Boolean network models. Phys Rev Lett 93(4):048701(1–4)

Sievers D, von Kiedrowski G (1994) Self-replication of complementary nucleotide-based oligomers. Nature 369:221–224

Silberstein M (2003) Reduction, emergence and explanation. In: Peter M, Michael S (eds) The Blackwell guide of the philosophy of science. Blackwell Publishing, Williston, VT

Smith E, Harold M (2004) Universality in intermediary metabolism. Santa Fe Institute Working Paper. http://www.santafe.edu/research/publications/workingpapers/04-07-024.pdf

Sterelny K, Griffiths PE (1999) Sex and death. University of Chicago, Chicago

Journal : 10539 | Dispatch : 26-2-2007 | Pages : 19
Article No. : 9066 | LE | TYPESET
MS Code : BIPH55R1 | CP | DISK

Chapter 1 Jan 3, 07

Beyond Reductionism: Reinventing the Sacred

Batter my heart, three-person'd God; for you
As yet but knock, breathe, shine, and seek to mend;
That I may rise, and stand, o'erthrow me and bend
Your force, to break, blow, burn and make me new.
I, like an usurpt town, to another due,
Labour to admit you, but Oh, to no end,
Reason your viceroy in me, me should defend,
But is captiv'd, and proves weak or untrue.
Yet dearly I love you, and would be loved fain,
But am betroth'd unto your enemy:
Divorce me, untie, or break that knot again,
Take me to you, imprison me, for I
Except you enthrall me, never shall be free,
Nor ever chaste, except you ravish me.

John Donne's exquisite Sacred Sonnet, written about 1590AD when he was a high Anglican Churchman, speaks to one of the most poignant schisms in Western society, and more broadly in the world, that between faith and reason. Today this schism finds enormous voice in the vehement disagreements between the religious fundamentalists in the United States, or Islamic Fundamentalists who believe in a transcendent Creator God, and "secular humanists" who do not believe in a transcendent God. These diverse beliefs are profoundly held. Our senses of the sacred have been with us for thousands of years, at least from the presumptive female earth goddess of Europe ten thousand years ago, to Egyptian, Greek, Abrahamic, Aztec, Mayan, Incan and Hindu gods, Buddhism, Taoism, and more. I recently learned of an aboriginal tribe unwilling to allow its DNA to be sampled as part of a world wide study on the origins and evolution of humanity for fear that its view of its own sacred origins would be challenged. Ways of life in the world hang in the balance. This book, Beyond Reductionism: Reinventing the Sacred, hopes to address this schism in a fruitful way. Part of the project of this book is to discuss newly discovered limitations to the reigning scientific world view, reductionism, that has dominated Western science at least since Newton, but that leaves us in a meanningless world of facts devoid of values. In its place I will discuss a newly glimmered scientific world view, beyond reductionism, in which we are members of a vastly creative universe in which life, agency, meaning, value, consciousness and the full richness of human action and creativity have emerged. But even beyond this emergence, we will find grounds to radically alter our understanding of what science itself appears able to tell us: I hope to show that science cannot foretell the evolution of the biosphere, foretell the evolution of human technologies, let alone human culture or history. A central implication of this new world view is that we are partial co-creators of a ceaselessly novel, creative universe, biosphere, and culture. That this appears to be true

is, simply stated, awesome. I shall want to say: is it more amazing to think that an Abrahamic transcendent God created all around us, all that we participate in, in six days, or that all arose with no transcendent Creator God, all on its wondrous own. I believe the latter is so stunning, so overwhelming, so worthy of awe, gratitude and respect, that it is God enough for me, and I hope many of us. I shall want to say that God is the very creativity of the universe.

God is the most powerful symbol we have created. Do we use the "God" word or not? I believe we should. The word "God" carries with it thousands of years of the most profound respect, awe and wonder. But that wonder is due a creative universe viewed beyond reductionism. "You see", we can say, "God is the name we use to mean the creativity of the universe we share and help create - the universe, the emergence of life and agency and consciousness, the evolving biosphere and human culture, are worthy of awe and respect. What more could one want of a God?"

This God is not one to which we can pray, nor does this God invite faith in a heaven and hell. Yet a view in which life is meaning-laden, and creative, may afford us a different view of ourselves and our place in the universe. If we can find awe, reverence, respect and a global responsibility to all of life and the globe on which we live in this possible new world view, it may form the basis for a transnational mythic structure to sustain the global civilization that is emerging.

Billions of people across our world have faith in a God. Billions of us, including myself, do not. What is before us therefore, is the next step in this global civilization that is emerging. The work of working out our views is a long term task. In recent response to the Christian fundamentalist right in the United States, a response has been heard from Richard Dawkins in "The God Delusion", Daniel Dennett in "Breaking the Spell", and Sam Harris in "The End of Faith" and "Letter to a Christian Nation". Where freedom of expression is to be valued, these as well as the opinions of those of faith, are to be heard. I hope in this book to be exploring what may be a third path, afforded by the new implications of a new scientific world view and its spiritual and cultural implications.

Unraveling this new scientific world view and its implications for our unity with nature, with all of life, is task enough. But the project before us appears to be even larger. T.S. Elliot once wrote that with Donne and the other "metaphysical poets" of the Elizabethean Age, for the first time in the Western mind, a split arose between reason and our other human sensibilities. The anguish between faith and reason in Donne's Sacred Sonnet, is but one of the schisms that was emerging if Elliot is correct. The Western mind that now dominates much of secular society, placed its faith in reason, and decried the rest of our humanity, Elliot's "other sensibilities", the fullness of human life. Poetry, once thought of as a path to truth, became instead merely a path to appreciation and surprise. Donne wrote roughly in the time of Copernicus and Kepler. Within a hundred years Newton had given us his three laws of motion and universal gravitation, uniting rest and motion, earth and the heavens, and the full foundations of modern science. With Newton, as I shall discuss, reductionism began its profound reign. In the ensuing centuries, science, which I love as a practicing scientist, and the Elightenment, have given birth to secular society. More, science, in particular, redutionistic physics, has emerged as the gold standard for learning about the world. Indeed, science has almost become a kind of secular religion.

Almost un-sensed, our science based secular modern society suffers at least four injuries which split our humanity down the center. I wish to try to address them in this book, for they are part of very much larger cultural issues even than the secular versus religious split in modern society. What the metaphyisical poets began to split asunder, reason and the remaining human sensibilities, we must now attempt to reintegrate to re-understand our full humanity. This understanding is also part of reinventing the sacred. This task is too large for me to encompass. I can speak with some authority about science. I am not an expert in these broader domains. The Quakers have a beautiful view. It is that the universe is so vast and complex that no one can understand it all. Thus humility is necessary. I approach these broader topics with the deepest humility. Perhaps the task is too large for any single author. Thus I ask of you, my reader, that you understand the limitations of this one author, but that you think with me, and beyond me as well. If we begin to discuss these issues, we may find our way to a new view of our selves and our lives. It is hard to imagine how much may be at stake in this discussion. It may prove part of a global ethic, and part of how we evolve a global civilization in the coming centuries.

The first injury is between science and the humanities. C.P. Snow wrote a famous essay in 1959, "The Two Cultures", decrying the fact that science and the humanities are split apart. Einstein or Shakespeare, we seem to be told, but not both in a single framework. This split is itself a further expression of the split between reason and sensibilities from the metaphysical poets of which T.S. Elliot wrote. It is, in fact, precisely a fracture down the middle of our integrated humanity. For example, humanities and the arts are told they are engaged in the soft sciences and often feel themselves to be second class intellectual citizens. I believe it is important that this view is deeply wrong. Science itself is more limited than we have realized and in any case is not the only path to knowledge and more broadly understanding. The humanities and the arts are among such pathways. Most of us live lives in vastly diverse ways that are non-scholarly. Part of what I shall discuss is that science cannot explain the intricate context depended specificities of much of human action and invention. Beyond the reign of science is the older reign of practical action and with it, practical reason. And beyond reason itself is the rest of our humanity and how we live our lives. This matters because, as yet, we have no unified view of our own humanity. Nor do we appear to have a unified discussion or view of our entire humanity, how we live our lives and therefore the global civilization that it may be wise to evolve toward.

A second injury derives from our current reigning scientific world view. Reductionism has taught us that, at its base the real world we live in a world of fact, without values. I shall describe this in detail below. Thus, our reigning scientific world view, reductionism, further cuts us off from our own humanity, where we live lives full of value and meaning, yet have no secure place for these facets of our humanity along with science in a single framework.

A third injury is that we secular humanists have been quietly taught that spirituality is at least questionable, if not foolish. Some of us secular humanists are spiritual, but most of us are not. We are, therefore, unknowingly cut off from a deep aspect of ourselves. Humans have, in one form or another, been spiritual for thousands of years and we secular humanists are bereft of it.

And the fourth injury is that we secular humanists lack a global ethic. We believe in love of family, friend, fairness, place our faith in democracy. But we are largely reduced to consumers. It is telling that Nobel Laureate economist, Kenneth Arrow, when asked to help evaluate the U.S. National Parks, was stymied because he could not compute the utility of these Parks for U.S. consumers. Even in our lives in nature we are reduced to consumers. But the value of U.S. Parks and beyond is life itself and our participation in it. It is this very materialism that so profoundly dismays many thoughtful believers in both the Islamic world and the religious in the Western world.

Part of reinventing the sacred may be to heal these injuries, injuries that we secular humanists hardly know we suffer.

Part of the healing of the injuries I have noted, may derive, unexpectedly, from the apparent limitations of science itself as I have already suggested. What if what I shall write is, as I deeply believe, correct: that we cannot foretell the future evolution of the biosphere, technology, culture, human action or history? We are unable to predict, or can do so at most in very limited ways. When these thoughts first occurred to me, I felt apprehensive. "How can I know what to do?" I thought. Then I felt, hopefully, life on Earth has been solving this problem for 3.8 billion years or so. After all, we do not deduce our lives, we live them forward in time, a point the philosopher Kierkegaard made as well. Or, as Niels Bohr, one of the founders of Quantum Mechanics put it so wonderfully, "Prediction is so difficulty, particularly about the future."

In living our lives forward in time, we bring to bear all the tools that we have accumulated in the 3.8 billion year history since the presumed advent of life on Earth, certainly in the history of animal life, vertebrate life, mammalian life, and the five million years of hominid evolution. Reason, celebrated since Socrates as of precedent value, is but one of the means by which we make our way. In what sense of reason, to be wondered about, does the CEO make his or her decision to act or not?

In this regard it is deeply interesting that in the Medieval period, legal reasoning was held to be the height of human rationality. Legal reasoning deals with the situated specific actions of human agents. It is with Newton's triumph that scientific reasoning came to reign supreme. And in its reductionistic version, that same scientific reasoning denies the very existence of the human agency that legal reasoning struggles to adjudicate. Here again is the split down the heart of humanity.

In raising the issues described above, I shall begin by basing my discussion on science itself. Much of the science that I shall discuss in this book, where I do have some authority, is new, and part of what the new emerging scientific world view appears to be. Part of what I shall discuss is the emergence of life itself, and of agency in very simple systems, that is the amazing fact that organisms can act on their own behalf, and can alter the universe in doing so. With agency, value, meaning, action, and doing enter the universe for the first time. These are genuine parts of our real world. And I shall be at pains to elaborate a possible theory of consciousness and situated human conscious action. Life and consciousness are pre-eminately topics upon which the religious feel the profound need for a Creator God. Thus this is first a book of science, often new science, sometimes, as with consciousness, problematic science. For example, I shall discuss the fact that the universe is vastly non-repeating at levels of complexity from molecules to operas. The universe will never have time to create all complex molecules, let alone operas. This leads into a persistent advance into what I shall call the Adjacent Possible,

certainly in the chemical evolution of this planet, the evolution of the biosphere, and the evolution of the economy and human history. It is also a book of philosophy, for I wish to show why we must go beyond reductionism. And I wish to discuss the inventive creativity of full human action and whether it is, as I think, not "algorithmic", a term I will explain later, and not predictable in its situated details. But above all it is a book aimed at finding the start of a single view of our entire humanity in a universe made sacred by its awesome creativity. And made sacred to us by our membership in it.

Thus, this book has a broad task: To place humanity as part of a universe, biosphere, and evolving culture of persistent creativity. To attempt to frame our whole humanity, science, practical human action, the humanities, law, as a to-be integrated vision of ourselves as we co-create our human and partially our biological and physical worlds. These discussions will, I hope, be part of reinvineting the sacred, finding a global ethic, and beginning to envision the global civilization that we are creating.

Michael Travisano

Statement

and

Readings

Most biological systems appear complex, which has been both a motivation to study for some and cause for appeal to divinity for others. A major goal of biological research has been to reduce the apparent complexity into a clearly understandable set of principles. The goal is given some parameter values, involving variables such as those associated with environmental conditions, body size and perhaps life-history, the complexity of an organism could be predicted with some degree of confidence. By some measures, this research program has been fantastically successful. Understanding the organization of genes into operons in Eubacteria, environmentally induced patterns of gene expression, and predator-prey dynamics are all examples of complex systems for which reasonably good predictions can sometimes be made without precise information on the particulars of the system.

Natural Selection is a major hurdle for universal predictions on the specifics of complexity. Since the outcome of evolution and adaptation are contingent on the prior state of the system, predictions on the specifics of evolutionary outcomes will necessarily be difficult without substantial information on the initial conditions. Moreover, the greater the biological complexity of a system, the greater the potential for initial conditions to affect evolutionary outcomes. Thus, generality in understanding complex systems in Biology is unlikely to be achieved simply by studying the parts of biological systems in isolation.

Reading:

Direct demonstration of an adaptive constraint.
S. P. Miller, M. Lunzer, and A. M. Dean (2006)
Science **314**, 458-461

The effects of evolution are local: Evidence from experimental evolution in *Drosophila*
M. R. Rose, H. B. Passananti, A. K. Chippindale, J. P. Phelan, M. Matos, H. Teotónio and L. D. Mueller. Integrative & Comparative Biology **45,** 486-491.

Life's Complexity Pyramid
Z. N. Oltvai and A.-L. Barabasi
Science 298, 763-764.

the CCR5 agonist RANTES induced interleukin-6 (IL-6) release from wild-type but not CCR5Δ32 iDCs (Fig. 4C). Furthermore, pretreatment of iDCs with pertussis toxin or TAK-779 inhibited the ability of peptide-loaded myHsp70 to stimulate DCs to generate influenza peptide–specific CTLs (Fig. 4D). To examine the effect of CCR5-mediated signaling in mycobacterial infection, we used the model pathogen *M. bovis* BCG-lux (*12*). As reported for murine DCs (*13*), human DCs from immune people were unable to kill internalized mycobacteria, even in the presence of autologous T cells. TAK-779 inhibition of CCR5 led to a dose-dependent enhancement of intracellular mycobacterial replication (Fig. 4E and fig. S6A) at all concentrations of mycobacteria tested (fig. S6B), suggesting an important role for this receptor in controlling mycobacterial infection.

The identification of CCR5 as the critical receptor for myHsp70-mediated DC stimulation has implications for both mycobacterial infection and the therapeutic use of myHsp70. CCR5 is important in immune cell cross talk. Interaction with its naturally occurring ligand MIP-1β promotes the recruitment of cells to sites of inflammation (*14*), facilitates immune synapse formation (*15*), orchestrates T cell interactions within lymph nodes (*16*), and controls the activation and differentiation of T cells (*17*). Our finding that a mycobacterial lysate, as well as purified myHsp70, stimulated a CCR5-dependent calcium response indicates a further connection between the innate and adaptive immune responses during mycobacterial infection. The cellular aggregation induced by myHsp70 signaling through CCR5 may play an important role in the formation of granulomas, the hallmark of mycobacterial infection.

Microbial-induced DC responses need to be highly regulated, reflecting a balance between a rapid and appropriate response to invading microbes and the inducement of immunopathology (*2*)—a particular problem in mycobacterial infection. An increasing number of human pathogens, including HIV (*18*), toxoplasma (*19*), and *M. tuberculosis* (as described here), target the CCR5 receptor. This role of CCR5 as a pattern-recognition receptor for myHsp70 may, at least in part, be responsible for the maintenance of the high CCR5Δ32 allele frequency (10 to 15%) in Northern European populations (*20*) and may alter the pattern of disease seen in people with the CCR5Δ32 allele.

References and Notes

1. J. L. Flynn, J. Chan, *Trends Microbiol.* **13**, 98 (2005).
2. J. L. Flynn, J. Chan, *Annu. Rev. Immunol.* **19**, 93 (2001).
3. Y. Wang *et al.*, *J. Immunol.* **169**, 2422 (2002).
4. M. J. Gething, J. Sambrook, *Nature* **355**, 33 (1992).
5. P. Srivastava, *Annu. Rev. Immunol.* **20**, 395 (2002).
6. P. A. MacAry *et al.*, *Immunity* **20**, 95 (2004).
7. M. Baba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5698 (1999).
8. M. Benkirane, D. Y. Jin, R. F. Chun, R. A. Koup, K. T. Jeang, *J. Biol. Chem.* **272**, 30603 (1997).
9. J. E. Hildreth, F. M. Gotch, P. D. Hildreth, A. J. McMichael, *Eur. J. Immunol.* **13**, 202 (1983).
10. P. Revy, M. Sospedra, B. Barbour, A. Trautmann, *Nat. Immunol.* **2**, 925 (2001).
11. M. L. Dustin, T. G. Bivona, M. R. Philips, *Nat. Immunol.* **5**, 363 (2004).
12. B. Kampmann *et al.*, *J. Infect. Dis.* **182**, 895 (2000).
13. K. A. Bodnar, N. V. Serbina, J. L. Flynn, *Infect. Immun.* **69**, 800 (2001).
14. J. G. Cyster, *Annu. Rev. Immunol.* **23**, 127 (2005).
15. M. Nieto *et al.*, *J. Exp. Med.* **186**, 153 (1997).
16. F. Castellino *et al.*, *Nature* **440**, 890 (2006).
17. S. A. Luther, J. G. Cyster, *Nat. Immunol.* **2**, 102 (2001).
18. T. Dragic *et al.*, *Nature* **381**, 667 (1996).
19. J. Aliberti *et al.*, *Nat. Immunol.* **4**, 485 (2003).
20. E. de Silva, M. P. Stumpf, *FEMS Microbiol. Lett.* **241**, 1 (2004).

21. We are grateful to M. Mahaut-Smith for the use of the Cairn Spectrophotometer and A. Betz for CCR5$^{-/-}$ mice. Funded by the Wellcome Trust (P.J.L.), the Medical Research Council (R.A.F.), the Swiss National Science Foundation (J.P.), and FEBS (E.H.). P.J.L. holds a Lister Institute Research Prize.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5798/454/DC1
Materials and Methods
Figs. S1 to S7
Table S1
Videos S1 to S4
References

26 May 2006; accepted 8 September 2006
10.1126/science.1133515

Direct Demonstration of an Adaptive Constraint

Stephen P. Miller,[1] Mark Lunzer,[1] Antony M. Dean[1,2]*

The role of constraint in adaptive evolution is an open question. Directed evolution of an engineered β-isopropylmalate dehydrogenase (IMDH), with coenzyme specificity switched from nicotinamide adenine dinucleotide (NAD) to nicotinamide adenine dinucleotide phosphate (NADP), always produces mutants with lower affinities for NADP. This result is the correlated response to selection for relief from inhibition by NADPH (the reduced form of NADP) expected of an adaptive landscape subject to three enzymatic constraints: an upper limit to the rate of maximum turnover (k_{cat}), a correlation in NADP and NADPH affinities, and a trade-off between NAD and NADP usage. Two additional constraints, high intracellular NADPH abundance and the cost of compensatory protein synthesis, have ensured the conserved use of NAD by IMDH throughout evolution. Our results show that selective mechanisms and evolutionary constraints are to be understood in terms of underlying adaptive landscapes.

The old notion of natural selection as an omnipotent force in biological evolution has given way to one where adaptive processes are constrained by physical, chemical, and biological exigencies (*1*–*4*). Whether constraint and/or stabilizing selection explain phenotypic stasis, in the fossil record and in phylogenies, remains an open question (*5*). Direct experimental tests of constraint are scarce (*6*–*9*). Even tight correlations among traits, at once suggestive of constraint, can be broken by artificial selection to produce new phenotypic combinations (*8*, *9*). Despite all circumstantial evidence, results from direct experimental tests imply that selection is largely unconstrained.

The direct experimental test for constraint is conceptually simple. A phenotype is subjected to selection (natural or artificial) in an attempt to break the postulated constraint (*6*–*9*). A response to selection indicates a lack of constraint. No response to selection indicates the presence of a constraint. However, the cause of a constraint is rarely specified because the etiologies of most phenotypes are not well understood, their relationships to fitness are usually opaque, and a lack of response to selection may reflect nothing more than a lack of heritable variation (*4*, *7*). If the cause of a constraint is to be elucidated, it must be for a simple phenotype whose relationship to fitness is understood.

Coenzyme use by β-isopropylmalate dehydrogenase (IMDH) is a simple phenotype whose etiology and relationship to fitness are understood (*10*, *11*). IMDHs catalyze the oxidative decarboxylation of β-isopropylmalate to α-ketoisocaproate during the biosynthesis of leucine, an essential amino acid. All IMDHs use nicotinamide adenine dinucleotide (NAD) as a coenzyme (cosubstrate). This invariance of function among IMDHs hints at the presence of ancient constraints, even though some related isocitrate dehydrogenases (IDHs) use NADP instead (*12*, *13*).

Structural comparisons with related NADP-using IDHs identify amino acids controlling coenzyme use (*14*–*16*) (Fig. 1A). Introducing five replacements (Asp^{236} → Arg, Asp^{289} → Lys, Ile^{290} → Tyr, Ala^{296} → Val, and Gly^{337} → Tyr) into the coenzyme-binding pocket of *Escherichia coli leuB*–encoded IMDH by site-directed mutagenesis causes a complete reversal in specificity (*10*, *11*): NAD performance (k_{cat}^{NAD}/K_m^{NAD}, where K_m is the Michaelis constant) is reduced by a factor of 340, from 68×10^3 M^{-1} s^{-1} to 0.2×10^3 M^{-1} s^{-1}, whereas NADP performance ($k_{cat}^{NADP}/K_m^{NADP}$) is increased by a factor of 70, from 0.49×10^3 M^{-1} s^{-1} to 34×10^3 M^{-1} s^{-1}. The engineered LeuB[RKYVYR] mutant (the final R represents Arg^{341}, already present in wild-type *E. coli* IMDH) is as active and as specific toward NADP as the wild-type enzyme is

[1]BioTechnology Institute, [2]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN 55108, USA.

*To whom correspondence should be addressed. E-mail: deanx024@umn.edu

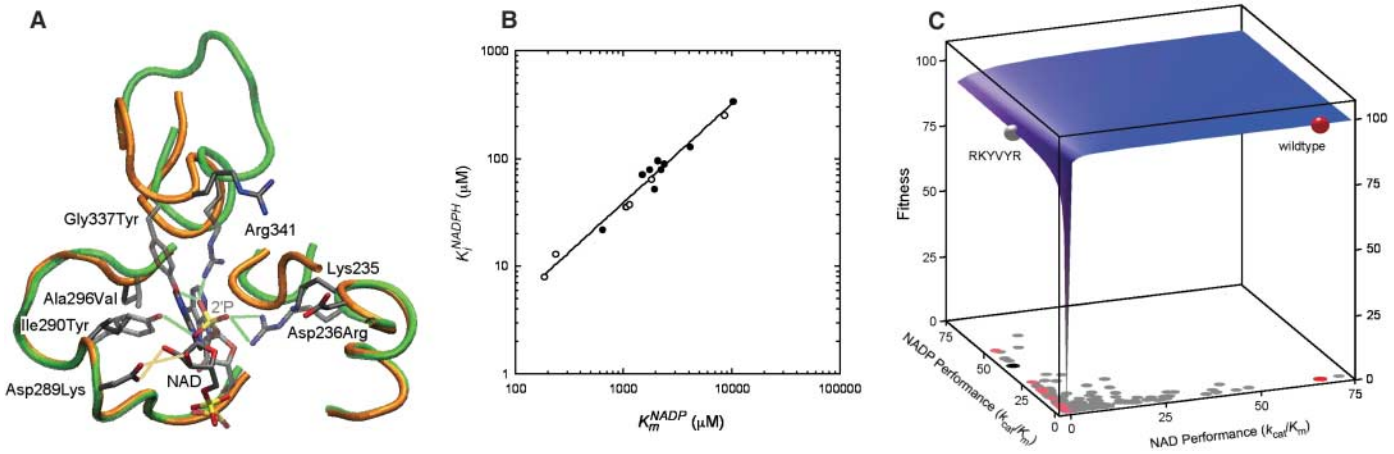


Fig. 1. The molecular anatomy of an adaptive constraint. (**A**) Structural alignment of the coenzyme binding pockets of *E. coli* IMDH (*14*) (brown main chain) with bound NAD modeled from *Thermus thermophilus* IMDH (*15*) and *E. coli* IDH (*16*) (green main chain) with bound NADP. Only key residues are shown (gray, carbon; red, oxygen; blue, nitrogen; yellow, phosphorus) with labels designating the amino acid and site number in IMDH followed by the amino acid in IDH. Coenzyme use is determined by H bonds to NAD (brown lines) and to the 2′-phosphate (2′P) of NADP (green lines). (**B**) Mutations in the coenzyme binding pocket that stabilize NADP also stabilize NADPH. The ensuing correlation in affinities for NADP (K_m^{NADP}) and NADPH (K_i^{NADPH}) seen with engineered mutants (circles) (*11*) is retained in the screened mutants (dots). (**C**) The adaptive landscape for coenzyme use by IMDH (*11*) showing the phenotype-fitness map (blue surface) determined in chemostat competition and described by equation S5 (*11*, *18*) and the predicted distribution of performances for 512 mutants in the coenzyme binding pocket (gray dots), with wild type (red dot), LeuB[RKYVYR] (black dot), and single amino acid replacements in LeuB[RKYVYR] coenzyme binding pocket (pink dots). The fitness of LeuB[RKYVYR] is hypothesized to be lower than in the wild type because of strong inhibition by abundant intracellular NADPH (*11*).

toward NAD. Evidently, protein architecture has not constrained IMDH to use NAD since the last common ancestor.

Despite similar in vitro performances, the NADP-specific LeuB[RKYVYR] mutant is less fit than the NAD-specific wild type (*11*). IMDHs, wild-type and mutant alike, display a factor of 30 higher affinity for the reduced form of NADP (NADPH) (coproduct) than for NADP (Fig. 1B). We suggested the LeuB[RKYVYR] mutant is subject to intense inhibition by intracellular NADPH, which is far more abundant in vivo than is NADP (*11*, *17*). The inhibition slows leucine biosynthesis, reduces growth rate, and lowers Darwinian fitness (Fig. 1C). The wild type retains high fitness because its affinity for NADPH is low, whereas inhibition by NADH, which is far less abundant than NAD in vivo, is ineffective. Perhaps as a consequence of differences in Michaelis complex structure (*18*), IDH is not subject to such intense NADPH inhibition and hence could evolve NADP use.

We hypothesize that IMDH is constrained to use NAD because the strong inhibition associated with NADP use reduces fitness. However, identifying the mechanism of selection (NADPH inhibition) is not synonymous with identifying the causes of constraint. Mutations that increase k_{cat}^{NADP} (maximum rate of NADP turnover), that break the correlation in NADP and NADPH affinities (Fig. 1B), or that eliminate the trade-off in coenzyme performances (Fig. 1C) could each benefit the LeuB[RKYVYR] mutant without compromising its performance with NADP (*18*). We therefore hypothesize that IMDH is constrained to use NAD by three causes: an upper limit to k_{cat}^{NADP}, an unbreakable correlation in the affinities of NADP and NADPH, and an inescapable trade-off in coenzyme performance.

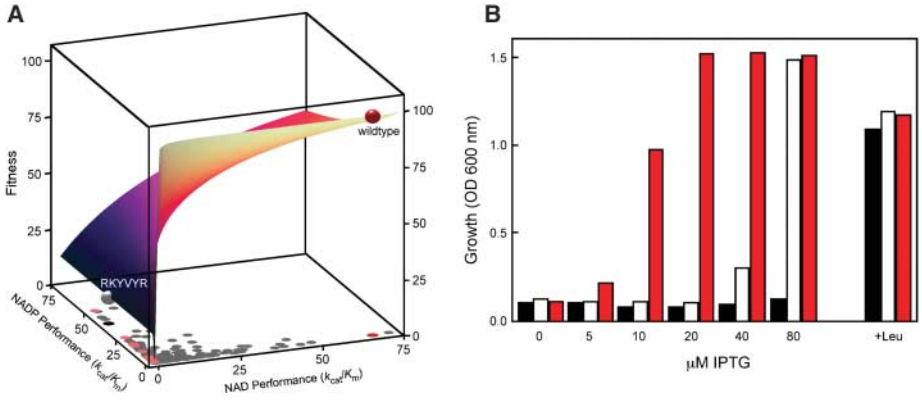


Fig. 2. The phenotypic basis of the genetic screen. Lowered expression in the presence of excess glucose brings IMDH to saturation with isopropylmalate, increasing coenzyme affinities. (**A**) Lowering IMDH expression in the chemostat-derived adaptive landscape [lower concentration of *E* in equation S5 (*11*, *18*)] is predicted to reduce the fitness of LeuB[RKYVYR] (white sphere) far more than that of the wild type (red sphere). (**B**) IPTG-controlled expression of IMDHs ligated downstream of the T7 promoter in pETcoco in strain RFS[DE3] (*leuA*$^+$*B*^{am}*C*$^+$, with T7 RNA polymerase expressed from a chromosomal *lacUV5* promoter) confirms that lower expression affects growth in minimal glucose medium of the LeuB[RKYVYR] (white) more than in the wild type (red). Plasmids lacking LeuB (black) are incapable of growth except in the presence of leucine.

We used directed evolution (targeted random mutagenesis and selective screening) (*19*–*21*) to test whether NADP-specific IMDHs with higher fitness could be isolated. Random substitutions were introduced into *leuB[RKYVYR]* by means of error-prone polymerase chain reaction (*18*). Mutated alleles were ligated downstream of the T7 promoter in pETcoco (a stable single-copy vector) and transformed into strain RFS[DE3] (*leuA*$^+$*B*^{am}*C*$^+$, with T7 RNA polymerase expressed from a chromosomal *lacUV5* promoter). Sequencing unscreened plasmids revealed that, on average, each 1110–base pair *leuB[RKYVYR]* received two nucleotide substitutions. From the pattern of base substitutions in these mutants and assuming Poisson statistics, we estimate that only 10.5 (0.4%) of the 2431 possible amino acid replacements were missing in our mutant library (*18*).

Our experimental design used decreased IMDH expression to provide a simple selective screen for mutations in *leuB[RKYVYR]* that increase growth and/or fitness. As predicted from the phenotype-fitness map (Fig. 2A), selection against *leuB[RKYVYR]* intensified as IMDH expression was lowered in the presence of excess glucose (Fig. 2B). Using 40 μM isopropyl-β-D-thiogalactopyranoside (IPTG) to induce a low level of IMDH expression, we found that cells harboring *leuB[WT]* formed large colonies at 24 hours, whereas cells harboring *leuB[RKYVYR]*

barely formed pinprick colonies at 48 hours. We chose colony formation at 48 hours as the least stringent criterion compatible with reliably identifying beneficial mutations in *leuB[RKYVYR]*. Longer periods of growth (or higher concentrations of IPTG) allowed unmutated *leuB[RKYVYR]* to form colonies, whereas shorter periods (or lower concentrations of IPTG) produced no colonies.

Of the 100,000 mutated plasmids screened, 134 (representing 107 distinct isolates) formed colonies within 48 hours (table S1). Each had either a substitution in the 5′ leader sequence upstream of *leuB[RKYVYR]*, an amino acid replacement in the coenzyme binding pocket, or both. Upstream substitutions occurred at three nucleotide positions: –3, –9, and –14 relative to the *leuB* AUG start codon (fig. S1). Ten amino acid replacements were found at three codons in the coenzyme binding pocket.

At first glance, the positive response to selection might suggest that coenzyme use by IMDH is unconstrained. Upstream substitutions in the Shine-Dalgarno sequence (positions –9 and –14), as well as a new AUG start codon (position –3) that replaces the less efficient GUG start codon, presumably derive their benefits through increases in expression because their kinetics are unchanged. Unexpectedly, however, beneficial amino acid replacements in the coenzyme binding pocket eliminate H bonds to the 2′-phosphate of NADP to cause striking reductions in NADP performance (Table 1). Although some mutants have improved NAD performance, others remain unchanged and several show reduced NAD performance. Isolated amino acid replacements outside the coenzyme binding pocket, which might have been expected to increase k_{cat}^{NADP} or to break the correlation in affinities between NADP and NADPH (K_m^{NADP} and K_i^{NADPH}), have no detectable functional effects (table S2, those associated with beneficial 5′-leader mutations in Table 1). No doubt they, along with 126 silent substitutions (table S1), hitchhiked through the genetic screen with the beneficial mutations.

Table 1. Kinetic effects of amino acid replacements in LeuB[RKYVYR] isolated at 48 hours growth. Standard errors are <13% of estimates. Single-letter abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Enzyme	NADP			NAD			Preference (A/B)	K_i^{NADPH} (μM)
	K_m (μM)	k_{cat} (s^{-1})	Performance (A) k_{cat}/K_m (mM^{-1} s^{-1})	K_m (μM)	k_{cat} (s^{-1})	Performance (B) k_{cat}/K_m (mM^{-1} s^{-1})		
Site								
DDIAGR (wild type)	8400	4.10	0.4900	101	6.90	68.3200	0.007	254.30
RKYVYR	183	6.20	33.8800	4108	0.83	0.2000	169.400	9.90
Beneficial replacements								
K235N	3919	5.95	1.5200	6356	0.50	0.0800	19.000	129.00
K235N,–3(M),A60V	1744	5.06	2.9000	5395	0.44	0.0800	35.700	56.00
K235R	554	6.39	11.5000	3292	0.36	0.1100	106.100	26.00
K235R,E63V	658	6.75	10.3000	3748	0.46	0.1200	84.100	28.00
K289E*	3449	2.71	0.7900	4897	1.98	0.4000	2.000	133.50
K289E*,D317E	3554	3.62	1.0200	5580	2.09	0.3700	2.800	112.00
K289E*,D87Y,S182F	2551	2.54	1.0000	5750	3.65	0.6300	1.600	105.20
K289E*,F102L	3891	3.66	0.9400	4823	1.10	0.2300	4.100	131.90
K289M	2216	4.13	1.8600	4034	0.48	0.1200	15.500	78.60
K289M,F170L	1945	2.50	1.2900	2947	0.48	0.1600	8.100	87.20
K289N*	1141	6.32	5.5400	4997	1.14	0.2300	24.300	44.00
K289N*,R187H	1146	5.83	5.0900	3069	0.79	0.2600	19.600	44.70
K289N*,H367Q	1248	5.27	4.2200	4555	1.05	0.2300	18.300	57.80
K289T	1596	6.18	3.8700	4056	0.78	0.1900	20.400	71.60
K289T,Q157H	1696	5.66	3.3400	5038	0.93	0.1800	18.600	72.60
Y290C	988	5.21	5.2700	3792	1.02	0.2700	19.600	49.00
Y290C,R152C	910	3.15	3.4600	3844	0.90	0.2300	14.800	43.00
Y290D	10213	3.13	0.3100	9040	0.82	0.0900	3.400	344.00
Y290F*	1658	5.59	3.3700	3110	0.65	0.2100	16.000	59.70
Y290F,P97S,D314E	1097	4.64	4.2300	4158	0.71	0.1700	24.800	57.40
Y290N,N52T	2381	3.44	1.4400	10660	0.26	0.0200	59.500	90.00
Replacements with beneficial 5′-leader mutations								
RKYVYR	183	6.20	33.8800	4108	0.83	0.2000	169.400	9.90
–3(M)†,E66D	208	2.97	14.2800	5002	0.30	0.0600	238.000	11.90
–3(M)†,E66D,E331D	267	7.99	29.9000	3403	0.34	0.1000	296.900	14.00
E82D	221	5.44	24.6200	4735	0.35	0.0700	351.700	14.10
K100R,A229T,L248M	137	5.58	40.3700	4079	0.45	0.1100	370.300	11.50
G156R,K289R,Y311H	292	5.40	18.4900	4464	0.54	0.1200	144.800	17.40
E173D	219	6.39	29.1800	5304	0.89	0.1700	171.600	6.60
Y337N,G131	171	5.71	33.3900	3367	0.24	0.0700	468.500	8.00
Y337H,P85S,E181K	318	5.21	16.3800	2409	0.22	0.0900	183.600	12.80

*Switch from NADP to NAD requires two base substitutions in one codon. This is a possible transitional amino acid produced by a single base substitution (*11*). †The –3(M) designates a G to A substitution at base –3 that creates a new start codon.

Our results suggest that increases in expression are beneficial, whereas increases in NADP performance are not. This seeming paradox is resolved if there are no mutations capable of breaking the upper limit to k_{cat}^{NADP}, the correlation in affinities for NADP and NADPH, or the trade-off in coenzyme performance. With these constraints, and with reduced expression in the genetic screen, the phenotype-fitness map near LeuB[RKYVYR] remains flat with respect to increases in NADP performance (Fig. 2A). By contrast, severe losses of NADP performance are predicted to be beneficial as correlated reductions in the affinities for NADPH free up IMDH for use with abundant NAD. As predicted, all beneficial LeuB[RKYVYR] mutants have reduced affinities for both NADP and NADPH (Table 1). Unaffected by constraints, increases in expression are unconditionally beneficial (*18*). These results support the hypothesis that NADP-specific IMDHs function poorly in vivo because of strong inhibition by abundant NADPH. That reductions in NADP performance and increases in expression are both beneficial are the predicted consequences of a phenotype-fitness map constrained by an upper limit to k_{cat}^{NADP}, a correlation in affinities for NADP and NADPH, and a trade-off in coenzyme performance.

Breaking any one constraint would allow NADP-specific IMDHs to evolve. Yet no mutant increases k_{cat}^{NADP}, no mutant uncouples the affinities for NADP and NADPH, and no mutant breaks the trade-off in coenzyme performance. These conclusions are not the result of a selective screen that is too stringent. Of 35 colonies, representing 26 distinct mutants, that appeared after the 48-hour limit, 17 had no nucleotide substitutions in the 5′ leader sequence or amino acid replacements in LeuB[RKYVYR] (table S3). Amino acid replacements in the other mutants neither improved enzyme performance nor decreased NADPH inhibition (table S4). That no additional beneficial mutations were recovered with relaxed criteria demonstrates that the selective screen was not overly stringent.

Nor are the results a consequence of inadequate sampling of protein sequence space. Natural adaptive evolution fixes advantageous mutations sequentially (*22*–*24*). Indeed, experimental evolution demonstrates that advantageous double mutants in the evolved β-galactosidase of *E. coli* are not evolutionarily accessible and perforce must be accumulated as sequential advantageous mutations (*25*). Hence, screening

all single substitutions is a sufficient sampling of protein sequence space for robust evolutionary conclusions. We estimate that only 0.04 advantageous amino acid replacements are missing from the mutant *leuB[RKYVYR]* library (*18*). We conclude that mutations capable of breaking the limit, the correlation, or the trade-off are unlikely to ever be fixed in populations because they are exceedingly rare (they may not exist), because they are minimally advantageous, or both.

The two remaining ways to evolve an NADP-specific IMDH are to reduce intracellular NADPH pool and, as our results show, to increase expression. Reducing intracellular NADPH relieves the inhibition but, as experiments deleting sources of NADPH show (*13*), the disruption to the rest of metabolism costs far more than the benefit to be gained. The phenotype-fitness map (Fig. 1C) imposes a law of diminishing returns such that LeuB[RKYVYR] must be expressed above wild-type levels by a factor of 100 to overcome the inhibition by NADPH (*18*). Diverting resources away from other metabolic needs toward compensatory protein synthesis would impose a protein burden (*26*–*29*) sufficient to prevent the evolution of NADP-specific IMDHs.

The production of unnatural phenotypes, by artificial selection or molecular engineering, is not sufficient to conclude that evolutionary constraints are absent entirely. Rather, potential constraints underlying a conserved phenotype can be identified from the relationships among genotype, phenotype, and fitness that define an adaptive landscape. Experimental evolution can then be used to test their existence. Using this approach, we have shown how certain structure-function relationships in IMDH have constrained its coenzyme phenotype since the last common ancestor.

References and Notes

1. S. J. Gould, R. C. Lewontin, *Proc. R. Soc. London Ser. B* **205**, 581 (1979).
2. J. Antonovics, P. H. van Tienderen, *Trends Ecol. Evol.* **6**, 166 (1991).
3. M. R. Rose, G. V. Lauder, *Adaptation* (Academic Press, San Diego, CA, 1996).
4. M. Pigliucci, J. Kaplan, *Trends Ecol. Evol.* **15**, 66 (2000).
5. N. Eldredge *et al.*, *Paleobiology* **31**, 133 (2005).
6. M. Travisano, J. A. Mongold, A. F. Bennett, R. E. Lenski, *Science* **267**, 87 (1995).
7. H. Teotónio, M. R. Rose, *Nature* **408**, 463 (2000).
8. P. Beldade, K. Koops, P. M. Brakefield, *Nature* **416**, 844 (2002).
9. W. A. Frankino, B. J. Zwaan, D. L. Stern, P. M. Brakefield, *Science* **307**, 718 (2005).
10. R. Chen, A. Greer, A. M. Dean, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 12171 (1996).
11. M. Lunzer, S. P. Miller, R. Felsheim, A. M. Dean, *Science* **310**, 499 (2005).
12. A. M. Dean, G. B. Golding, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3104 (1997).
13. G. Zhu, G. B. Golding, A. M. Dean, *Science* **307**, 1279 (2005); published online 13 January 2005 (10.1126/science.1106974).
14. G. Wallon *et al.*, *J. Mol. Biol.* **266**, 1016 (1997).
15. J. H. Hurley, A. M. Dean, *Structure* **2**, 1007 (1994).
16. J. H. Hurley, A. M. Dean, D. E. Koshland Jr., R. M. Stroud, *Biochemistry* **30**, 8671 (1991).
17. T. Penfound, J. W. Foster, in Escherichia coli *and* Salmonella: *Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (ASM Press, Washington, DC, ed. 2, 1996), pp. 721–730.
18. See supporting material on *Science* Online.
19. O. Kuchner, F. H. Arnold, *Trends Biotechnol.* **58**, 1 (1997).
20. F. H. Arnold, P. L. Wintrobe, K. Miyazaki, A. Gershenson, *Trends Biochem. Sci.* **26**, 100 (2001).
21. L. G. Otten, W. J. Quax, *Biomol. Eng.* **22**, 1 (2005).
22. S. Wright, in *Proceedings of the Sixth International Genetics Congress*, D. F. Jones, Ed. (Brooklyn Botanic Garden, Menasha, WI, 1932), pp. 356–366.
23. J. Maynard Smith, *Nature* **225**, 563 (1970).
24. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, *Science* **312**, 111 (2005).
25. B. G. Hall, *Antimicrob. Agents Chemother.* **46**, 3035 (2002).
26. S. Zamenhof, H. H. Eichhorn, *Nature* **216**, 456 (1967).
27. K. J. Andrews, G. D. Hegeman, *J. Mol. Evol.* **8**, 317 (1976).
28. D. E. Dykhuizen, *Evolution* **32**, 125 (1978).
29. A. L. Koch, *J. Mol. Evol.* **19**, 455 (1983).

30. Supported by NIH grant GM060611 (A.M.D.). We thank three anonymous reviewers whose constructive criticisms we found invaluable.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5798/458/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
Tables S1 to S5
References

4 August 2006; accepted 20 September 2006
10.1126/science.1133479

An Essential Role for LEDGF/p75 in HIV Integration

Manuel Llano, Dyana T. Saenz, Anne Meehan, Phonphimon Wongthida, Mary Peretz, William H. Walker, Wulin Teo, Eric M. Poeschla*

Chromosomal integration enables human immunodeficiency virus (HIV) to establish a permanent reservoir that can be therapeutically suppressed but not eradicated. Participation of cellular proteins in this obligate replication step is poorly understood. We used intensified RNA interference and dominant-negative protein approaches to show that the cellular transcriptional coactivator lens epithelium–derived growth factor (LEDGF)/p75 (p75) is an essential HIV integration cofactor. The mechanism requires both linkages of a molecular tether that p75 forms between integrase and chromatin. Fractionally minute levels of endogenous p75 are sufficient to enable integration, showing that cellular factors that engage HIV after entry may elude identification in less intensive knockdowns. Perturbing the p75-integrase interaction may have therapeutic potential.

Integration enables human immunodeficiency virus type 1 (HIV-1) to establish a permanent genetic reservoir that can initiate new virion production, evade immune surveillance, and replicate through mitosis. Integrated proviruses that persist in long-lived T cells ensure rapid HIV recrudescence if antiviral drugs are withdrawn. Integration is catalyzed by the viral integrase (IN). When expressed as a free protein in cells rather than within its normal context as an intravirion cleavage product of the HIV Gag-Pol precursor, IN becomes tethered to chromatin by cellular lens epithelium–derived growth factor/p75 (p75) (*1*–*3*), which is a transcriptional coactivator (*4*). Accordingly, both proteins display tight colocalization with chromatin throughout the cell cycle; short hairpin RNA (shRNA)–mediated knockdown of p75 untethers IN, redistributing it from an entirely nuclear to an entirely cytoplasmic location (*3*). Molecular tethering results from specific linkages formed by p75's discrete functional modules: the N-terminal Pro-Trp-Trp-Pro (PWWP) and A/T-hook elements bind to chromatin (*5*), and a C-terminal integrase-binding domain (IBD) binds to IN (*6*, *7*). p75 also protects the HIV-1 IN protein from rapid degradation in the 26S proteasome (*8*). In the bona fide viral context, drastic knockdown of p75 changed the genomic pattern of HIV-1 integration by reducing the viral bias for active genes, which suggests that p75 influences integration targeting (*9*). However, changes in overall levels of HIV integration and replication have been either absent or modest, and single-cycle infection analyses in cell lines have consistently detected no effect, which has led to questions about the overall importance of p75 in the viral life cycle (*3*, *7*, *9*–*12*).

Previously, we observed that a nuclear localization signal–mutant p75 protein became constitutively chromatin-trapped in stable cell lines (*7*). In the present work, we hypothesized the existence in previous severely RNA interference (RNAi)–depleted HIV-susceptible cells of a very small yet virologically potent chromatin-associated p75 residuum. We reasoned that a fractionally minute residual pool with a spatially favorable location (colocalized with chromatin) could explain the inability to demonstrate substantial, reproducible impairments in integration or viral replication in cells lacking detectable p75. Such a reservoir would be inadequate to affect observable properties of ectopically expressed IN but might be sufficient to engage the vastly less abundant incoming viral preintegration complex.

To test this hypothesis, we performed subcellular fractionation and interrogated chromatin, using a deoxyribonuclease (DNase) I– and salt-based extraction protocol (*13*). These methods

Molecular Medicine Program, Mayo Clinic College of Medicine, Rochester, MN 55905, USA.

*To whom correspondence should be addressed. E-mail: emp@mayo.edu

INTEGR. COMP. BIOL., 45:486–491 (2005)

The Effects of Evolution are Local: Evidence from Experimental Evolution in *Drosophila*[1]

M. R. ROSE,[2] H. B. PASSANANTI, A. K. CHIPPINDALE,[3] J. P. PHELAN,[4] M. MATOS,[5] H. TEOTÓNIO,[6] AND L. D. MUELLER

Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525

SYNOPSIS. One of the enduring temptations of evolutionary theory is the extrapolation from short-term to long-term, from a few species to all species. Unfortunately, the study of experimental evolution reveals that extrapolation from local to general patterns of evolution is not usually successful. The present article supports this conclusion using evidence from the experimental evolution of life-history in *Drosophila*. The following factors demonstrably undermine evolutionary correlations between functional characters: inbreeding, genotype-by-environment interaction, novel foci of selection, long-term selection, and alternative genetic backgrounds. The virtual certainty that at least one of these factors will arise during evolution shreds the prospects for global theories of the effects of adaptation. The effects of evolution apparently don't generalize, even though evolution is a global process.

INTRODUCTION

Few goals are more fervently espoused by scientists than the creation of a general scientific theory that is predictive over a wide range of circumstances. Many of us would point to the power and elegance of Darwin's theory of evolution by natural selection and claim that we have just such a theory. Is this a fair claim? It might be, if we could make content-laden predictions concerning the long-term outcome of evolution. But can we?

A number of evolutionary theories have been advanced which seem to claim, implicitly or explicitly, to predict long-term and general features of evolution. For example, several researchers have put forward theories of phenotypic evolution that offer a complete predictive package, providing one knows the genetic and phenotypic variances, covariances, and higher order moments between characters (*e.g.*, Lande, 1979, 1980; Lande and Arnold, 1983; Turelli and Barton, 1990; see also Barton and Turelli, 1987, 1991; Kirkpatrick *et al.*, 2002). One thing evolutionary biology does not lack is an abundance of theories.

But do these theories really hold up? An advantage for evolutionary theorists is that few people ever apply strong inference (*cf.*, Platt, 1964) to evolution as a process (but see Lenski *et al.*, 1991, for an important example), leaving few of the predictions or assumptions of most evolutionary theories at much risk of experimental refutation. It is difficult to study evolution. A large amount of time is required in most cases, and many samples have to be taken, creating problems of physical scale in the housing of experimental organisms. Thus there are only a few cases where critical tests of evolutionary theories have actually been performed, and several of these tests have used experimental evolution. The microbial work of Lenski (*e.g.*, Lenski *et al.*, 1991; Lenski and Travisano, 1994; Cooper *et al.*, 2001) and others (*e.g.*, Rainey and Travisano, 1998; Burch and Chao, 1999; Dahlberg and Chao, 2003) has strong tests of evolutionary theory (see reviews in Bell, 1997; Travisano and Rainey, 2000).

We have an outbreeding experimental evolution system, laboratory-evolved *Drosophila* populations (*vid.* Rose *et al.*, 2004). In the first ten years that they were studied, we developed a simple consensus model for the effects of adaptation in these fruit flies. The interesting point for the present purpose is that this model was to be annihilated by the next decade's worth of work. From this destruction, we learned a great deal about the robustness of evolutionary findings, as we will now adumbrate. We are sure that few of our colleagues will mind if we only set ourselves up to be demolished, thereby sparing them the injury or insult.

DROSOPHILA LIFE-HISTORY: THE STANDARD MODEL

The study of *Drosophila* life-history in the laboratory goes back to the 1920s (*e.g.*, Pearl and Parker, 1922), if not earlier (Loeb and Northrop, 1917). Much of this work used mutant or inbred stocks, creating problems that we will discuss shortly. Reasonable fruit fly work on life-history is not much older than the 1960s (*e.g.*, Wattiaux, 1968). A major feature of the modern era of fruit fly life-history research is the use of large quantitative genetics experiments (*e.g.*, Rose and Charlesworth, 1981; Hutchinson and Rose, 1991; Hutchinson *et al.*, 1991; Hughes and Charlesworth, 1994) and replicated experimental evolution (*e.g.*, Rose, 1984*a*; Luckinbill *et al.*, 1984; Service *et al.*, 1988; Partridge and Fowler, 1992; Mueller *et al.*, 1993). Sometimes, the findings from these two types

[1] From the Symposium *Selection Experiments as a Tool in Evolutionary and Comparative Physiology: Insights into Complex Traits* presented at the Annual Meeting of the Society for Integrative and Comparative Biology, 5–9 January 2004, at New Orleans, Louisiana.

[2] E.mail: mrrose@uci.edu

[3] Dept. of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.

[4] Present address: Dept. of Organismal Biology, Ecology, and Evolution, University of California, Los Angeles, California 90095-1606.

[5] Present address: Centro de Biologia Ambiental/Dep. Biologia Animal, Fac. Ciências Univ. Lisboa, Lisboa, Portugal.

[6] Present address: Instituto Gulbenkian de Ciência, 2780–156 Oeiras, Portugal.

TABLE 1. *The Irvine* Drosophila *experimental evolution system.*

- started from a SINGLE endemic fruit fly population, IV, in 1975; large sample
- Reasonable Ne's-about 1,000 plus
- Adapted to lab for about five years first
- All selection with five populations
- All selection regimes paired with controls
- Long sustained selection regimes

Some of the Laboratory Evolution Regimes:

- B-selected for day 14 fertility in vials
- O-selected for day 70 fertility in cages
- D-selected to survive extreme desiccation
- C-selected to survive moderate starvation
- SO-selected to survive extreme starvation
- CO-selected for day 28 fertility in cages
- SB-selected to survive extreme starvation
- CB-selected for day 28 fertility in cages
- ACO-selected for early (day 7–9) fertility
- ACB-selected for early (day 7–9) fertility
- RSO-relaxed selection, like CO's
- NDO-new D stocks
- NDCO-new C stocks

of experiments reinforce each other, as in cases of negative genetic correlations and antagonistic indirect responses to selection. However, most of our knowledge of the functional interrelations between *Drosophila* life-history characters has come from studies of experimental evolution (*cf.* Rose *et al.*, 2004); see Table 1 for a brief summary.

From the mid-seventies to 1990, the overall pattern in the results from *Drosophila* work on life-history was fairly clear. In outbred fruit flies, early fecundity generally traded-off with longevity. Longer-lived flies had reduced early fecundity, and vice versa. Longer-lived flies had increased later fecundity. Longer-lived flies were more robust under several stressors: starvation, desiccation, and ambient ethanol. Starvation resistance appeared to trade-off with fitness, while desiccation resistance did not. A number of additional subtleties could be added to this model, but it contains the highlights. (See Table 2 for a rough summary of the initially inferred patterns.) Several labs contributed to these basic findings, and quite a few more individual investigators, using flies of different origins. (Rose *et al.* [2004] supply an introduction to this research.) It would have been reasonable to conclude that some fundamental truths had been discovered about *Drosophila* life-history, and perhaps life-history in general.

Below we will destroy this standard model, mostly using the data that we collected in the period after 1990. Since we are undermining our own pet hypotheses, rather than anyone else's, we can afford to be brutally matter-of-fact.

Inbreeding

The ideal finding in science is one that applies regardless of initial conditions. Many of the theories of physics apply with absolute generality, such as the velocity-dependent transformations of special relativity. Some theories in biology have this property. Darwinian evolution implies that extinct species will never reappear unaltered millions of years later. It would be nice if, for example, such an important idea as a negative genetic correlation between early reproduction and later survival was always true (*cf.* Williams, 1957). Such a conclusion is especially attractive when a result of this kind has been obtained repeatedly (Rose and Charlesworth, 1980; Rose, 1984*a*; Luckinbill *et al.*, 1984) and there are explicit mathematical studies that predict the occurrence of such negative correlations at evolutionary equilibrium (*e.g.*, Rose, 1982).

But it was not to be. One of the anomalies facing this result is that a common observation in studies of fruit fly life-history has been generally positive correlations among life-history characters, especially positive genetic correlations (*e.g.*, Giesel *et al.*, 1982). Perhaps there is no general trade-off pattern?

In a sense, this conclusion was correct. When Rose (1984*b*) derived inbred flies from the stock that had shown a trade-off (*e.g.*, Rose, 1984*a*), the genetic correlations became generally positive. Rose interpreted this result as a reflection of inbreeding depression, with some inbred lines more inbred than others and so generally having reduced life-history characters. As inbred lines vary in their degree of inbreeding, and thus in the depression of their life-history characters, life-history characters will positively co-vary. This effect apparently swamps the negative genetic correlation between early reproduction and longevity. Such trade-offs are not robust under inbreeding. Rather, they are ''local'' to outbred populations.

This was the first demonstration of the lack of universality of the ''standard model'' for *Drosophila* life-history evolution. However, at the time it was felt that barring cases of inbred flies was a reasonable qualification to the standard model. This lack of robustness was not treated as a source of concern. Worse was to follow.

TABLE 2. *The matrix of evolutionary genetic correlations that make up part of the Standard Model.*

	Longevity	Fecundity	Starv. resist.	Desic. resist.	Development	Viability
Longevity	-	neg	pos	pos	neg	pos
Fecundity	neg	-	neg	x	pos	neg
Starvation resistance	pos	neg	-	pos	pos	x
Desiccation resistance	pos	x	pos	-	x	x
Development	neg	pos	pos	x	-	pos
Viability	pos	neg	x	x	pos	-

(x - no correlation inferred; - same character).

Genotype-by-environment interaction

Flies that have been recently sampled from nature are not near evolutionary equilibrium with respect to the laboratory (*vid.* Matos *et al.*, 2000). They undergo a process of rapid adaptation to the laboratory during which several life-history characters improve. Since fly populations inevitably vary in the degree to which they initially are adapted to the laboratory, they will vary up and down for many of their life-history characters in laboratory assays, again producing positive genetic correlations between life-history characters. This was shown in laboratory-adapted fruit flies by giving them a novel environment, and comparing genetic correlations in their normal lab environment *versus* the novel environment (Service and Rose, 1985). As expected, under novel environmental conditions the genetic correlation between fecundity and starvation resistance shifted toward positive values.

This illustrated the dependence of genetic correlations on the environment to which organisms are exposed, in addition to the dependence of these correlations on the degree of inbreeding. Change the environment and the genetic correlation changes. If the environment is novel, there is a tendency to express positive genetic correlations. This result is probably not as robust as the inbreeding result—some novel environments might preserve negative genetic correlations by chance. Still, there is a circumscription of the standard model.

Further evolution of stocks selected for postponed aging led to a reduction in the trade-off between longevity and early fecundity (Hutchinson and Rose, 1991; Chippindale *et al.*, 1993; Leroi *et al.*, 1994*a*). Eventually, the longer-lived stocks even exhibited *increased* early fecundity, compared to the ancestral type of stock. This posed an obvious problem for our understanding of trade-offs in life-history. No inbreeding or novel environment appeared to be involved. But extensive testing for genotype-by-environment interaction revealed that the early fecundity of long-lived stocks was nonetheless reduced specifically under the environmental conditions used to culture the ancestral fruit fly stock: crowding, bad food, and a short opportunity for egg laying (Leroi *et al.*, 1994*a*). Under appropriate environmental conditions, the original trade-offs would reappear (Leroi *et al.*, 1994*a*, *b*).

Novel and long-term selection

Up to this point, it was still possible to regard these difficulties for the standard *Drosophila* life-history model as experimental artifacts (*cf.* Rose, 1991, Ch. 3–4). But greater difficulties were to come.

One of the areas that the standard model was extended to was the evolution of development. We found an apparent trade-off between rate of development and viability (Chippindale *et al.*, 1994). This was a natural elaboration of the standard model in that it suggests a trade-off between rapidly developing an adult and the survival of the larva. Chippindale *et al.* (1997) successfully selected for accelerated development in the *Drosophila* stocks that had been used to develop the standard model. The rapidly developing flies had reduced viability, too. Borash *et al.* (2000) also found that these faster developing flies were more vulnerable to noxious environments. In these respects, the larval evolutionary patterns seemed to fit the kind of trade-off pattern built into our standard model of *Drosophila* life-history evolution.

It was only when more detailed analyses of growth rate were performed that problems appeared. When Chippindale and collaborators analyzed growth rate using measurements of dry body mass instead of thorax length, this trade-off disappeared (Chippindale *et al.*, 2004). The correlation between growth rate and viability went from negative to positive, as a function of the specific trait that was measured, say mass *versus* thorax length. In other words, the evidence for a trade-off was dependent on how the traits were characterized.

The populations that were originally used to develop the standard model underwent continued selection. The total number of generations of selection came to exceed 100 for most of these stocks, as opposed to 20 or 30 generations, the number of generations of selection that characterized the stocks when they were first studied. At that earlier point in the evolution of our populations, we had a positive genetic correlation between stress resistance and longevity. (See Table 2.) After more than 100 generations of experimental evolution, we re-analyzed the relationship between stress resistance and longevity (Phelan *et al.*, 2003), finding that the positive correlation between stress resistance and longevity disappeared at high levels of stress resistance. There was even evidence for a negative relationship between high levels of starvation resistance and longevity. Because this correlation breakdown arose in a miscellany of stocks, we proceeded to select specifically for very high levels of stress resistance to determine its effects on longevity (Archer *et al.*, 2003), without confounding selection. Again, the positive correlation built into the standard model broke down.

The pattern of the selection results was fairly simple. So long as selection didn't push functional characters too far, our standard ideas about viability, development, fecundity, longevity, and stress resistance held up fairly well. But if we pushed selection hard, producing substantial enhancements in these functional characters, the standard model collapsed. In other words, our standard model was only a local finding.

Genetic background

There are some findings that do seem to be highly robust. For example, the effect of delayed reproduction on the laboratory evolution of *Drosophila* appears quite reliable: longevity increases (Wattiaux, 1968; Rose and Charlesworth, 1980; Rose, 1984*a*; Luckinbill *et al.*, 1984; Partridge and Fowler, 1992). Passananti (2000) performed such a late-reproduction study using hybrids of four inbred *rosy D. melanogaster* stocks:

TABLE 3. *The evolutionary effects of postponed reproduction in* rosy *stocks, generation 22.*

	B_{ry}	O_{ry}
Male longevity* (days)	46.52 ± 1.22	54.76 ± 1.14
Female longevity* (days)	36.98 ± 0.59	47.63 ± 1.49
Early fecundity	31.80 ± 3.51	17.64 ± 2.17
Male starvation resistance (hours)	24.87 ± 2.28	22.08 ± 1.29
Female starvation resistance (hours)	33.68 ± 3.78	28.11 ± 1.22

(* Indicates $P < 0.05$ in paired *t*-tests for treatment differences with 5 replicates; results are given as means ± standard errors).

Canton-S, Oregon-R, Swedish-C, and Lausanne. These stocks and their controls were created using independent crosses. Once the starting stocks were created, five populations were subjected to selection for early reproduction, the B *ry*, while the other five were subject to selection for late reproduction, the O *ry*. Selection proceeded for 22 O generations, and many more B generations, before samples were taken for assay. Two generations of standardized rearing were used before data were collected. The results are shown in Table 3. (Note that all statistical comparisons are between treatments, so that the number of replicate lines [not individuals] limits the degrees of freedom, which in turn makes the inference of statistical significance quite conservative.)

As in earlier studies, average longevity significantly increased in the O *ry* stocks. This is in keeping with the findings of Rose (1984*a*), a study that employed fewer generations. The chief interest of the results of Table 3, however, is that the indirect responses of starvation resistance and early fecundity are not in keeping with the standard model. There is no statistically significant decrease in fecundity or increase in starvation resistance at generation 22. While the linear regression of average fecundity in O *ry* stocks does significantly trend downward when multiple generations of data are used ($P < 0.05$; data not shown here; Passananti, 2000), the starvation resistance results are not even in the right direction. Here, as in the findings of Phelan *et al.* (2003) and Archer *et al.* (2003), the qualitative correlation between starvation resistance and longevity is undermined. Using a different genetic background breaks the positive correlation between starvation resistance and longevity.

CONCLUSION: WHERE DO WE GO FROM HERE?

The standard *Drosophila* model for life-history evolution arose first in the 1970s. It is now more than 30 years old. Much of its history is outlined in Rose *et al.* (2004). But our recent research is inimical to the standard model. As a set of precepts about life-history evolution in a particular system, the standard model should be abandoned.

What is the general import of this conclusion? There is the question of whether or not other evolutionary systems will have the same features. In general, we do not know the answer to this question. There aren't many studies of experimental evolution. Of these studies, very few compare with our *Drosophila* work in terms of the number of generations, replicates, or distinct selection regimes utilized. The Luckinbill laboratory has performed somewhat similar research (*e.g.*, Luckinbill *et al.*, 1984). Interestingly, one of their studies demonstrated the existence of a genotype-by-environment interaction involving rearing density (Clare and Luckinbill, 1985), a finding that was later corroborated in our system (Service *et al.*, 1988).

An experimental evolution system that has been even more replicated is the *Escherichia coli* model system established by Lenski and his colleagues, beginning with Lenski *et al.* (1991). This system has been studied for thousands of generations, and some additional lines have been created that focus on particular characters, such as adaptation to temperature (*e.g.*, Bennett *et al.*, 1992). Like the original standard model for *Drosophila* life-history evolution, it would be fair to say that Lenski and colleagues have created a standard model for *E. coli*. But how global is it? Will it too breakdown as they learn more?

Consider the possibility that the destruction of the *Drosophila* model will prove generally true, if not for all organisms perhaps, then at least for metazoa. That is, what if none of the patterns that we adduce for the effects of evolution on animals hold up when we learn more? Natural selection may be a process that rapaciously exploits new advantageous alleles and allele combinations to increase fitness, often in ways that undermine antecedent limits on adaptation. If so, then it is only appropriate to expect that simple evolutionary just-so stories will not be sustained when enough is learned about the range of pertinent evolutionary dynamics. What can we do about this prospect?

We could track the accomplishments of evolution the way market analysts track the stock market, always searching for the latest pattern. Ever more complex models could explain observed patterns with increasing precision, without gaining predictive power. There might be an alternative, however: focusing only on the dynamical machinery of evolution independently of the outcome of evolution. With this approach, the *workings* of the process would be studied, eschewing any prospect of generally characterizing the *effects* of the evolutionary process. This leads to a focus on testable predictions concerning the evolutionary mechanisms involved: 1) We might test whether standing genetic variation plays a predominant role in the response to selection (*cf.* Teotónio and Rose, 2000); 2) Similarly, we could determine if new mutations were involved in the response to selection (*e.g.*, Mackay, 1985), and if their effect is dependent on population structure (*e.g.*, Estes and Lynch, 2003); 3) The relative role of additive and non-additive gene interactions can be tested for their role in inbreeding depression (*e.g.*, Vassilieva *et al.*, 2000); 4) Hypotheses about how genetic drift and selection change the patterns of genetic variance and covariance can be addressed (*e.g.*, Whitlock *et al.*, 2002); 5) Specific forms of natural selection, such as density and frequency-dependent selec-

tion, can be examined for their prevalence (*e.g.*, Mueller *et al.*, 1993; Reznick *et al.*, 1996); 6) The relationship between evolutionary rate and initial differentiation is sometimes strong, linear, and negative in slope (*e.g.*, Teotónio and Rose, 2000), perhaps because of the greater magnitude of selection differentials when there is more differentiation—a testable finding that may not depend on local features of evolution. But we should always be prepared to discover, and document, that our expectations are not met, even for hypotheses about basic mechanisms of evolution.

Some might conclude that we have shown that experimental evolution is of little value for evolutionary research. On the contrary, we propose that experimental evolution is one of the most powerful techniques in evolutionary biology, powerful enough to reveal the unreliability of most conclusions that have been adduced concerning evolution.

ACKNOWLEDGMENTS

We are grateful for the encouragement of T.J. Bradley, T. Garland, and J.G. Swallow. The manuscript was read in draft by C.L. Rauser. The research described herein was supported over the years by the Natural Science and Engineering Research of Canada, the National Institutes of Health of the USA, and the National Science Foundation of the USA.

REFERENCES

Archer, M. A., J. P. Phelan, K. A. Beckman, and M. R. Rose. 2003. Breakdown in correlations during laboratory evolution. II. Selection on stress resistance in *Drosophila* populations. Evolution 57:536–543.

Barton, N. H. and M. Turelli. 1987. Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet. Res. Camb. 49:157–173.

Barton, N. H. and M. Turelli. 1991. Natural and sexual selection on many loci. Genetics 127:229–255.

Bell, G. 1997. *Selection*: *The mechanism of evolution*. Chapman and Hall, New York.

Bennett, A. F., R. E. Lenski, and J. E. Mittler. 1992. Evolutionary adaptation to temperature. I. Fitness responses of *Escherichia coli* to changes in its thermal environment. Evolution 46:16–30.

Borash, D. J., H. Teotónio, M. R. Rose, and L. D. Mueller. 2000. Density-dependent natural selection in *Drosophila*: Correlations between feeding rate, development time, and viability. J. Evol. Biol. 13:181–187.

Burch, C. L. and L. Chao. 1999. Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics 151:921–927.

Chippindale, A. K., J. A. Alipaz, H-W. Chen, and M. R. Rose. 1997. Experimental evolution of accelerated development in *Drosophila*. 1. Larval development speed and survival. Evolution 51:1536–1551.

Chippindale, A. K., J. A. Alipaz, and M. R. Rose. 2004. Experimental evolution of accelerated development in *Drosophila*. 2. Adult fitness and the fast development syndrome. *In* M. R. Rose, H. B. Passananti, and M. Matos (eds.), *Methuselah flies: A case study in the evolution of aging*. World Scientific Publishing, Singapore. (In press)

Chippindale, A. K., D. T. Hoang, P. M. Service, and M. R. Rose. 1994. The evolution of development in *Drosophila melanogaster* selected for postponed senescence. Evolution 48:1880–1899.

Chippindale, A. K., A. M. Leroi, S. B. Kim, and M. R. Rose. 1993. Phenotypic plasticity and selection in *Drosophila* life-history evolution. I. Nutrition and the cost of reproduction. J. Evol. Biology 6:171–193.

Clare, M. J. and L. S. Luckinbill. 1985. The effects of gene-environment interaction on the expression of longevity. Heredity 55: 19–29.

Cooper, T. F., D. E. Rozen, and R. E. Lenski. 2001. Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli. PNAS 100:1072–1077.

Dahlberg, C. and L. Chao. 2003. Amelioration of the cost of conjugative plasmid carriage in Eschericha coli K12. Genetics 165: 1641–1649.

Estes, S. and M. Lynch. 2003. Rapid fitness recovery in mutationally degraded lines of Caenorhabditis elegans. Evolution 57:1022–30.

Giesel, J. T., P. A. Murphy, and M. N. Manlove. 1982. The influence of temperature on genetic interrelationships of life history traits in a population of *Drosophila melanogaster*: What tangled data sets we weave. Am. Nat. 119:464–479.

Hughes, K. A. and B. Charlesworth. 1994. A genetic analysis of senescence in *Drosophila*. Nature 367:64–66.

Hutchinson, E. W. and M. R. Rose. 1991. Quantitative genetics of postponed aging in *Drosophila melanogaster*. I. Analysis of outbred populations. Genetics 127:719–727.

Hutchinson, E. W., A. J. Shaw, and M. R. Rose. 1991. Quantitative genetics of postponed aging in *Drosophila melanogaster*. II. Analysis of selected lines. Genetics 127:729–737.

Kirpatrick, M., T. Johnson, and N. H. Barton. 2002. General models of multilocus evolution. Genetics 161:1727–1750.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: Body size allometry. Evolution 33:402–416.

Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. Genetics 94:203–215.

Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Lenski, R. E., M. R. Rose, S. E. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. Am. Nat. 138: 1315–1341.

Lenski, R. E. and M. Travisano. 1994. Dynamics of adaptation and diversification: A 10,000- generation experiment with bacterial populations. Proc. Nat. Acad. Sci. U.S.A. 91:6808–6814.

Leroi, A. M., A. K. Chippindale, and M. R. Rose. 1994*a*. Long-term laboratory evolution of a genetic trade-off in *Drosophila melanogaster*. I. The role of genotype × environment interaction. Evolution 48:1244–1257.

Leroi, A. M., W. R. Chen, and M. R. Rose. 1994*b*. Long-term laboratory evolution of a genetic trade-off in *Drosophila melanogaster*. II. Stability of genetic correlations. Evolution 48:1258–1268.

Loeb, J. and J. H. Northrop. 1917. On the influence of food and temperature upon the duration of life. J. Biol. Chem. 32:103–121.

Luckinbill, L. S., R. Arking, M. J. Clare, W. C. Cirocco, and S. A. Buck. 1984. Selection for delayed senescence in *Drosophila melanogaster*. Evolution 38:996–1003.

Mackay, T. F. 1985. Transposable element-induced response to artificial selection in Drosophila melanogaster. Genetics 111:351–374.

Matos, M., T. Avelar, and M. R. Rose. 2002. Variation in the rate of convergent evolution: Adaptation to a laboratory environment in *Drosophila subobscura*. J. Evol. Biol. 15:673–682.

Mueller, L. D., J. L. Graves, and M. R. Rose. 1993. Interactions between density-dependent and age-specific selection in *Drosophila melanogaster*. Funct. Ecol. 7:469–479.

Partridge, L. and K. Fowler. 1992. Direct and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. Evolution 46:76–91.

Passananti, H. B. 2000. Studies of postponed aging in *Drosophila melanogaster*. Ph.D. Diss., Biological Sciences, University of California, Irvine.

Pearl, R. and S. L. Parker. 1922. Experimental studies on the duration of life. II. Hereditary differences in duration of life in live-breed strains of *Drosophila*. Am. Nat. 58:71–82.

Phelan, J. P., M. A. Archer, K. A. Beckman, A. K. Chippindale, T.

J. Nusbaum, and M. R. Rose. 2003. Breakdown in correlations during laboratory evolution. I. Comparative analyses of *Drosophila* populations. Evolution 57:527–535.
Platt, J. R. 1964. Strong inference. Science 146:347–353.
Rainey, P. B. and M. Travisano. 1998. Adaptive radiation in a heterogeneous environment. Nature 394(6688):69–72.
Reznick, D. N., M. J. Butler, IV, F. H. Rodd, and P. Ross. 1996. Life-history evolution in guppies Poecilia reticulata. VI. Differential mortality as a mechanism for natural selection. Evolution 50:1651–1660.
Rose, M. R. 1982. Antagonistic pleiotropy, dominance, and genetic variation. Heredity 48:63–78.
Rose, M. R. 1984*a*. Laboratory evolution of postponed senescence in *Drosophila melanogaster*. Evolution 38:1004–1010.
Rose, M. R. 1984*b*. Genetic covariation in *Drosophila* life history: Untangling the data. Am. Nat. 123:565–569.
Rose, M. R. 1991. *Evolutionary biology of aging*. Oxford University Press, New York.
Rose, M. R. and B. Charlesworth. 1980. A test of evolutionary theories of senescence. Nature 287:141–142.
Rose, M. R. and B. Charlesworth. 1981. Genetics of life history in Drosophila melanogaster. I. Sib analysis of adult females. Genetics 97:173–186.
Rose, M. R., H. B. Passananti, and M. Matos. (eds.) 2004. *Methuselah flies: A case study in the evolution of aging*. World Scientific Publishing, Singapore.
Service, P. M. and M. R. Rose. 1985. Genetic covariation among life history components: The effect of novel environments. Evolution 39:943–945.
Service, P. M., E. W. Hutchinson, and M. R. Rose. 1988. Multiple genetic mechanisms for the evolution of senescence in *Drosophila melanogaster*. Evolution 42:708–716.
Turelli, M. and N. H. Barton. 1990. Dynamics of polygenic characters under selection. Theor. Popul. Biol. 38:1–57.
Teotónio, H. and M. R. Rose. 2000. Variation in the reversibility of evolution. Nature 408:463–466.
Travisano, M. and P. B. Rainey. 2000. Studies of adaptive radiation using model microbial systems. Am. Nat. 156:S35–S44 Suppl. S.
Vassilieva, L. L., A. M. Hook, and M. Lynch. 2000. The fitness effect of spontaneous mutations in *Caenorhabditis elegans*. Evolution 54:1234–1246.
Wattiaux, J. M. 1968. Cumulative parental effects in *Drosophila subobscura*. Evolution 22:406–421.
Whitlock, M. C., P. C., Phillips, and K. Fowler. 2002. Persistence of changes in the genetic covariance matrix after a bottleneck. Evolution 56:1968–75.
Williams, G. C. 1957. Pleiotropy, natural selection, and the evolution of senescence. Evolution 11:398–411.

This inevitably leads to widespread haploinsufficiency at several gene loci, only a fraction of which provide the nascent tumor cell with some degree of selective advantage. Do tumor suppressor genes exist for which haploinsufficiency is more strongly selected for than complete inactivation? Only accurate and quantitative genome-wide expression profiling by microarray or proteomic analysis will enable such gene-dosage defects to be identified. Analyzing targeted hypomorphic alleles in experimental animals should facilitate the identification of modifier genes, their tissue-specific dosage thresholds, and their interaction with more penetrant tumor suppressor genes and environmental mutagens.

References

1. A. G. Knudson Jr., *Proc. Natl. Acad. Sci. U.S.A.* **68**, 820 (1971).
2. S. B. Gruber *et al.*, *Science* **297**, 2013 (2002).
3. K. H. Goss *et al.*, *Science* **297**, 2051 (2002).
4. K. Spring *et al.*, *Nature Genet.* **32**, 185 (2002).
5. S. Venkatachalam *et al.*, *EMBO J.* **17**, 4657 (1998).
6. W. J. Song *et al.*, *Nature Genet.* **23**, 166 (1999).
7. R. Smits *et al.*, *Gastroenterology* **119**, 1045 (2000).
8. M. L. Fero *et al.*, *Nature* **396**, 177 (1998).
9. M. Kucherlapati *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9924 (2002).
10. M. Swift *et al.*, *N. Engl. J. Med.* **325**, 1831 (1991).
11. N. G. Howlett *et al.*, *Science* **297**, 606 (2002).
12. H. Meijers-Heijboer *et al.*, *Nature Genet.* **31**, 55 (2002).
13. K. Inoue *et al.*, *Genes Dev.* **15**, 2934 (2001).
14. H. Miyoshi *et al.*, *Cancer Res.* **62**, 2261 (2002).
15. Y. Zhu *et al.*, *Science* **296**, 920 (2002).
16. C. Wetmore *et al.*, *Cancer Res.* **60**, 2239 (2000).
17. R. H. Zurawel *et al.*, *Genes Chr. Cancer* **28**, 77 (2000).
18. B. Kwabi-Addo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11563 (2001).

PERSPECTIVES: SYSTEMS BIOLOGY

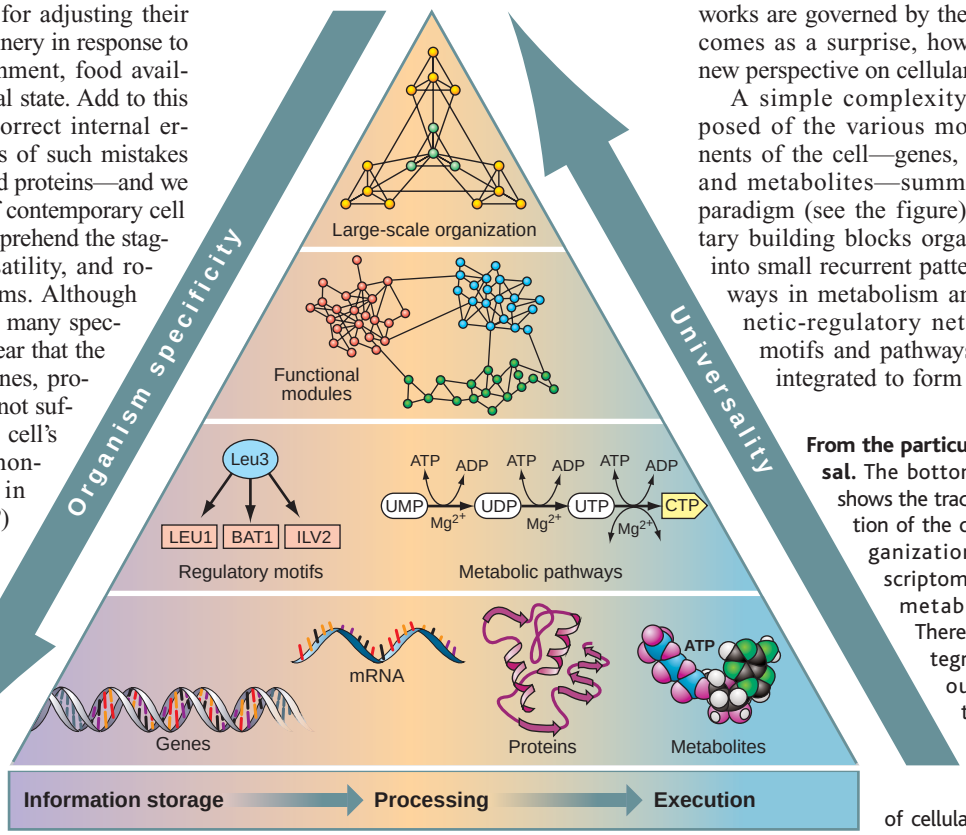
Life's Complexity Pyramid

Zoltán N. Oltvai and Albert-László Barabási

Cells and microorganisms have an impressive capacity for adjusting their intracellular machinery in response to changes in their environment, food availability, and developmental state. Add to this an amazing ability to correct internal errors—battling the effects of such mistakes as mutations or misfolded proteins—and we arrive at a major issue of contemporary cell biology: our need to comprehend the staggering complexity, versatility, and robustness of living systems. Although molecular biology offers many spectacular successes, it is clear that the detailed inventory of genes, proteins, and metabolites is not sufficient to understand the cell's complexity (*1*). As demonstrated by two papers in this issue—Lee *et al.* (*2*) on page 799 and Milo *et al.* (*3*) on page 824—viewing the cell as a network of genes and proteins offers a viable strategy for addressing the complexity of living systems.

According to the basic dogma of molecular biology, DNA is the ultimate depository of biological complexity. Indeed, it is generally accepted that information storage, information processing, and the execution of various cellular programs reside in distinct levels of organization: the cell's genome, transcriptome, proteome, and metabolome. However, the distinctness of these organizational levels has recently come under fire. For example, although long-term information is stored almost exclusively in the genome, the proteome is crucial for short-term information storage (*4, 5*) and transcription factor–controlled information retrieval is strongly influenced by the state of the metabolome. This integration of different organizational levels increasingly forces us to view cellular functions as distributed among groups of heterogeneous components that all interact within large networks (*6, 7*). There is clear evidence for the existence of such cellular networks: For example, the proteome organizes itself into a protein interaction network and metabolites are interconverted through an intricate metabolic web (*7*). The finding that the structures of these networks are governed by the same principles comes as a surprise, however, offering a new perspective on cellular organization.

A simple complexity pyramid composed of the various molecular components of the cell—genes, RNAs, proteins, and metabolites—summarizes this new paradigm (see the figure). These elementary building blocks organize themselves into small recurrent patterns, called pathways in metabolism and motifs in genetic-regulatory networks. In turn, motifs and pathways are seamlessly integrated to form functional mod-



CREDIT: KATHARINE SUTLIFF/*SCIENCE*

From the particular to the universal. The bottom of the pyramid shows the traditional representation of the cell's functional organization: genome, transcriptome, proteome, and metabolome (level 1). There is remarkable integration of the various layers both at the regulatory and the structural level. Insights into the logic of cellular organization can be achieved when we view the cell as a complex network in which the components are connected by functional links. At the lowest level, these components form genetic-regulatory motifs or metabolic pathways (level 2), which in turn are the building blocks of functional modules (level 3). These modules are nested, generating a scale-free hierarchical architecture (level 4). Although the individual components are unique to a given organism, the topologic properties of cellular networks share surprising similarities with those of natural and social networks. This suggests that universal organizing principles apply to all networks, from the cell to the World Wide Web.

Z. N. Oltvai is in the Department of Pathology, Northwestern University, Chicago, IL 60611, USA. E-mail: zno008@nwu.edu A.-L. Barabási is in the Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: alb@nd.edu

ules—groups of nodes (for example, proteins and metabolites) that are responsible for discrete cellular functions (*6*). These modules are nested in a hierarchical fashion and define the cell's large-scale functional organization (*8*).

The papers by Lee *et al.* (*2*) and Milo *et al*. (*3*) offer key support for the cellular organization suggested by the complexity pyramid (see the figure). Using 106 tagged transcription factors of the budding yeast *Saccharomyces cerevisiae*, Lee *et al*. have systematically identified the genes to whose promoter regions these transcription factors (regulators) bind. After establishing transcription factor binding at various confidence levels, they uncovered from 4000 to 35,000 genetic-regulatory interactions, generating the most complete map of the yeast regulatory network to date. The map allows the authors to identify six frequently appearing motifs, ranging from multi-input motifs (in which a group of regulators binds to the same set of promoters) to regulatory chains (alternating regulator-promoter sequences generating a clear temporal succession of information transfer). A similar set of regulatory motifs was recently uncovered in the bacterium *Escherichia coli* by Alon and co-workers (*9*). In their new study, Milo, Alon and colleagues provide evidence that motifs are not unique to cellular regulation but emerge in a wide range of networks, such as food webs, neural networks, computer circuits, and even the World Wide Web (*3*). They identified small subgraphs that appear more frequently in a real network than in its randomized version. This enabled them to distinguish coincidental motifs from recurring significant patterns of interconnections.

An important attribute of the complexity pyramid is the gradual transition from the particular (at the bottom level) to the universal (at the apex). Indeed, the precise repertoire of components—genes, metabolites, proteins—is unique to each organism. For example, 43 organisms for which relatively complete metabolic information is available share only ~4% of their metabolites (*7*). Key metabolic pathways are frequently shared, however, and—as demonstrated in this issue (*2, 3*) and elsewhere (*9*)—so are some of the motifs. An even higher degree of universality is expected at the module level; although quantitative evidence is lacking, it is generally believed that key properties of functional modules are shared across most species. The hierarchical relationship among modules, in turn, appears to be quite universal, shared by all examined metabolic (*8*) and protein interaction networks. Finally, the scale-free nature (*7*) of the network's large-scale organization is known to characterize all intracellular relationships documented in metabolic, protein interaction, genetic, and protein domain networks. The Milo *et al*. study now raises the possibility that the complexity pyramid might not be specific only to cells. Indeed, scale-free connectivity with embedded hierarchical modularity has been documented for a wide range of nonbiological networks. Motifs are now known to be abundant in networks as different as ecosystems and the World Wide Web.

These results highlight some of the challenges systems biology will face in the coming years. Lately, we have come to appreciate the power of maps—reliable depositories of molecular interactions. Yet existing maps are woefully incomplete; key links between different organizational levels are missing. For example, we lack the systematic tools to map out lipid-protein or metabolite–transcription factor interactions in vivo. The topological relationships among pathways, motifs, modules, and the full network will also have to be studied in much more detail. Most important, maps must be complemented with detailed measurements of cellular dynamics, recording the timing of processes that take place along the links. This topic is increasingly studied within isolated motifs and modules (*10*) but has received relatively scant attention at the whole-network level. Despite all of these recent challenges, an initial framework offering a rough roadmap appears to have been established. As we seek further insights, we increasingly understand that our quest to capture the system-level laws governing cell biology in fact represents a search for the deeper patterns common to complex systems and networks in general. Therefore, cell biologists, engineers, physicists, mathematicians, and neuroscientists will need to equally contribute to this fantastic voyage.

References

1. H. Kitano, *Science* **295**, 1662 (2002).
2. T. I. Lee *et al.*, *Science* **298**, 799 (2002).
3. R. Milo *et al.*, *Science* **298**, 824 (2002).
4. D. Bray, *Nature* **376**, 307 (1995).
5. U. S. Bhalla, R. Iyengar, *Science* **283**, 381 (1999).
6. L. H. Hartwell *et al.*, *Nature* **402**, C47 (1999).
7. H. Jeong *et al.*, *Nature* **407**, 651 (2000).
8. E. Ravasz *et al.*, *Science* **297**, 1551 (2002).
9. S. S. Shen-Orr *et al.*, *Nature Genet.* **31**, 64 (2002).
10. J. Hasty *et al.*, *Nature Rev. Genet.* **2**, 268 (2001).

PERSPECTIVES: ARCHAEOLOGY

Climate and Human Migrations

Tom D. Dillehay

Archaeological records are affected by a variety of natural and cultural processes at a variety of spatial and temporal scales (*1*). A given cultural phenomenon may appear across a broad range of environments, or may be limited to a narrow range of environments and time periods. Paleoecological studies can help to discriminate between these cases. But most reconstructions of early human ecosystems are based on the excavation and interpretation of individual archaeological sites. Paleoecological studies of long-term climatic change are also often limited in scope (*2*).

Integrative studies of multiple sites, multiple records, and larger areas over long time periods can dramatically change the interpretation (*3*–*7*). On page 821of this issue, Núñez *et al*. (*8*) demonstrate the power of such a comprehensive approach. They closely integrate paleoecological and archaeological analysis to study the long-term interaction between hunter-gatherers and changing environments over the last 15,000 years in the Atacama desert of northern Chile.

The authors examine why initial human occupation occurred about 2000 years later in this hyperarid region than in more humid forested regions in south central Chile (*9*), and several centuries later than in less arid areas in the central and southern Andes. They also ask why a long "Silencio Arqueologico" (a cultural hiatus in the archaeological record) took place between 9500 and 4500 calendar years before the present (cal yr B.P.).

The possible reasons for these variations in human presence considered by Núñez *et al.* include migration lags, inhospitable late Pleistocene environments, biased survey and visibility, and rapid and long-term abandonment of the region. The study illustrates the importance of integrating local environmental and archaeological information in studying regional human ecosystems and in comparing the findings with other regions at a larger scale.

The authors assume that high-altitude ancient lakes (paleolakes), mid-altitude grasslands (puna), and low-altitude wetlands best indicate changes in habitat ex-

The author is in the Department of Anthropology, University of Kentucky, Lexington, KY 40506, USA. E-mail: dilleha@uky.edu