The Music of Life Sourcebook

The Music of Life

Sourcebook

(version 6, August 2016)

The Music of Life is a short book. Deliberately so. The aim was to use metaphorical stories and surprising ways of explaining modern biology to jolt the reader away from many of the serious misunderstandings of biological science that developed during the second half of the 20th century as molecular biology came to dominate the scene.

There is nothing wrong with molecular biology as science. In leading up to the sequencing of the human genome it represents one of man's greatest scientific achievements. But there is everything wrong in the way in which it is often presented. The genome is not 'The Book of Life'.

To judge from the reviews, *The Music of Life* has succeeded. It has worldwide sales exceeding 20,000 copies, and translations have been published in nine foreign languages, with more in preparation. The reviews are enthusiastic. Some of them are very perceptive and represent good literature in their own right. You can find many of them on the website www.musicoflife.website

This sourcebook responds to a growing demand by readers who are hungry for more. What they want is chapter and verse on the sources for the claims in *The Music of Life*. What you will find here is an extensive series of review articles written after *The Music of Life* was published that spell out in detail what the sources are, and with full references to the literature. Each and every claim in the book is not only justified by these articles, those claims are extended as new evidence has appeared since the book was published and which establishes some of the revolutionary ideas even more firmly than was possible in 2006.

Denis Noble, August 2016.

Contents

Noble, D. (2012). A Biological Theory of Relativity. No privileged level of causation. *Journal of the Royal Society Interface Focus* **2**, 55- 74.

This article provides a more complete statement of the Biological Theory of Relativity.

Noble, D. (2013). Physiology is rocking the foundations of evolutionary biology, *Experimental Physiology*, **98**, 1235-1243.

Based on lectures given to several international congresses between 2010 and 2013, this article follows on from the Neo- Darwinism article to explain why experimental findings are breaking the rules of the modern synthesis and proposes an extended synthesis in which physiological function may have a role in determining genetic changes.

Noble, D, Jablonka, E, Joyner, MJ, Muller, GB & Omholt, SW. (2014). Evolution evolves: physiology returns to centre stage. *Journal of Physiology*, **592**, 2237-2244.

This is an extensive Editorial for a Special Issue of the *Journal of Physiology* devoted to *The integration of evolutionary biology with physiological science*.

Noble, D. (2015) Evolution beyond neo-Darwinism: a new conceptual framework. *Journal of Experimental Biology*, **218**, 7-13.

This article analyses the conceptual basis of neo-Darwinism to reveal its inconsistencies and why it needs replacement by an integrative, multi-mechanism approach to evolutionary biology.

Noble, D. (2015) Conrad Waddington and the origin of epigenetics. *Journal of Experimental Biology*, **218**, 816-818.

Conrad Waddington invented the term epigenetics. He also performed experiments that demonstrated the assimilation of epigenetic changes into the genome, thus demonstrating the existence of the inheritance of acquired characteristics.

Experimental Physiology – *Paton Lecture*

Claude Bernard, the first systems biologist, and the future of physiology

Denis Noble

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

The first systems analysis of the functioning of an organism was Claude Bernard's concept of the constancy of the internal environment (*le milieu intérieur*), since it implied the existence of control processes to achieve this. He can be regarded, therefore, as the first systems biologist. The new vogue for systems biology today is an important development, since it is time to complement reductionist molecular biology by integrative approaches. Claude Bernard foresaw that this would require the application of mathematics to biology. This aspect of Claude Bernard's work has been neglected by physiologists, which is why we are not as ready to contribute to the development of systems biology as we should be. In this paper, I outline some general principles that could form the basis of systems biology as a truly multilevel approach from a physiologist's standpoint. We need the insights obtained from higher-level analysis in order to succeed even at the lower levels. The reason is that higher levels in biological systems impose boundary conditions on the lower levels. Without understanding those conditions and their effects, we will be seriously restricted in understanding the logic of living systems. The principles outlined are illustrated with examples from various aspects of physiology and biochemistry. Applying and developing these principles should form a major part of the future of physiology.

(Received 4 August 2007; accepted after revision 3 October 2007; first published online 26 October 2007)
Corresponding author D. Noble: Department of Physiology, Anatomy and Genetics, Parks Road, Oxford OX1 3PT, UK. denis.noble@dpag.ox.ac.uk

Historical introduction

Claude Bernard was Sir William Paton's great physiological hero. When the Physiological Society celebrated its centenary in 1976, Bill contributed a paper to the historical part of the meeting concerning one of Bernard's experiments on curare and drawing attention to the important role his ideas played in the foundation of the Society in 1876 (Paton, 1976). The reasons for his admiration of Claude Bernard are not hard to find. Bernard was a superb experimentalist, as the history of his work on digestion shows (Holmes, 1974). He also displayed his skills in many other areas of physiology and he laid out the principles of his science in his highly influential *Introduction à l'étude de la Médecine Expérimentale* (Bernard, 1865, 1984), in which he revealed himself to be a great thinker as well as a great experimentalist. The theoretical problem he addressed is one that is very relevant both to my claim that he was the first systems biologist and to the challenge that physiology faces today.

What was Claude Bernard's problem? It was that the chemists had created 'organic' molecules. This was a major development, since people had thought since Lémery's *Cours de Chymie* (published in 1675) that there were three completely separate classes of compounds: mineral, vegetable and animal. The first break in this idea came from the work of Lavoisier (1784), who showed that all compounds from vegetable and animal sources always contained at least carbon and hydrogen, and frequently nitrogen and phosphorus. This work bridged the vegetable–animal chemical boundary, but it left intact the boundary between the living and non-living. In fact, Berzelius (1815) even proposed that organic compounds were produced by laws different from inorganic compounds; the idea that there was a specific vital force that could not operate outside living systems. In 1828, however, Wöhler succeeded in creating urea from ammonium cyanate. The distinction between organic and non-organic origins was further weakened by Kolbe who, in 1845, synthesized acetic acid from its elements. Many

This article is based on the Paton Lecture delivered with the same title to the Life Sciences 2007 meeting in Glasgow in July 2007.

DOI: 10.1113/expphysiol.2007.038695

other discoveries of this kind (Finar, 1964) led to the idea that life itself could be reduced to chemistry and physics.

This was the challenge that physiologists such as Claude Bernard faced. His answer was precise. Neither vitalism nor chemical reductionism characterized living organisms. To the challenge that 'There are . . . chemists and physicists who . . . try to absorb physiology and reduce it to simple physico-chemical phenomena', Bernard responded, 'Organic individual compounds, though well defined in their properties, are still not active elements in physiological phenomena. They are only passive elements in the organism.' The reason, he explained, is that 'The living organism does not really exist in the *milieu extérieur* but in the liquid *milieu intérieur* . . .. a complex organism should be looked upon as an assemblage of simple organisms . . .. that live in the liquid *milieu intérieur*.'

His response to vitalism was equally robust: 'Many physicians . . .. assume a vital force in opposition to physico-chemical forces. I propose therefore to prove that the science of vital phenomena must have the same foundations as the science of the phenomena of inorganic bodies, and that there is no difference between the principles of biological science and those of physico-chemical science.'

By 'principles' here Bernard meant the laws governing the behaviour of the components. The control of the *milieu intérieur* meant not that the individual molecules did anything different from what they would do in non-living systems, but rather that the *ensemble* behaves in a controlled way, the controls being those that maintain the constancy of the internal environment. How could that be formalized? Could there be a theoretical physiology? Physical scientists had long since used mathematics to formalize their theories. Could that also be done in physiology? Bernard's answer to this question was 'yes, but not yet.' He cautioned, 'The most useful path for physiology and medicine to follow now is to seek to discover new facts instead of trying to reduce to equations the facts which science already possesses.' I believe that this view has been in part responsible for the broadly antitheoretical stance of British and American Physiology. It is important, therefore, to recognize that it represents only half of Bernard's views on the matter. For the emphasis in that statement should be on the word *now*. He also wrote that it was necessary to 'fix numerically the relations' between the components. He continued: 'This application of mathematics to natural phenomena is the aim of all science, because the expression of the laws of phenomena should always be mathematical.' His caution, therefore, was purely practical and temporal. In 1865 he saw, correctly of course, that physiology simply did not have enough data to make much mathematical application worthwhile *at that time*. But he clearly foresaw that the day would come when there would be sufficient data and that mathematical analysis would then become necessary.

The problem physiology faces today both resembles that faced by Bernard and differs from it. We face a new form of reductionism: that of genetic determinism, exemplified by the idea that there is a genetic program, what Jacob and Monod called '*le programme génétique*' (Monod & Jacob, 1961; Jacob, 1970). This challenge strongly resembles that of 'reducing life to physics and chemistry', the chemical being DNA. The major difference from Bernard's day is that we now have more facts than we can handle. There is a data explosion at all levels of biology. The situation is almost the reverse of that in Bernard's time. I have no doubt, therefore, that if he were alive today he would be championing his 'application of mathematics to natural phenomena.' I will illustrate why this is necessary and how it can be achieved by outlining some principles of systems biology from a physiologist's viewpoint. The principles are derived from my book on systems biology, *The Music of Life* (Noble, 2006), but their arrangement as a set of 10 was first presented by Noble (2007).

The principles of systems biology

First principle: biological functionality is multilevel. I start with this principle because it is obviously true, all the other principles can be shown to follow from it, and it is therefore the basis on which a physiological understanding of the phenomenon of life must be based. It is also a more general statement of the insight contained in Claude Bernard's idea of the constancy of the internal environment. That functionality is attributable to the organism as a whole and it controls all the other levels. This is the main reason why I describe Bernard as the first systems biologist. It is hard to think of a more important overall systems property than the one Bernard first identified.

Yet, the language of modern reductionist biology often seems to deny this obvious truth. The enticing metaphor of the 'book of life' made the genome into the modern equivalent of the 'embryo-homunculus', the old idea that each fertilized egg contains within it a complete organism in miniature (Mayr, 1982; p. 106). That the miniature is conceived as a digital 'map' or 'genetic program' does not avoid the error to which I am drawing attention, which is the idea that the living organism is simply the unfolding of an already-existing program, fine-tuned by its interaction with its environment, to be sure, but in all essentials, already there in principle as a kind of zipped-up organism. In its strongest form, this view of life leads to gene-selectionism and to gene-determinism: 'They [genes] created us body and mind' (Dawkins, 1976).

Dawkins himself does not really believe that. In a more recent book, he entitles one chapter 'Genes aren't us' (Dawkins, 2003) and, even in *The Selfish Gene*, the bold, simple message of the early chapters is qualified at the

end. My reservations, however, go much further than his. For, in truth, the stretches of DNA that we now call genes do nothing on their own. They are simply databases used by the organism as a whole. This is the reason for replacing the metaphor of the 'selfish' gene by genes as 'prisoners' (Noble, 2006; chapter 1). As Maynard Smith & Szathmáry (1999) express it, 'Co-ordinated replication prevents competition between genes within a compartment, and forces co-operation on them. They are all in the same boat.' From the viewpoint of the organism, genes as DNA molecules are therefore captured entities, no longer having a life of their own independent of the organism.

Second principle: transmission of information is not one way. The central dogma of molecular biology (Crick, 1970) is that information flows from DNA to RNA, from RNA to proteins, which can then form protein networks, and so on up through the biological levels to that of the whole organism. Information does not flow the other way. This is the dogma that is thought to safeguard modern neo-Darwinian theory from the spectre of 'Lamarckism', the inheritance of acquired characteristics. Applied to all the levels, this view is illustrated in Fig. 1. It encourages the bottom-up view of systems biology, the idea that if we knew enough about genes and proteins we could reconstruct all the other levels. Bioinformatics alone would be sufficient.

There are two respects in which the dogma is at least incomplete. The first is that it defines the relevant information uniquely in terms of the DNA code, the sequence of C, G, A, T bases. But the most that this information can tell us is *which* protein will be made. It does not tell us *how much* of each protein will be made. Yet, this is one of the most important characteristics of any living cell. Consider the speed of conduction of a nerve or muscle impulse, which depends on the density of rapidly activated sodium channels: the larger the density, the greater the ionic current and the faster the conduction. But this relationship applies only up to a certain optimum density, since the channel gating also contributes to the cell capacitance, which itself slows conduction, so there is a point beyond which adding more channel proteins is counter-productive (Hodgkin, 1975; Jack *et al.* 1975; p. 432). A feedback mechanism must therefore operate between the electrical properties of the nerve and the expression levels of the sodium channel protein. We now refer to such feedback mechanisms in the nervous system, which take many forms, as electro-transcription coupling (e.g. Deisseroth *et al.* 2003).

Similar processes must occur in the heart (e.g. Bers & Guo, 2005) and all the other organs. One of the lessons I have learnt from many attempts to model cardiac electrophysiology (Noble, 2002) is that, during the slow phases of repolarization and pacemaker activity, the ionic currents are so finely balanced that it is inconceivable that nature arrives at the correct expression and activity levels without some kind of feedback control. We don't yet know what that control might be, but we can say that it must exist. Nature cannot be as fragile as our computer models are! Robustness is an essential feature of successful biological systems.

There is nothing new in the idea that such feedback control of gene expression must exist. It is, after all, the basis of cell differentiation. All nucleated cells in the body contain exactly the same genome (with the exception of course of the germ cells, with only half the DNA). Yet the expression pattern of a cardiac cell is completely different from, say, a hepatic or bone cell. Moreover, whatever is determining those expression levels is accurately inherited during cell division. This cellular inheritance process is robust; it depends on some form of gene marking. It is this information on relative gene expression levels that is critical in determining each cell type.

By what principle could we possibly say that this is not relevant information? In the processes of differentiation and growth it is just as relevant as the raw DNA sequences. Yet, it is clear that this information *does* travel 'the other way'. The genes are told by the cells and tissues what to do, how frequently they should be transcribed and when to stop. There is 'downward causation' (Noble, 2006; chapter 4) from those higher levels that determines how the genome is 'played' in each cell (Fig. 2). Moreover, the possible number of combinations that could arise from so many gene components is so large (Feytmans *et al.* 2005) that there wouldn't be enough material in the whole universe for nature to have tried more than a small fraction

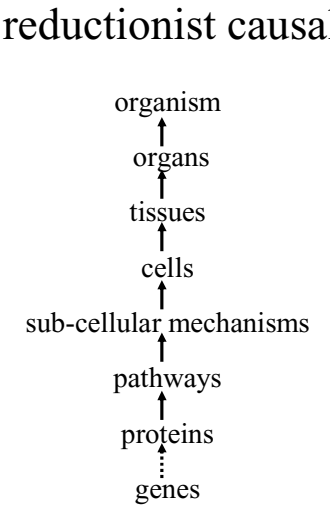


Figure 1. The reductionist 'bottom-up' causal chain (reproduced with permission from Noble, 2006)
This begins with the central dogma that information flows from DNA to proteins (bottom dotted arrow), never the other way, and extends the same concept through all the higher levels.

of the possible combinations even over the billions of years of evolution (Noble, 2006; chapter 2).

So the dogma is at least incomplete. But I also think it is incorrect in several important ways. Sure, protein sequences are not back-translated to form DNA sequences. In this limited original form, as formulated by Crick (1970), the central dogma is correct. But there is growing evidence from work on plants and microbes that environmental factors *do* change the genome, particularly by gene transfer (Goldenfeld & Woese, 2007). We cannot, therefore, use the original central dogma to exclude information transfer *into* the genome, determined by the organism and its environment.

Moreover, the DNA code itself is marked by the organism. This is the focus of the rapidly growing field of epigenetics (Qiu, 2006). At least two such mechanisms are now known at the molecular level: methylation of cytosine bases and control by interaction with the tails of histones around which the DNA is wound. Both of these processes modulate gene expression. The terminological question then arises: do we regard this as a form of code-modification? Is a cytosine, the C of the code, a kind of C* when it is methylated? That is a matter of definition of code, and one which I will deal with in the next section, but what is certain is that it is relevant information determining levels of gene expression, and that this information does flow against the direction of the central dogma. In fact, a form of inheritance of acquired characteristics (those of specific cell types) is rampant within all multicellular organisms with very different specialized cell types (Noble, 2006; chapter 7). At the least we have to say that, during the lifetime of the individual organism, transmission of information is far from being one way.

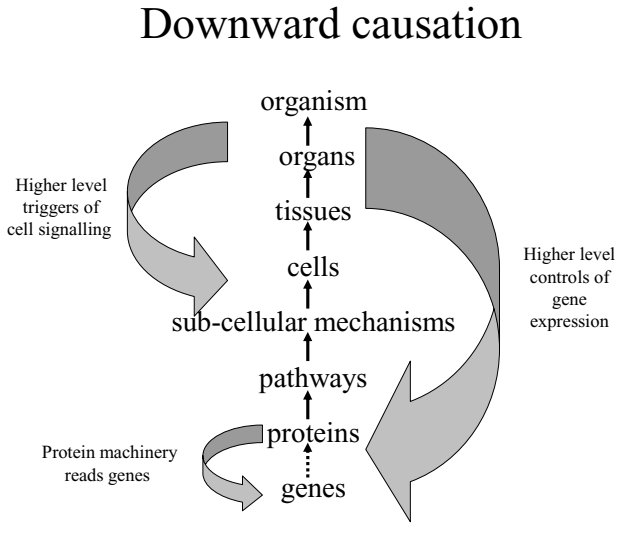


Figure 2. Figure 1 has been completed by adding the downward forms of causation, such as higher levels triggering cell signalling and gene expression
Note the downward-pointing arrow connecting from proteins to genes to indicate that it is protein machinery that reads and interprets gene coding. Loops of interacting downward and upward causation can be built between all levels of biological organization. Reproduced with permission from Noble (2006).

Third principle: DNA is not the sole transmitter of inheritance. The defenders of the original version of the central dogma would argue that, while my conclusions regarding the second principle are correct, what happens when information is transmitted to the next generation through the germ-line nevertheless involves wiping the slate clean of epigenetic effects. Methylation of cytosine bases and other forms of genome marking are removed. The genome is reset so that 'Lamarckism' is impossible.

But this is to put the matter the wrong way round. We need to explain *why* the genome (usually) reverts to an unmarked state. We don't explain that by appealing to the central dogma, for that dogma is simply a restatement of the same idea. We are in danger of circular logic here. Later, I will suggest a plausible reason why, at least most of the time, the resetting is complete, or nearly so. In order to do that, we first need to analyse the idea that genetics, as originally understood, is just about DNA.

This is not the original biological meaning of 'gene'. The concept of a gene has changed (Kitcher, 1982; Mayr, 1982; Dupré, 1993; Pichot, 1999). Its original biological meaning was an inheritable phenotype characteristic, such as eye/hair/skin colour, body shape and weight, number of legs/arms, to which we could perhaps add more complex traits like intelligence, personality, sexuality, etc. Genes, as originally conceived, are not just the same as stretches of DNA unless we subscribe to the view that the inheritance of all such characteristics is attributable entirely to DNA sequences. That is clearly false, since the egg cell is also inherited, together with any epigenetic characteristics transmitted by sperm (Anway *et al.* 2005), perhaps via RNA in addition to its DNA, and all the epigenetic influences of the mother and environment. Of course, the latter (environment) begins to be about 'nurture' rather than 'nature', but one of my points is that this distinction is fuzzy. The proteins that initiate gene transcription in the egg cell and impose an expression pattern on the genome are initially from the mother, and other such influences continue throughout development in the womb. Where we draw the line between nature and nurture is not at all obvious. There is an almost seamless transition from one to the other. 'Lamarckism', the inheritance of acquired characteristics, lurks in this fuzzy crack to a degree yet to be defined (Jablonka & Lamb, 1995, 2005). As the evolutionary geneticist Maynard Smith says, 'It [Lamarckism] is not so obviously false as is sometimes made out' (Maynard Smith, 1998).

Inheritance of the egg cell is important for two reasons. First, it is the egg cell DNA-reading machinery (a set of around 100 proteins and the associated cellular ribosome architecture) that enables the DNA to be used as a

template to make more proteins. Second, the set of other cellular elements, mitochondria, endoplasmic reticulum, microtubules, nuclear and other membranes, and a host of chemicals arranged specifically in cellular compartments, is also inherited. Most of this is not coded for by DNA sequences. Lipids certainly are not so coded. But they are absolutely essential to all the cell architecture. There would be no cells, nuclei, mitochondria, endoplasmic reticulum, ribosomes and all the other cellular machinery and compartments without the lipids. The specific details of all this cellular machinery matter. We can't make any old DNA do its thing in any old egg cell. Most attempts at interspecies cloning simply don't work. Invariably, a block occurs at an early stage in development. The only successful case so far is that of a wild ox (*Bos javanicus*) cloned in a domestic cow egg. The chances are that it will work only in very closely related species. The egg cell information is therefore also species specific.

Could epigenetic inheritance and its exclusion from the germ cell line be a requirement of multicellular harmony? The exact number of cell types in a human is debatable. It is partly a question of definition. A project that seeks to model all the cell types in the body, the Human Physiome Project (Crampin *et al.* 2004), estimates that there are around 200, all with completely different gene expression patterns. There would be even more if one took account of finer variations, such as those that occur in various regions of the heart and which are thought to protect the heart against fatal arrhythmias.

The precise number is not too important. The important fact is that it is large and that the range of patterns of gene expression is therefore also large and varied. Their patterns must also be harmonious in the context of the organism as a whole. They are all in the same boat; they sink or swim together. Disturbing their harmony would have serious consequences. It was arrived at after more than 2 billion years of experimentation.

Each cell type is so complex that the great majority of genes are expressed in many cell types. So it makes sense that all the cells in the body have the same gene complement, and that the coding for cell type is transmitted by gene marking, rather than by gene complement. I think that this gives the clue to the purpose of re-setting in germ-line inheritance. Consider what would happen if germ-line inheritance reflected adaptive changes in individual cell types. Given that all cell types derive ultimately from the fused germ-line cells, what would the effect be? Clearly, it would be to alter the patterns of expression in nearly all the cell types. There would be no way to transmit an improvement in, say, heart function to the next generation via gene marking of the germ cells without *also* influencing the gene expression patterns in many other types of cell in the body. And of course there is no guarantee that what is beneficial for a heart cell will be so in, say, a bone cell or a liver cell. On the contrary, the chances are that an adaptation beneficial in one cell type would be likely to be deleterious in another.

Much better, therefore, to let the genetic influences of natural selection be exerted on undifferentiated cells, leaving the process of differentiation to deal with the fine-tuning required to code for the pattern of gene expression appropriate to each type of cell. If this explanation is correct, we would not necessarily expect it to be 100% effective. It is conceivable that some germ-line changes in gene expression patterns might be so beneficial for the organism as a whole, despite deleterious effects on a few cell lines, that the result would favour selection. This could explain the few cases where germ-line 'Lamarckian' inheritance seems to have occurred. It also motivates the search for other cases. The prediction would be that it will occur in multicellular species only when beneficial to overall intercellular harmony. It might be more likely to occur in simpler species. That makes sense in terms of the few examples that we have so far found (Maynard Smith, 1998). Notice that, in contrast to the central dogma, this explanation is a systems level explanation.

Finally, in this section, I will comment on the concept of code. Applied to DNA, this is clearly metaphorical. It is also a useful metaphor, but we should beware of its limitations. One of these is to imply that only information that is coded is important, as in talk of the genome as the 'book of life'. The rest of cellular inheritance is not so coded; in fact, it is not even digital. The reason is very simple. The rest of the cellular machinery doesn't need to 'code for' or get 'translated into' anything else for the simple reason that it 'represents' itself; cells divide to form more cells, to form more cells, and so on. In this sense, germ-line cells are just as 'immortal' as DNA but a lot of this information is transmitted directly without having to be encoded. We should beware of thinking that only digitally 'coded' information is what matters in genetic inheritance.

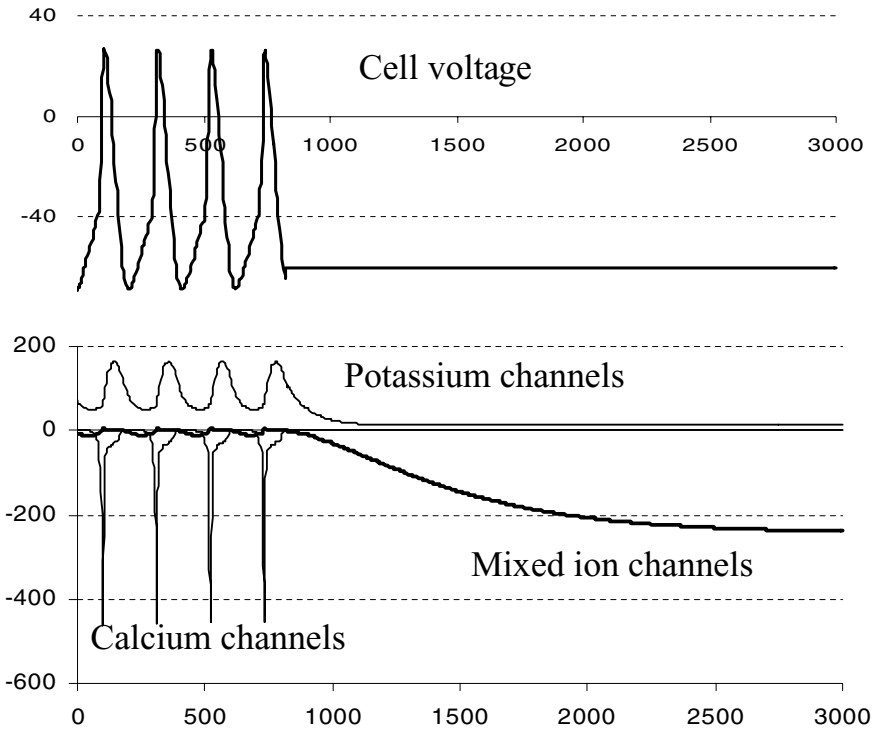
Fourth principle: the theory of biological relativity; there is no privileged level of causality. A fundamental property of systems involving multiple levels between which there are feedback control mechanisms is that there is no privileged level of causality. Consider, as an example, the cardiac pacemaker mechanism. This depends on ionic current generated by a number of protein channels carrying sodium, calcium, potassium and other ions. The activation, de-activation and inactivation of these channels proceed in a rhythmic fashion in synchrony with the pacemaker frequency. We might therefore be tempted to say that their oscillations generate that of the overall cell electrical potential, i.e. the higher-level functionality. But this is not the case. The kinetics of these channels varies with the electrical potential. There is therefore feedback between the higher-level property, the cell potential, and

the lower level property, the channel kinetics (Noble, 2006; chapter 5). This form of feedback was originally identified by Alan Hodgkin working on the nerve impulse, so it is sometimes called the Hodgkin cycle. If we remove the feedback, e.g. by holding the potential constant, as in a voltage clamp experiment, the channels no longer oscillate (Fig. 3). The oscillation is therefore a property of the system as a whole, not of the individual channels or even of a set of channels unless they are arranged in a particular way in the right kind of cell.

Nor can we establish any priority in causality by asking which comes first, the channel kinetics or the cell potential. This fact is also evident in the differential equations we use to model such a process. The physical laws represented in the equations themselves, and the initial and boundary conditions, operate *at the same time* (i.e. during every integration step, however infinitesimal), not sequentially.

It is simply a prejudice that inclines us to give some causal priority to lower-level, molecular events. The concept of level in biology is itself metaphorical. There is no literal sense in which genes and proteins lie *underneath* cells, tissues and organs. It is a convenient form of biological classification to refer to different levels, and we would find it very hard to do without the concept (Fig. 4). But we should not be fooled by the metaphor into thinking that 'high' and 'low' here have their normal meanings. From the metaphor itself, we can derive no justification for referring to one level of causality as privileged over others. That would be a misuse of the metaphor of level.

One of the aims of my book, *The Music of Life* (Noble, 2006), is to explore the limitations of biological metaphors. This is a form of linguistic analysis that is rarely applied in science, though a notable exception is Steven J. Gould's monumental work on the theory of evolution



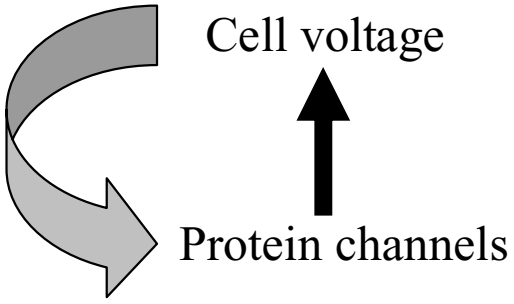


Figure 3. Computer model of pacemaker rhythm in the heart (reproduced with permission from Noble & Noble, 1984)
For the first four beats, the model is allowed to run normally and generates rhythm closely similar to a real heart. Then the feedback from cell voltage to protein channels is interrupted. All the protein channel oscillations then cease. They slowly change to steady, constant values. The diagram shows the causal loop involved. Protein channels carry current that changes the cell voltage (upward arrow), while the cell voltage changes the protein channels (downward arrow). In the simulation, this downward arrow was broken at 800 ms.

(Gould, 2002), in which he analyses the arguments for the multiplicity of levels at which natural selection operates.

These points can be generalized to any biological function. The only sense in which a particular level might be said to be privileged is that, in the case of each function, there is a level at which the function is integrated, and it is one of our jobs as biological scientists to determine what that level may be.

The idea that there is no privileged level of causality has a much wider range of applications than purely biological ones (Dupré, 1993; Cartwright, 1999; Keller, 2002), though the idea is rarely expressed in this bold, relativistic form. I use the word 'relativity' in formulating the principle because it shares certain features with theories of scale relativity proposed by some theoretical physicists, in particular the idea that there is no privileged scale, which is at the foundation of the theory of scale relativity (Nottale, 1993). There is an obvious correlation between scale and level, since lower and higher levels in any system operate at different scales. For this reason, some have proposed the application of the scale relativity theory framework and its associated mathematical tools to tackle the challenge of multiscale integration in systems biology (Nottale, 2000; Auffray & Nottale, 2008; Nottale & Auffray, 2008). But it is too early to judge whether this can provide a firm basis to a fully fledged theory of systems biology. Although the theory of scale relativity has already delivered a number of predictions in the realm of astrophysics which have been validated by subsequent observations, it still has to establish fully its position within theoretical physics. Nor is it possible yet to decide which principles are specific to systems biology and which are of general importance beyond the boundaries of biology.

Fifth principle: gene ontology will fail without higher-level insight. Genes, as defined by molecular genetics to be the coding regions of DNA, code for proteins. Biological function then arises as a consequence of multiple interactions between different proteins in the context of the rest of the cell machinery. Each function therefore depends on many genes, while many genes play roles in multiple functions. What then does it mean to give genes names in terms of functions? The only unambiguous labelling of genes is in terms of the proteins for which they code. Thus, the gene for the sodium–calcium exchange protein is usually referred to as *ncx*. Ion channel genes are also often labelled in this way, as in the case of sodium channel genes being labelled *scn*.

This approach, however, naturally appears unsatisfactory from the viewpoint of a geneticist, since the original question in genetics was not which proteins are coded for by which stretches of DNA [in fact, early ideas on where the genetic information might be found (Schrödinger, 1944) favoured the proteins], but rather what is responsible for higher-level phenotype characteristics. There is no one-to-one correspondence between genes or proteins and higher-level biological functions. Thus, there is no 'pacemaker' gene. Cardiac rhythm depends on many proteins interacting within the context of feedback from the cell electrical potential.

Let's do a thought experiment. Suppose we could knock out the gene responsible for L-type calcium channels and still have a living organism (perhaps because a secondary pacemaker takes over and keeps the organism viable – and something else would have to kick-in to enable excitation–contraction coupling, and so on throughout the body because L-type calcium channels are ubiquitous!). Since

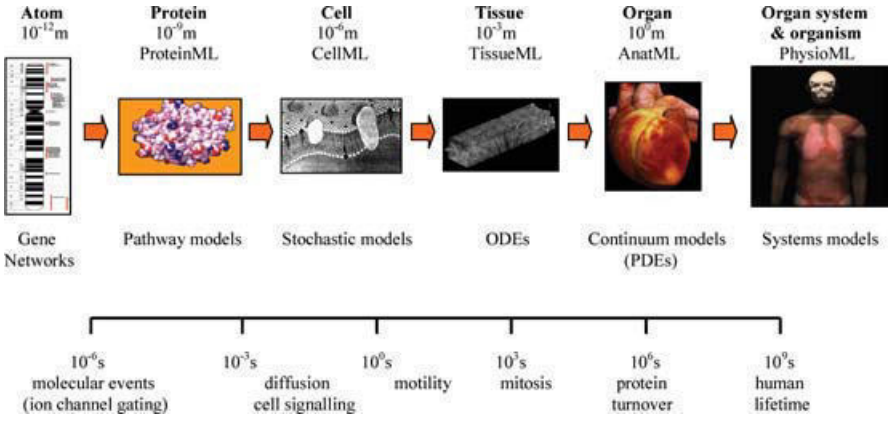


Figure 4. Spatial (top) and temporal (bottom) scales encompassed by the Human Physiome Project
The types of mathematical model appropriate to each spatial scale are also indicated. The last two images on the right in this figure, and all subsequent anatomical images, are from anatomically based models developed by the Auckland Bioengineering group. The tissue image is a three-dimensional confocal microscopy reconstruction of a transmural segment of rat heart by the Auckland group led by Peter Hunter (Hunter *et al.* 2002). Abbreviations: ML, markup language; ODE, ordinary differential equations; PDE, partial differential equations. Reproduced with Permission from Hunter *et al.* (2002).

L-type calcium current is necessary for the upstroke of the action potential in the SA node of most species, we would find that we had abolished normal pacemaker rhythm. Do we then call the gene for L-type calcium channels the 'pacemaker' gene? The reason why this is unsatisfactory, even misleading, to a systems-level biologist is obvious. Yet it is the process by which we label many genes with high-level functions. The steadily growing list of 'cancer genes' have been identified in this way, by determining which mutations (including deletions) change the probability of cancer occurring. We can be fairly sure though that this characteristic is not why they were selected during the evolutionary process. In this sense, there are no 'cancer genes'. As the Gene Ontology (GO) Consortium (http://geneontology.org/) puts it, 'oncogenesis is not a valid GO term because causing cancer is not the normal function of any gene'.

Another good example of this approach is the discovery of what are called clock genes, involved in circadian rhythm. Mutations in a single gene (now called the *period* gene) are sufficient to abolish the circadian period of fruit flies (Konopka & Benzer, 1971). This discovery of the first 'clock gene' was a landmark, since it was the first time that a single gene had been identified as playing such a key role in a high-level biological rhythm. The expression levels of this gene are clearly part of the rhythm generator. They vary (in a daily cycle) in advance of the variations in the protein for which they code. The reason is that the protein is involved in a negative feedback loop with the gene that codes for it (Hardin *et al.* 1990). The idea is very simple. The protein levels build up in the cell as the *period* gene is read to produce more protein. The protein then diffuses into the nucleus, where it inhibits further production of itself by binding to the promoter part of the gene sequence. With a time delay, the protein production falls off and the inhibition is removed so that the whole cycle can start again. So, we not only have a single gene capable of regulating the biological clockwork that generates circadian rhythm, it is itself a key component in the feedback loop that forms the rhythm generator.

However, such rhythmic mechanisms do not work in isolation. There has to be some connection with light-sensitive receptors (including the eyes). Only then will the mechanism lock on to a proper 24 h cycle rather than free-running at say 23 or 25 h. In the mouse, for example, many other factors play a role. Moreover, the clock gene itself is involved in other functions. That is why Foster and Kreitzman have written 'What we call a clock gene may have an important function within the system, but it could be involved in other systems as well. Without a complete picture of all the components and their interactions, it is impossible to tell what is part of an oscillator generating rhythmicity, what is part of an input, and what is part of an output. In a phrase, it ain't that simple!' (Foster & Kreitzman, 2004).

Indeed not. The *period* gene has also been found to be implicated in embryonic development as the adult fly is formed over several days, and it is deeply involved in coding for the male love songs generated by wing-beat oscillations which are specific to each of around 5000 species of fruit fly and ensure that courtship is with the right species. Perhaps it should be renamed the 'fruit fly love gene'!

The point is obvious. We should not be misled by gene ontology. The first function a gene is found to be involved in is rarely, if ever, the only one and may not even be the most important one. Gene ontology will require higher-level insight to be successful in its mission. Moreover, current methods of relating genotype to phenotype suffer from a major methodological limitation: by determining the effects of *changes* (mutations) in the genome, we can say little *a priori* on the direct causal relations between wild-type genes and the phenotype. They reveal simply the *differences* produced as a result of the *change* in genotype. All the causal effects *common* to both the wild-type and the mutated gene are hidden. What is observed may be just the tip of the iceberg.

Gene ontology in its fullest sense, as originally conceived by geneticists to relate genes to high-level features, is therefore very difficult and subject to many traps for the unwary. This would explain why projects such as the GO Consortium are more limited in their scope. Thus, GO assigns three categories to a gene, namely molecular function, biological process and cellular component, which are not intended to deal with higher-level function. It specifically excludes protein domains or structural features, protein–protein interactions, anatomical or histological features above the level of cellular components, including cell types, and it excludes the environment, evolution and expression. In other words, it excludes virtually all of what we classically understand by physiology and most aspects of evolutionary biology.

Sixth principle: there is no genetic program. No genetic programs? Surely, they are all over the place! They are the crown jewels of the molecular genetic revolution, invented by none other than the famous French Nobel Prize winners, Monod and Jacob (Monod & Jacob, 1961; Jacob, 1970). Their enticing idea was born during the early days of electronic computing, when computers were fed with paper tape or punched cards coded with sequences of instructions. Those instructions were clearly separate from the machine itself that performed the operations. They dictated those operations. Moreover, the coding is digital. The analogy with the digital code of DNA is obvious. So, are the DNA sequences comparable to the instructions of a computer program?

An important feature of such computer programs is that the program is separate from the activities of the machine that it controls. Originally, the separation was

physically complete, with the program on the tape or cards only loaded temporarily into the machine. Nowadays, the programs are stored within the memory of the machine, and the strict distinction between the program, the data and the processes controlled may be breaking down. Perhaps computers are becoming more like living systems, but in any case the concept of a genetic program was born in the days when programs were separate, identifiable sets of instructions.

So, what do we find when we look for genetic programs in an organism? We find no genetic programs! There are no sequences of instructions in the genome that could possibly play a role similar to that of a computer program. The reason is very simple. A database, used by the system as a whole, is not a program. To find anything comparable to a program we have to extend our search well beyond the genome itself. Thus, as we have seen above, the sequence of events that generates circadian rhythm includes the *period* gene, but it necessarily also includes the protein for which it codes, the cell in which its concentration changes and the nuclear membrane across which it is transported with the correct speed to effect its inhibition of transcription. This is a gene–protein–lipid–cell network, not simply a gene network. The nomenclature matters. Calling it a gene network fuels the misconception of genetic determinism. In the generation of a 24 h rhythm, none of these events in the feedback loop is privileged over any other. Remove any of them, not just the gene, and you no longer have circadian rhythm.

Moreover, it would be strange to call this network of interactions a program. The network of interactions *is itself the circadian rhythm process*. As Enrico Coen, the distinguished plant geneticist, put it, 'Organisms are not simply manufactured according to a set of instructions. There is no easy way to separate instructions from the process of carrying them out, to distinguish plan from execution' (Coen, 1999). In short, the concept of a program here is completely redundant. It adds nothing to what a systems approach to such processes can reveal.

Seventh principle: there are no programs at any other level. I have introduced the analogy of the genome as a database and the metaphor of 'genes as prisoners' in order to provoke the change in mindset that is necessary for a fully systems approach to biology to be appreciated. The higher levels of the organism 'use the database' and 'play the genome' to produce functionality. If the genome can be likened to a huge pipe organ (Noble, 2006; chapter 2), then it seems correct to ask who is the player, who was the composer? If we can't find the program of life at the level of the genome, at what level do we find it? The answer is 'nowhere'!

We should view all such metaphors simply as ladders of understanding. Once we have used them we can, as it were, throw them away. This way of thinking can seem strange to some scientists for whom there must be just one correct answer to any scientific question. I explore this important issue in *The Music of Life* by analysing the 'selfish gene' and 'prisoner gene' metaphors linguistically to reveal that no conceivable experiment could decide which is correct (Noble, 2006; chapter 1). They highlight totally different aspects of the properties of genes. This philosophy is applied throughout the book as it answers questions like 'where is the program of life?' The conclusion is simply that there are no such programs at any level. At all levels, the concept of a program is redundant since, as with the circadian rhythm network, the networks of events that might be interpreted as programs are themselves the functions we are seeking to understand. Thus, there is no program for the heart's pacemaker separate from the pacemaker network itself.

While causality operates within and between all levels of biological systems, there are certain levels at which so many functions are integrated that we can refer to them as important levels of abstraction. Sydney Brenner wrote, 'I believe very strongly that the fundamental unit, the correct level of abstraction, is the cell and not the genome' (unpublished Lecture, Columbia University, 2003). He is correct, since the development of the eukaryotic cell was a fundamental stage in evolutionary development, doubtless requiring at least a billion years to be achieved. To systems physiologists though there are other important levels of abstraction, including whole organs and systems.

Eighth principle: there are no programs in the brain. In his book *The Astonishing Hypothesis*, Francis Crick proclaimed, 'You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behaviour of a vast assembly of nerve cells and their associated molecules' (Crick, 1994). This is a variation of the idea that in some sense or other, the mind is just a function of the brain. The pancreas secretes insulin, endocrine glands secrete hormones ... and the brain 'secretes' consciousness! All that's left is to find out how and where in the brain that happens. In one of his last statements, Crick has even hinted at where that may be: 'I think the secret of consciousness lies in the claustrum' (Francis Crick, 2004, quoted by V. S. Ramachanran, in *The Astonishing Francis Crick*, Edge, 18 October, 2004, http://www.edge.org/3rd_culture/crick04/crick04_index.html). This structure is a thin layer of nerve cells in the brain. It is very small and it has many connections to other parts of the brain, but the details are of no importance to the argument. The choice of brain location for the 'secret of consciousness' varies greatly according to the author. Descartes even thought that it was in the pineal gland. The mistake is always the same, which is to think that in some way or other the brain is a kind of performance space in which the world of perceptions is reconstructed

inside our heads and presented to us as a kind of Cartesian theatre. But that way of looking at the brain leaves open the question: where is the 'I', the conscious self that sees these reconstructions? Must that be another part of the brain that views these representations of the outside world?

We are faced here with a mistake similar to that of imagining that there must be programs in the genomes, cells, tissues and organs of the body. There are no such programs, even in the brain. The activity of the brain and of the rest of the body simply *is* the activity of the person, the self. Once again, the concept of a program is superfluous. When a guitarist plays the strings of his guitar at an automatic speed that comes from frequent practice, there is no separate program that is making him carry out this activity. The patterns and processes in his nervous system and the associated activities of the rest of his body simply *are* him playing the guitar. Similarly, when we deliberate intentionally, there is no nervous network 'forcing' us to a particular deliberation. The nervous networks, the chemistry of our bodies, together with all their interactions within the social context in which any intentional deliberation makes sense, *are* us acting intentionally. Looking for something in addition to those processes is a mistake.

Ninth principle: the self is not an object. In brief, the mind is not a separate object competing for activity and influence with the molecules of the body. Thinking in that way was originally the mistake of the dualists, such as Sherrington and Eccles, led by the philosophy of Descartes. Modern biologists have abandoned the separate substance idea, but many still cling to a materialist version of the same mistake (Bennett & Hacker, 2003), based on the idea that somewhere in the brain the self is to be found as some neuronal process. The reason why that level of integration is too low is that the brain, and the rest of our bodies which are essential for attributes such as consciousness to make sense (Noble, 2006; chapter 9), are tools (back to the database idea again) in an integrative process that occurs at a higher level involving social interactions. We cannot attribute the concept of self-ness to ourselves without also doing so to others (Strawson, 1959). Contrary to Crick's view, therefore, our selves are indeed much 'more than the behaviour of a vast assembly of nerve cells and their associated molecules' precisely because the social interactions are essential even to understanding what something like an intention might be. I analyse an example of this point in much more detail in chapter 9 of *The Music of Life*. This philosophical point is easier to understand when we take a systems view of biology, since it is in many ways an extension of that view to the highest level of integration in the organism.

Conclusions

Tenth principle: there are many more to be discovered; a genuine 'theory of biology' does not yet exist. Well, of course, choosing just 10 principles was too limiting. This last one points the way to many others of whose existence we have only vague ideas. We do not yet have a genuine theory of biology. The Theory of Evolution is not a theory in the sense in which I am using the term. It is more an historical account, itself standing in need of explanation. We don't even know yet whether it consists of events that are difficult, if not impossible, to analyse fully from a scientific perspective, or whether it was a process that would have homed in to the organisms we have, regardless of the conditions. My own suspicion is that it is most unlikely that, if we could turn the clock right back and let the process run again, we would end up with anything like the range of species we have today on earth (Gould, 2002).

But, whichever side of this particular debate you may prefer, the search for general principles that could form the basis of a genuine theory of biology is an important aim of systems biology. Can we identify the logic by which the organisms we find today have succeeded in the competition for survival? In searching for that logic, we should not restrict ourselves to the lower levels. Much of the logic of living systems is to be found at the higher levels, since these are often the levels at which selection has operated (Keller, 1999; Gould, 2002) and determined whether organisms live or die. This is the level at which physiology works. Physiology therefore has a major contribution to make to systems biology.

In conclusion, I return to the theme with which this article began. Claude Bernard's concept of the constancy of the internal environment was the first example of multilevel functionality. It was critical in defining physiology as a subject distinct from the applications of physics and chemistry. The challenge we face today resembles that faced by Bernard in the mid-nineteenth century, but the chemistry involved is that of the molecule DNA. The answer though should be much the same. Higher-level control cannot be reduced to lower-level databases like the genome. A major part of the future of physiology surely lies in returning to our roots. Higher-level systems biology is, I suggest, classical physiology by another name.

References

Anway MD, Cupp AS, Uzumcu M & Skinner MK (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science* **308**, 1466–1469.

Auffray C & Nottale L (2008). Scale relativity theory and integrative systems biology 1. Founding principles and scale laws. *Prog Biophys Mol Biol*, in press.

Bennett MR & Hacker PMS (2003). *Philosophical Foundations of Neuroscience*. Blackwell Publishing, Oxford.
Bernard C (1865, 1984). *Introduction a L'étude de la Médecine Expérimentale*. Flammarion, Paris.
Bers DM & Guo T (2005). Calcium signaling in cardiac ventricular myocytes. *Ann New York Acad Sci* **1047**, 86–98.
Berzelius (1815). *Afhandlingar I Fysik, Kemi och Mineralogi*, Stockholm: **4**, 307.
Cartwright N (1999). *The Dappled World. A Study of the Boundaries of Science*. Cambridge University Press, Cambridge.
Coen E (1999). *The Art of Genes*. Oxford University Press, Oxford.
Crampin EJ, Halstead M, Hunter PJ, Nielsen P, Noble D, Smith N & Tawhai M (2004). Computational physiology and the physiome project. *Exp Physiol* **89**, 1–26.
Crick FHC (1970). Central dogma of molecular biology. *Nature* **227**, 561–563.
Crick FHC (1994). *The Astonishing Hypothesis: the Scientific Search for the Soul*. Simon and Schuster, London.
Dawkins R (1976). *The Selfish Gene*. Oxford University Press, Oxford.
Dawkins R (2003). *A Devil's Chaplain*. Weidenfeld and Nicolson, London.
Deisseroth K, Mermelstein PG, Xia H & Tsien RW (2003). Signaling from synapse to nucleus: the logic behind the mechanisms. *Curr Opin Neurobiol* **13**, 354–365.
Dupré J (1993). *The Disorder of Things*. Harvard, Cambridge, MA, USA.
Feytmans E, Noble D & Peitsch M (2005). Genome size and numbers of biological functions. *Trans Comput Systems Biol* **1**, 44–49.
Finar IL (1964). *Organic Chemistry*. Longmans, London.
Foster R & Kreitzman L (2004). *Rhythms of Life*. Profile Books, London.
Frankland E & Kolbe H (1845). Upon the chemical constitution of metacetonic acid, and some other bodies related to it. *Mem. Proc. Chem. Soc.* **1865**, 386–391.
Goldenfeld N & Woese C (2007). Biology's next revolution. *Nature* **445**, 369.
Gould SJ (2002). *The Structure of Evolutionary Theory*. Harvard, Cambridge, MA, USA.
Hardin PE, Hall JC & Rosbash M (1990). Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* **343**, 536–540.
Hodgkin AL (1975). The optimum density of sodium channels in an unmyelinated nerve. *Proc Royal Soc Lond B Biol Sci* **270**, 297–300.
Holmes FL (1974). *Claude Bernard and Animal Chemistry. The Emergence of a Scientist*. Harvard, Cambridge, MA, USA.
Hunter PJ, Robbins P & Noble D (2002). The IUPS human physiome project. *Pflugers Arch* **445**, 1–9.
Jablonka E & Lamb M (1995). *Epigenetic Inheritance and Evolution. The Lamarckian Dimension*. Oxford University Press, Oxford.
Jablonka E & Lamb M (2005). *Evolution in Four Dimensions*. MIT Press, Boston, MA, USA.
Jack JJB, Noble D & Tsien RW (1975). *Electric Current Flow in Excitable Cells*. Oxford University Press, Oxford.
Jacob F (1970). *La Logique Du Vivant, une Histoire de L'hérédité*. Gallimard, Paris.
Keller EF (2002). *Making Sense of Life. Explaining Biological Development with Models, Metaphors and Machines*. Harvard, Cambridge, MA, USA.
Keller L (1999). *Levels of Selection in Evolution*. Princeton University Press, Princeton, NJ, USA.
Kitcher P (1982). Genes. *Br J Philosophy Sci* **33**, 337–359.
Konopka RJ & Benzer S (1971). Clock mutants of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* **68**, 2112–2116.
Lémery N (1675). *Cours de Chymie*. Paris: Michallet.
Lavoisier A (1784). *Traité élémentaire de chimie, présenté dans un ordre nouveau et d'après les découvertes modernes*, 2 vols. Paris: Chez Cuchet.
Maynard Smith J (1998). *Evolutionary Genetics*. Oxford University Press, New York.
Maynard Smith J & Szathmáry E (1999). *The Origins of Life*. Oxford University Press, New York.
Mayr E (1982). *The Growth of Biological Thought*. Harvard, Cambridge, MA, USA.
Monod J & Jacob F (1961). Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb Symp Quant Biol* **26**, 389–401.
Noble D (2002). Modelling the heart: insights, failures and progress. *Bioessays* **24**, 1155–1163.
Noble D (2006). *The Music of Life*. Oxford University Press, Oxford.
Noble D (2007). Mind over molecule: activating biological demons. *Annals N Y Acad Sci*, in press.
Noble D & Noble SJ (1984). A model of sino-atrial node electrical activity using a modification of the DiFrancesco-Noble (1984) equations. *Proc Royal Soc Lond B Biol Sci* **222**, 295–304.
Nottale L (1993). *Fractal Space-Time and Microphysics: Towards a Theory of Scale Relativity*. World Scientific, Singapore.
Nottale L (2000). *La Relativité Dans Tous Ses Etats. Du Mouvements Aux Changements D'échelle*. Hachette, Paris.
Nottale L & Auffray C (2008). Scale relativity and integrative systems biology 2. Macroscopic quantum-type mechanics. *Prog Biophys Mol Biol*, in press.
Paton WDM (1976). An experiment of Claude Bernard on curare: the origins of the Physiological Society. *J Physiol* **263**, 26*P*–29*P*.
Pichot A (1999). *Histoire de la Notion de Gène*. Flammarion, Paris.
Qiu J (2006). Unfinished symphony. *Nature* **441**, 143–145.
Schrödinger E (1944). *What Is Life?* Cambridge University Press, Cambridge, UK.
Strawson PF (1959). *Individuals*. Routledge, London.
Wöhler F (1828). Ueber künstliche Bildung des Harnstoffs. *Ann. Chim. Phys.* **37**, 330.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2008) **366**, 3001–3015
doi:10.1098/rsta.2008.0086
Published online 17 June 2008

Review

Genes and causation

By Denis Noble*

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

Relating genotypes to phenotypes is problematic not only owing to the extreme complexity of the interactions between genes, proteins and high-level physiological functions but also because the paradigms for genetic causality in biological systems are seriously confused. This paper examines some of the misconceptions, starting with the changing definitions of a gene, from the cause of phenotype characters to the stretches of DNA. I then assess whether the 'digital' nature of DNA sequences guarantees primacy in causation compared to non-DNA inheritance, whether it is meaningful or useful to refer to genetic programs, and the role of high-level (downward) causation. The metaphors that served us well during the molecular biological phase of recent decades have limited or even misleading impacts in the multilevel world of systems biology. New paradigms are needed if we are to succeed in unravelling multifactorial genetic causation at higher levels of physiological function and so to explain the phenomena that genetics was originally about. Because it can solve the 'genetic differential effect problem', modelling of biological function has an essential role to play in unravelling genetic causation.

Keywords: genes; genetic causation; genetic program; digital coding; analogue representation; cell inheritance

1. Introduction: what is a gene?

At first sight, the question raised by this paper seems simple. Genes transmit inherited characteristics; so in each individual they must be the cause of those characteristics. And so it was when the idea of a gene was first mooted. The word itself was coined by Johannsen (1909), but the concept already existed and was based on 'the silent assumption [that] was made almost universally that there is a 1:1 relation between genetic factor (gene) and character' (Mayr 1982).

Since then, the concept of a gene has changed fundamentally (Kitcher 1982; Mayr 1982; Dupré 1993; Pichot 1999; Keller 2000*a*,*b*), and this is a major source of confusion when it comes to the question of causation. Its original biological meaning referred to the cause of an inheritable phenotype characteristic, such as

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

One contribution of 12 to a Theme Issue 'The virtual physiological human: building a framework for computational biomedicine I'.

eye/hair/skin colour, body shape and weight, number of legs/arms/wings, to which we could perhaps add more complex traits such as intelligence, personality and sexuality.

The molecular biological definition of a gene is very different. Following the discovery that DNA codes for proteins, the definition shifted to locatable regions of DNA sequences with identifiable beginnings and endings. Complexity was added through the discovery of regulatory elements, but the basic cause of phenotype characteristics was still the DNA sequence since that determined which protein was made, which in turn interacted with the rest of the organism to produce the phenotype.

But unless we subscribe to the view that the inheritance of all phenotype characteristics is attributable entirely to DNA sequences (which I will show is just false) then genes, as originally conceived, are not the same as the stretches of DNA. According to the original view, genes were necessarily the cause of inheritable phenotypes since that is how they were defined. The issue of causation is now open precisely because the modern definition identifies them instead with DNA sequences.

This is not a point that is restricted to the vexed question of the balance of nature versus nurture. Even if we could separate those out and arrive at percentages attributable to one or the other (which I believe is misconceived in a system of nonlinear interactions and in which either on its own is equal to zero), we would still be faced with the fact that not all the 'nature' characteristics are attributable to DNA alone. Indeed, as we will see as we come to the conclusion of this paper, strictly speaking no genetic characteristics as originally defined by geneticists in terms of the phenotype could possibly be attributable to DNA alone.

My first point therefore is that the original concept of a gene has been taken over and significantly changed by molecular biology. This has undoubtedly led to a great clarification of *molecular* mechanisms, surely one of the greatest triumphs of twentieth-century biology, and widely acknowledged as such. But the more philosophical consequences of this change for higher level biology are profound and they are much less widely understood. They include the question of causation by genes. This is also what leads us to questions such as 'how many genes are there in the human genome?', and to the search to identify 'genes' in the DNA sequences.

2. Where does the genetic code lie?

Of course, it is an important question to ask which stretches of DNA code for proteins, and that is a perfectly good *molecular biological* question. It also leads us to wonder what the other stretches of DNA are used for, a question to which we are now beginning to find answers (Pearson 2006). But genetics, as originally conceived, is not just about what codes for each protein. Indeed, had it turned out (as in very simple organisms) that each coding stretch of DNA translates into just one protein, then it would have been as valid to say that the genetic code lies in the protein sequences, as was originally thought (Schrödinger 1944). We are then still left with the question 'how do these sequences, whether DNA or protein, generate the phenotypic characteristics that we wish to explain?' Looked at from this viewpoint, modern molecular biology, starting with Watson and

Crick's work, has succeeded brilliantly in mapping sequences of DNA to those of amino acids in proteins, but not in explaining phenotype inheritance. Whether we start from DNA or protein sequences, the question is still there. It lies in the complexity of the way in which the DNA and proteins are used by the organism to generate the phenotype. Life is not a soup of proteins.

The existence of multiple splice variants and genetic 'dark matter' (only 1–2% of the human genome actually codes for proteins, but much of the rest codes for non-protein coding RNA; Bickel & Morris 2006; Pearson 2006) has made this question more complicated in higher organisms, while epigenetics (gene marking) makes it even more so (Qiu 2006; Bird 2007), but the fundamental point remains true even for higher organisms. In a more complicated way, the 'code' could still be seen to reside in the proteins. Some (e.g. Scherrer & Jost 2007) have even suggested that we should redefine genes to be the completed mRNA before translation into a polypeptide sequence (see also Noble 2008, in press). In that case, there would be as many as 500 000 genes rather than 25 000. The more complex genome structure (of multiple exons and introns and the way in which the DNA is folded in chromosomes) could then be viewed as an efficient way of preserving and transmitting the 'real' causes of biological activity, the proteins. It is still true that, if we identify genes as just the stretches of DNA and identify them by the proteins they code for, we are already failing to address the important issues in relation to genetic determinism of the phenotype. By accepting the molecular biological redefinition of 'gene', we foreclose some of the questions I want to ask. For, having redefined what we mean by a gene, many people have automatically taken over the concept of necessary causation that was correctly associated with the original idea of a gene, but which I will argue is incorrectly associated with the new definition, except in the limited case of generating proteins from DNA. This redefinition is not therefore just an arcane matter of scientific history. It is part of the mindset that needs to change if we are to understand the full nature of the challenge we face.

3. Digital versus analogue genetic determinism

The main reason why it is just false to say that all nature characteristics are attributable to DNA sequences is that, by itself, DNA does nothing at all. We also inherit the complete egg cell, together with any epigenetic characteristics transmitted by sperm (in addition to its DNA), and all the epigenetic influences of the mother and environment. Of course, the latter begins to be about 'nurture' rather than nature, but one of my points in this paper is that this distinction is fuzzy. The proteins that initiate gene transcription in the egg cell and impose an expression pattern on the genome are initially from the mother, and other such influences continue throughout development in the womb and have influences well into later life (Gluckman & Hanson 2004). Where we draw the line between nature and nurture is not at all obvious. There is an almost seamless transition from one to the other. 'Lamarckism', the inheritance of acquired characteristics, lurks in this fuzzy crack to a degree yet to be defined (Jablonka & Lamb 1995, 2005).

This inheritance of the egg cell machinery is important for two reasons. First, it is the egg cell gene reading machinery (a set of approx. 100 proteins and the associated cellular ribosome architecture) that enables the DNA to be used to

make more proteins. Second, the complete set of other cellular elements, mitochondria, endoplasmic reticulum, microtubules, nuclear and other membranes and a host (billions) of chemicals arranged specifically in cellular compartments, is also inherited. Much of this is not coded for by DNA sequences since they code only for RNA and proteins. Lipids certainly are not so coded. But they are absolutely essential to all the cell architecture. The nature of the lipids also determines how proteins behave. There is intricate two-way interaction between proteins and lipids (see Roux *et al.* 2008).

One way to look at this situation therefore is to say that there are two components to molecular inheritance: the genome DNA, which can be viewed as digital information, and the cellular machinery, which can, perhaps by contrast, be viewed as analogue information. I will refer to both of these as 'molecular inheritance' to emphasize that the distinction at this point in my argument is not between genetic molecular inheritance and higher-level causes. The egg cell machinery is just as molecular as the DNA. We will come to higher-level causation later.

The difference lies elsewhere. Both are used to enable the organism to capture and build the new molecules that enable it to develop, but the process involves a coding step in the case of DNA and proteins, while no such step is involved in the rest of the molecular inheritance. *This is the essential difference.*

The coding step in the case of the relationship between DNA and proteins is what leads us to regard the information as digital. This is what enables us to give a precise number to the base pairs (3 billion in the case of the human genome). Moreover, the CGAT code could be completely represented by binary code of the kind we use in computers. (Note that the code here is metaphorical in a biological context—no one has determined that this should be a code in the usual sense. For that reason, some people have suggested that the word 'cipher' would be better.)

By contrast, we cannot put similar precise numbers to the information content of the rest of the molecular inheritance. The numbers of molecules involved (trillions) would be largely irrelevant since many are exactly the same, though their organization and compartmentalization also need to be represented. We could therefore ask how much digital information would be required to 'represent' the non-DNA inheritance but, as with encoding of images, that depends on the resolution with which we seek to represent the information digitally. So, there is no simple answer to the question of a quantitative comparison of the DNA and non-DNA molecular inheritance. But given the sheer complexity of the egg cell—it took evolution at least 1 or 2 billion years to get to the eukaryotic cellular stage—we can say that it must be false to regard the genome as a 'vast' database while regarding the rest of the cell as somehow 'small' by comparison. At fine enough resolution, the egg cell must contain even more information than the genome. If it needed to be coded digitally to enable us to 'store' all the information necessary to recreate life in, say, some distant extra-solar system by sending it out in an 'Earth-life' information capsule, I strongly suspect that most of that information would be non-genomic. In fact, it would be almost useless to send just DNA information in such a capsule. The chances of any recipients anywhere in the Universe having egg cells and a womb capable of permitting the DNA of life on Earth to 'come alive' may be close to zero. We might as well pack the capsule with the bar codes of a supermarket shelf!

4. Is digital information privileged?

Of course, quantity of information is not the only criterion we could choose. Whatever its proportion would be in my imagined Earth-life capsule, some information may be more important than others. So, which is privileged in inheritance? Would it be the cell or the DNA? 'How central is the genome?' as Werner puts the question (Werner 2007). On the basis of our present scientific knowledge, there are several ways in which many people would seek to give primacy to the DNA.

The first is the fact that, since it can be viewed as digital information, in our computer-oriented age, that can appear to give it more security, to ensure that it is more reliable, much as the music recorded on a CD is said to be 'clearer' and less 'noisy' than that on a vinyl disc. Digital information is discrete and fixed, whereas analogue information is fuzzy and imprecise. But I wonder whether that is entirely correct. Large genomes actually require correcting machinery to ensure their preciseness. Nevertheless, with such machinery, it clearly is secure enough to act as reliably inheritable material. By contrast, it could be said that attempting to reduce analogue information, such as image data, to digital form is always fuzzy since it involves a compromise over questions such as resolution. But this criterion already biases us towards the DNA. We need to ask the fundamental question 'why do we need to prioritize digital information?' After all, DNA needs a digital code simply and precisely because it does not code only for itself. It codes for another type of molecule, the proteins. The rest of the cellular machinery does not need a code, or to be reduced to digital information, *precisely because it represents itself*. To Dawkins' famous description of DNA as the eternal replicator (Dawkins 1976, ch. 2), we should add that egg cells, and sperm, also form an eternal line, just as do all unicellular organisms. DNA cannot form an eternal line on its own.

So, although we might characterize the cell information as analogue, that is only to contrast it with being digital. But it is *not* an analogue *representation*. It itself *is* the self-sustaining structure that we inherit and it reproduces itself directly. Cells make more cells, which make more cells (and use DNA to do so), …, etc. The inheritance is robust: liver cells make liver cells for many generations of liver cells, at each stage marking their genomes to make that possible. So do all the other 200 or so cell types in the body (Noble 2006, ch. 7). Yet, the genome is the same throughout. That common 'digital' code is made to dance to the totally different instructions of the specific cell types. Those instructions are 'analogue', in the form of continuous variations in imposed patterns of gene expression. The mistake in thinking of gene expression as digital lies in focusing entirely on the CGAT codes, not on the continuously variable degree of expression. It is surely artificial to emphasize one or the other. When it comes to the pattern of expression levels, the information is analogue.

So, I do not think we get much leverage on the question of privileged causality (DNA or non-DNA) through the digital–analogue comparison route. We might even see the digital coding itself as the really hazardous step—and indeed it does require complex machinery to check for errors in large genomes (Maynard Smith & Szathmáry 1995; Maynard Smith 1998). Having lipid membranes that automatically 'accept' certain lipids to integrate into their structure and so to grow, enable cells to divide and so on seems also to be chemically reliable. The lipid membranes

are also good chemical replicators. That process was probably 'discovered' and 'refined' by evolution long before cells 'captured' genes and started the process towards the full development of cells as we now know them. I suspect that primitive cells, probably not much more than lipid envelopes with a few RNA enzymes (Maynard Smith & Szathmáry 1995, 1999), 'knew' how to divide and have progeny long before they acquired DNA genomes.

5. An impossible experiment

Could we get a hold on the question by a more direct (but currently and probably always impossible; Keller 2000*a*,*b*) biological experiment? Would the complete DNA sequence be sufficient to 'resurrect' an extinct species? Could dinosaur DNA (let us forget about all the technical problems here), for example, be inserted into, say, a bird egg cell. Would it generate a dinosaur, a bird, or some extraordinary hybrids?

At first sight, this experiment seems to settle the question. If we get a dinosaur, then DNA is the primary, privileged information. The non-DNA is secondary. I suspect that this is what most 'genetic determinists' would expect. If we get a bird, then the reverse is true (this is *highly* unlikely in my or anyone else's view). If we get a hybrid, or nothing (I suspect that this would be the most likely outcome), we could maintain a view of DNA primacy by simply saying that there is, from the DNA's point of view, a fault in the egg cell machinery. But note the phrase 'DNA's point of view' in that sentence. It already gives the DNA primacy and so begs the question.

The questions involved in such experiments are important. Cross-species clones are of practical importance as a possible source of stem cells. They could also reveal the extent to which egg cells are species specific. This is an old question. Many early theories of what was called 'cytoplasm inheritance' were eventually proved wrong (Mayr 1982), though Mayr notes that 'The old belief that the cytoplasm is important in inheritance … is not dead, although it has been enormously modified.' I suspect that the failure of most cross-species clones to develop to the adult stage is revealing precisely the extent to which 'the elaborate architecture of the cytoplasm plays a greater role than is now realized' (Mayr 1982). Since we cannot have the equivalent of mutations in the case of the non-DNA inheritance, using different species may be our only route to answering the question.

Interspecies cloning has already been attempted, though not with extinct animals. About a decade ago, J. B. Cibelli of Michigan State University tried to insert his own DNA into a cow egg cell and even patented the technique. The experiment was a failure and ethically highly controversial. Cibelli has since failed to clone monkey genes in cow's eggs. The only successful case is of a wild ox (a banteng *Bos javanicus*) cloned in domestic cow's eggs. The chances are that the technique will work only on very closely related species. At first sight, a banteng looks very much like a cow and some have been domesticated in the same way. More usually, interspecies clones fail to develop much beyond the early embryo.

But however interesting these experiments are, they are misconceived as complete answers to the question I am raising. Genomes and cells have evolved together (Maynard Smith & Szathmáry 1995). Neither can do anything without

the other. If we got a dinosaur from the imagined experiment, we would have to conclude that dinosaur and bird egg cells are sufficiently similar to make that possible. The difference (between birds and dinosaurs) would then lie in the DNA not in the rest of the egg cell. Remember that eukaryotic cells evolved aeons before dinosaurs and birds and so all cells necessarily have much of their machinery in common. But that difference does not give us grounds for privileging one set of information over the other. If I play a PAL video tape on a PAL reading machine, surely, I get a result that depends specifically on the information on the tape, and that would work equally well on another PAL reader, but I would get nothing at all on a machine that does not read PAL coding. The egg cell in our experiment still ensures that we get an organism at all, if indeed we do get one, and that it would have many of the characteristics that are common between dinosaurs and birds. The egg cell inheritance is not limited merely to the *differences* we find. It is essential for the *totality* of what we find. Each and every high-level function depends on effects attributable to *both* the DNA and the rest of the cell. 'Studying biological systems means more than breaking the system down into its components and focusing on the digital information encapsulated in each cell' (Neuman 2007).

6. The 'genetic differential effect problem'

This is a version of a more general argument relating to genes (defined here as DNA sequences) and their effects. Assignment of functions to genes depends on observing differences in phenotype consequent upon *changes* (mutations, knockouts, etc.) in genotype. Dawkins made this point very effectively when he wrote 'It is a fundamental truth, though it is not always realized, that *whenever* a geneticist studies a gene 'for' any phenotypic character, he is always referring to a *difference* between two alleles' (Dawkins 1982).

But differences cannot reveal the totality of functions that a gene may be involved in, since they cannot reveal all the effects that are common to the wild and mutated types. We may be looking at the tip of an iceberg. And we may even be looking at the wrong tip since we may be identifying a gene through the pathological effects of just one of its mutations rather than by what it does for which it must have been selected. This must be true of most so-called oncogenes, since causing cancer is unlikely to be a function for which the genes were selected. This is why the Gene Ontology (GO) Consortium (http://geneontology.org/) excludes oncogenesis: 'oncogenesis is not a valid GO term because causing cancer is not the normal function of any gene'. Actually, causing cancer could be a function if the gene concerned has other overwhelming beneficial effects. This is a version of the 'sickle cell' paradigm (Jones 1993, p. 219) and is the reason why I do not think oncogenesis could *never* be a function of a gene: nature plays with balances of positive and negative effects of genes (see 'Faustian pacts with the devil'; Noble 2006, p. 109).

Identifying genes by *differences* in phenotype correlated with *those* in genotype is therefore hazardous. Many, probably most, genetic modifications are buffered. Organisms are robust. They have to be to have succeeded in the evolutionary process. Even when the function of the gene is known to be significant, a knockout or mutation may not reveal that significance. I will refer to this

problem as the genetic differential effect problem. My contention is that it is a very severe limitation in unravelling the causal effects of genes. I will propose a solution to the problem later in this paper.

It is also important to remember that large numbers (hundreds or more) of genes are involved in each and every high-level function and that, at that level, individual genes are involved in many functions. We cannot assume that the first phenotype–genotype correlation we found for a given gene is its only or even its main function.

7. Problems with the central dogma

The video reader is a good analogy so far as it goes in emphasizing that the reading machinery must be compatible with the coding material, but it is also seriously limited in the present context. It is best seen as an analogy for the situation seen by those who take an extension of the central dogma of biology as correct: information passes from the coded material to the rest of the system but not the other way. What we now know of epigenetics requires us to modify that view. The cell machinery does not just read the genome. It imposes extensive patterns of marking and expression on the genome (Qiu 2006). This is what makes the precise result of our imagined experiment so uncertain. According to the central dogma, if the egg cell is compatible, we will automatically get a dinosaur, because the DNA dictates everything. If epigenetic marking is important, then the egg cell also plays a determining, not a purely passive, role. There are therefore two kinds of influence that the egg cell exerts. The first is that it is totally necessary for any kind of organism at all to be produced. It is therefore a primary 'genetic cause' in the sense that it is essential to the production of the phenotype and is passed on between the generations. The second is that it exerts an influence on what kind of organism we find. It must be an empirical question to determine how large the second role is. At present, we are frustrated in trying to answer that question by the fact that virtually all cross-species clones do not develop into adults. As I have already noted, that result itself suggests that the second role is important.

It would also be an interesting empirical question to determine the range of species across which the egg cell machinery is sufficiently similar to enable different genomes to work, but that tells us about similarities of the match of different genomes with the egg cells of different species, and their mutual compatibility in enabling development, not about the primacy or otherwise of DNA or non-DNA inheritance. In all cases, the egg cell machinery is as necessary as the DNA. And, remember, as 'information' it is also vast.

Note also that what is transferred in cross-species cloning experiments is not just the DNA. Invariably, the whole nucleus is inserted, with all its machinery (Tian *et al.* 2003). If one takes the contribution of the egg cell seriously, that is a very serious limitation. The nucleus also has a complex architecture in addition to containing the DNA, and it must be full of transcription factors and other molecules that influence epigenetic marking. Strictly speaking, we should be looking at the results of inserting the raw DNA into a genome-free nucleus of an egg cell, not at inserting a whole nucleus, or even just the chromosomes, into an enucleated egg cell. No one has yet done that. And would we have to include

the histones that mediate many epigenetic effects? This is one of the reasons, though by no means the only one, why the dinosaur cloning experiment may be impossible.

To conclude this section, if by genetic causation we mean the totality of the inherited causes of the phenotype, then it is plainly incorrect to exclude the non-DNA inheritance from this role, and it probably does not make much sense to ask which is more important, since only an interaction between DNA and non-DNA inheritance produces anything at all. Only when we focus more narrowly on changes in phenotype attributable to differences in genotype (which is how functionality of genes is currently assessed) could we plausibly argue that it is all down to the DNA, and even that conclusion is uncertain until we have carried out experiments that may reveal the extent to which egg cells are species specific, since nuclear DNA marking may well be very important.

8. Genetic programs?

Another analogy that has come from comparison between biological systems and computers is the idea of the DNA code being a kind of program. This idea was originally introduced by Monod & Jacob (1961) and a whole panoply of metaphors has now grown up around their idea. We talk of gene networks, master genes and gene switches. These metaphors have also fuelled the idea of genetic (DNA) determinism.

But there are no purely gene networks! Even the simplest example of such a network—that discovered to underlie circadian rhythm—is not a gene network, nor is there a gene for circadian rhythm. Or, if there is, then there are also proteins, lipids and other cellular machinery for circadian rhythm.

The circadian rhythm network involves at least three other types of molecular structures in addition to the DNA code. The stretch of DNA called the period gene (*per*) codes for a protein (PER) that builds up in the cell cytoplasm as the cellular ribosome machinery makes it. PER then diffuses slowly through the nuclear (lipid and protein) membrane to act as an inhibitor of *per* expression (Hardin *et al.* 1990). The cytoplasmic concentration of PER then falls, and the inhibition is slowly removed. Under suitable conditions, this process takes approximately 24 hours. It is the whole network that has this 24 hour rhythm, not the gene (Foster & Kreitzman 2004). However else this network can be described, it is clearly not a gene network. At the least, it is a gene–protein–lipid–cell network. It does not really make sense to view the gene as operating without the rest of the cellular machinery. So, if this network is part of a 'genetic program', then the genetic program is not a DNA program. It does not lie within the DNA coding. Moreover, as Foster & Kreitzman emphasized, there are many layers of interactions overlaid onto the basic mechanism—so much so that it is possible to knock out the *CLOCK* gene in mice and retain circadian rhythm (Debruyne *et al.* 2006). I prefer therefore to regard the DNA as a database rather than as a program (Atlan & Koppel 1990; Noble 2006). What we might describe as a program uses that database, but is not controlled by it.

The plant geneticist Coen (1999) goes even further. I will use my way of expressing his point, but I would like to acknowledge his ideas and experiments as a big influence on my thinking about this kind of question. In the early days of

computing, during the period in which Monod & Jacob (1961) developed their idea of *le programme génétique*, a program was a set of instructions separate from the functionality it serves. The program was a complete piece of logic, a set of instructions, usually stored on cards or tapes, that required data to work on and outputs to produce. Pushing this idea in relation to the DNA/non-DNA issue, we arrive at the idea that there is a program in the DNA, while the data and output is the rest: the cell and its environment. Jacob was quite specific about the analogy: 'The programme is a model borrowed from electronic computers. It equates the genetic material with the magnetic tape of a computer' (Jacob 1982). That analogy is what leads people to talk of the DNA 'controlling' the rest of the organism.

Coen's point is that there is no such distinction in biological systems. As we have seen, even the simplest of the so-called gene networks are not 'gene programs' at all. The process is the functionality itself. There is no separate program. I see similar conclusions in relation to my own field of heart rhythm. There is no heart rhythm program (Noble 2008, in press), and certainly not a heart rhythm genetic program, separate from the phenomenon of heart rhythm itself. Surely, we can refer to the functioning networks of interactions involving genes, proteins, organelles, cells, etc. as programs if we really wish to. They can also be represented as carrying out a kind of computation (Brenner 1998), in the original von Neumann sense introduced in his theory of self-reproducing machines. But if we take this line, we must still recognize that this computation does not tell something else to carry out the function. It is itself the function.

Some will object that computers are no longer organized in the way they were in the 1960s. Indeed not, and the concept of a program has developed to the point at which distinctions between data and instructions, and even the idea of a separate logic from the machine itself, may have become outdated. Inasmuch as this has happened, it seems to me that such computers are getting a little closer to the organization of living systems.

Not only is the *period* gene not the determinant of circadian rhythm, either alone or as a part of a pure gene network, but also it could be argued that it is incorrect to call it a 'circadian rhythm' gene. Or, if it is, then it is also a development gene, for it is used in the development of the fly embryo. And it is a courtship gene! It is used in enabling male fruitflies to sing (via their wing-beat frequencies) to females of the correct species of fruitfly (more than 3000 such species are known). Genes in the sense of the stretches of DNA are therefore like pieces of re-usable Lego. That is, in principle, why there are very few genes compared with the vast complexity of biological functions. Needless to say, human courtship uses other genes! And all of those will be used in many other functions. My own preference would be to cease using high-level functionality for naming genes (meaning here DNA sequences), but I realize that this is now a lost cause. The best we can do is to poke fun at such naming, which is why I like the Fruit Fly Troubadour Gene story (Noble 2006, p. 72).

9. Higher-level causation

I have deliberately couched the arguments so far in molecular terms because I wish to emphasize that the opposition to simplistic gene determinism, gene networks and genetic programs is not based only on the distinction between

higher- and lower-level causation, but also there are additional factors to be taken into account as a consequence of multilevel interactions.

The concept of level is itself problematic. It is a metaphor, and a very useful one in biology. Thus, there is a sense in which a cell, for example, and an organ or an immune system, is much more than its molecular components. In each of these cases, the molecules are constrained to cooperate in the functionality of the whole. Constrained by what? A physicist or an engineer would say that the constraints do not lie in the laws governing the behaviour of the individual components—the same quantum mechanical laws will be found in biological molecules as in molecules not forming part of a biological system. The constraints lie in the boundary and initial conditions: 'organisation becomes cause in the matter' (Strohman 2000; Neuman 2006). These conditions, in turn, are constrained by what? Well, ultimately by billions of years of evolution. That is why I have used the metaphor of evolution as the composer (Noble 2006, ch. 8). But that metaphor is itself limited. There may have been no direction to evolution (but for arguments against this strict view, see Jablonka & Lamb 2005). We are talking of a set of historical events, even of historical accidents. The information that is passed on through downward causation is precisely this set of initial and boundary conditions without which we could not even begin to integrate the equations representing molecular causality.

To spell this out in the case of the circadian rhythm process, this is what determines the cytoplasm volume in which the concentration of the protein changes, the speed with which it crosses the nuclear membrane, the speed with which ribosomes make new protein and so on. And those characteristics will have been selected by the evolutionary process to give a roughly 24 hour rhythm. Surely, each molecule in this process does not 'know' or represent such information, but the ensemble of molecules does. It behaves differently from the way in which it would behave if the conditions were different or if they did not exist at all. This is the sense in which molecular events are different as a consequence of the life process. Moreover, the boundary and initial conditions are essentially global properties, identifiable at the level at which they can be said to exist.

What is metaphorical here is the notion of 'up and down' (Noble 2006, ch. 10) —it would be perfectly possible to turn everything conceptually upside down so that we would speak of upward causation instead of downward causation. The choice is arbitrary, but important precisely because the principle of reductionism is always to look for 'lower-level' causes. That is the reductionist prejudice and it seems to me that it needs justification; it is another way in which we impose our view on the world.

Although the concept of level is metaphorical, it is nevertheless an essential basis for the idea of multilevel causation. The example I often give is that of pacemaker rhythm, which depends on another global property of cells, i.e. the electrical potential, influencing the behaviour of the individual proteins, the ionic channels, which in turn determine the potential. There is a multilevel feedback network here: channels → ionic current → electrical potential → channel opening or closing → ionic current and so on. This cycle is sometimes called the Hodgkin cycle, since it was Alan Hodgkin who originally identified it in the case of nerve excitation (Hodgkin & Huxley 1952).

Similarly, we can construct feedback networks of causation for many other biological functions. I see the identification of the level at which such networks are integrated, i.e. the highest level involved in the network, as being a primary aim of systems biology. This will also be the lowest level at which natural selection can operate since it is high-level functionality that determines whether organisms live or die. We must shift our focus away from the gene as the unit of selection to that of the whole organism (Tautz 1992).

But I also have hesitations about such language using the concepts of levels and causation. My book, in its last chapter, recommends throwing all the metaphors away once we have used them to gain insight (Noble 2006, ch. 10). In the case of the cycles involving downward causation, my hesitation is because such language can appear to make the causation involved be sequential in time. I do not see this as being the case. In fact, the cell potential influences the protein kinetics at exactly the same time as they influence the cell potential. Neither is primary or privileged as causal agency either in time or in space. This fact is evident in the differential equations we use. The physical laws represented in the equations themselves, and the initial and boundary conditions, operate *at the same time* (i.e. during every integration step, however infinitesimal), not sequentially.

This kind of conceptual problem (causality is one of our ways of making sense of the world, not the world's gift to us) underlies some knotty problems in thinking about such high-level properties as intentionality. As I show in *The music of life* (Noble 2006, ch. 9), looking for neural or, even worse, genetic 'causes' of an intention is such a will-of-the-wisp. I believe that this is the reason why the concept of downward causation may play a fundamental role in the philosophy of action (intentionality, free will, etc.).

I am also conscious of the fact that causality in any particular form does not need to be a feature of all successful scientific explanations. General relativity theory, for example, changes the nature of causality through replacing movement in space by geodesics in the structure of space–time. At the least, that example shows that a process that requires one form of causality (gravity acting at a distance between bodies) in one theoretical viewpoint can be seen from another viewpoint to be unnecessary. Moreover, there are different forms of causality, ranging from proximal causes (one billiard ball hitting another) to ultimate causes of the kind that evolutionary biologists seek in accounting for the survival value of biological functions and features. Genetic causality is a particularly vexed question partly not only because the concept of a gene has become problematic, as we have seen in this paper, but also because it is not usually a proximal cause. Genes, as we now define them in molecular biological terms, lie a long way from their phenotypic effects, which are exerted through many levels of biological organization and subject to many influences from both those levels and the environment. We do not know what theories are going to emerge in the future to cope with the phenomenon of life. But we can be aware that our ways of viewing life are almost certainly not the only ones. It may require a fundamental change in the mindset to provoke us to formulate new theories. I hope that this paper will contribute to that change in the mindset.

10. Unravelling genetic causation: the solution to the genetic differential effect problem

Earlier in this paper, I referred to this problem and promised a solution. The problem arises as an inherent difficulty in the 'forward' (reductionist) mode of explanation. The consequences of manipulations of the lowest end of the causal chain, the genes, can be hidden by the sheer cleverness of organisms to hide genetic mistakes and problems through what modern geneticists call genetic buffering and what earlier biologists would call redundancy or back-up mechanisms that kick in to save the functionality. The solution is not to rely solely on the forward mode of explanation. The backward mode is sometimes referred to as reverse engineering. The principle is that we start the explanation at the higher, functional level, using a model that incorporates the forward mode knowledge but, crucially, also incorporates higher level insights into functionality. For example, if we can successfully model the interactions between all the proteins involved in cardiac rhythm, we can then use the model to assess qualitatively and quantitatively the contribution that each gene product makes to the overall function. That is the strength of reverse engineering. We are no longer dealing just with differences. If the model is good, we are dealing with the totality of the gene function within the process we have modelled. We can even quantify the contribution of a gene product whose effect may be largely or even totally buffered when the gene is manipulated (see Noble 2006, p. 108). This is the reason why higher level modelling of biological function is an essential part of unravelling the functions of genes: 'Ultimately, *in silico* artificial genomes and *in vivo* natural genomes will translate into each other, providing both the possibility of forward and reverse engineering of natural genomes' (Werner 2005).

11. Conclusions

The original notion of a gene was closely linked to the causes of particular phenotype characteristics, so the question of causal relationships between genes and phenotype were circular and so hardly had much sense. The question of causality has become acute because genes are now identified more narrowly with particular sequences of DNA. The problem is that these sequences are uninterpretable outside the cellular context in which they can be read and so generate functionality. But that means that the cell is also an essential part of the inheritance and therefore was, implicitly at least, a part of the original definition of a gene. Depending on how we quantify the comparison between the contributions, it may even be the larger part. Genetic information is not confined to the digital information found in the genome. It also includes the analogue information in the fertilized egg cell. If we were ever to send out through space in an Earth-life capsule the information necessary to reconstruct life on Earth on some distant planet, we would have to include both forms of information. Now that we can sequence whole genomes, the difficult part would be encoding information on the cell. As Sydney Brenner has said, 'I believe very strongly that the fundamental unit, the correct level of abstraction, is the cell and not the genome' (Lecture to Columbia University in 2003). This fundamental insight has yet to be adopted by the biological science community in a way that will ensure

success in unravelling the complexity of interactions between genes and their environment. In particular, the power of reverse engineering using mathematical models of biological function to unravel gene function needs to be appreciated. Multilevel systems biology requires a more sophisticated language when addressing the relationships between genomes and organisms.

Work in the author's laboratory is supported by EU FP6 BioSim network, EU FP7 PreDiCT project, BBSRC and EPSRC. I would like to acknowledge valuable discussions with Jonathan Bard, John Mulvey, James Schwaber, Eric Werner and the critical comments of the referees.

References

Atlan, H. & Koppel, M. 1990 The cellular computer DNA: program or data. *Bull. Math. Biol.* **52**, 335–348. (doi:10.1007/BF02458575)

Bickel, K. S. & Morris, D. R. 2006 Silencing the transcriptome's dark matter: mechanisms for suppressing translation of intergenic transcripts. *Mol. Cell* **22**, 309–316. (doi:10.1016/j.molcel.2006.04.010)

Bird, A. 2007 Perceptions of epigenetics. *Nature* **447**, 396–398. (doi:10.1038/nature05913)

Brenner, S. 1998 Biological computation. In *The limits of reductionism in biology* (eds G. R. Bock & J. A. Goode). Novartis Foundation Symposium, no. 213, pp. 106–116. London, UK: Wiley.

Coen, E. 1999 *The art of genes*. Oxford, UK: Oxford University Press.

Dawkins, R. 1976 *The selfish gene*. Oxford, UK: Oxford University Press.

Dawkins, R. 1982 *The extended phenotype*. London, UK: Freeman.

Debruyne, J. P., Noton, E., Lambert, C. M., Maywood, E. S., Weaver, D. R. & Reppert, S. M. 2006 A clock shock: mouse CLOCK is not required for circadian oscillator function. *Neuron* **50**, 465–477. (doi:10.1016/j.neuron.2006.03.041)

Dupré, J. 1993 *The disorder of things*. Cambridge, MA: Harvard University Press.

Foster, R. & Kreitzman, L. 2004 *Rhythms of life*. London, UK: Profile Books.

Gluckman, P. & Hanson, M. 2004 *The fetal matrix. Evolution, development and disease*. Cambridge, UK: Cambridge University Press.

Hardin, P. E., Hall, J. C. & Rosbash, M. 1990 Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* **343**, 536–540. (doi:10.1038/343536a0)

Hodgkin, A. L. & Huxley, A. F. 1952 A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544.

Jablonka, E. & Lamb, M. 1995 *Epigenetic inheritance and evolution. The Lamarckian dimension*. Oxford, UK: Oxford University Press.

Jablonka, E. & Lamb, M. 2005 *Evolution in four dimensions*. Boston, MA: MIT Press.

Jacob, F. 1982 *The possible and the actual*. New York, NY: Pantheon Books.

Johannsen, W. 1909 *Elemente der exakten Erblichkeitslehre*. Jena, Germany: Gustav Fischer.

Jones, S. 1993 *The language of the genes*. London, UK: HarperCollins.

Keller, E. F. 2000*a* *The century of the gene*. Cambridge, MA: Harvard University Press.

Keller, E. F. 2000*b* Is there an organism in this text? In *Controlling our destinies. Historical, philosophical, ethical and theological perspectives on the human genome project* (ed. P. R. Sloan), pp. 273–288. Notre Dame, IN: University of Notre Dame Press.

Kitcher, P. 1982 Genes. *Br. J. Philos. Sci.* **33**, 337–359. (doi:10.1093/bjps/33.4.337)

Maynard Smith, J. 1998 *Evolutionary genetics*. New York, NY: Oxford University Press.

Maynard Smith, J. & Szathmáry, E. 1995 *The major transitions in evolution*. Oxford, UK: Oxford University Press.

Maynard Smith, J. & Szathmáry, E. 1999 *The origins of life*. New York, NY: Oxford University Press.

Mayr, E. 1982 *The growth of biological thought*. Cambridge, MA: Harvard University Press.

Monod, J. & Jacob, F. 1961 Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401.

Neuman, Y. 2006 Cryptobiosis: a new theoretical perspective. *Progr. Biophys. Mol. Biol.* **92**, 258–267. (doi:10.1016/j.pbiomolbio.2005.11.001)

Neuman, Y. 2007 The rest is silence. *Perspect. Biol. Med.* **50**, 625–628. (doi:10.1353/pbm.2007.0053)

Noble, D. 2006 *The music of life*. Oxford, UK: Oxford University Press.

Noble, D. 2008 Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* **93**, 16–26. (doi:10.1113/expphysiol.2007.038695)

Noble, D. In press. Commentary on Scherrer & Jost (2007) Gene and genon concept: coding versus regulation. *Theory Biosci.* **127**.

Pearson, H. 2006 Genetics: what is a gene? *Nature* **441**, 398–401. (doi:10.1038/441398a)

Pichot, A. 1999 *Histoire de la notion de gène*. Paris, France: Flammarion.

Qiu, J. 2006 Epigenetics: unfinished symphony. *Nature* **441**, 143–145. (doi:10.1038/441143a)

Roux, A., Cuvelier, D., Bassereau, P. & Goud, B. 2008 Intracellular transport. From physics to biology. *Ann. NY Acad. Sci.* **1123**, 119–125. (doi:10.1196/annals.1420.014)

Scherrer, K. & Jost, J. 2007 Gene and genon concept: coding versus regulation. *Theory Biosci.* **126**, 65–113. (doi:10.1007/s12064-007-0012-x)

Schrödinger, E. 1944 *What is life?* Cambridge, UK: Cambridge University Press.

Strohman, R. C. 2000 Organisation becomes cause in the matter. *Nat. Biotechnol.* **18**, 575–576. (doi:10.1038/76317)

Tautz, D. 1992 Redundancies, development and the flow of information. *Bioessays* **14**, 263–266. (doi:10.1002/bies.950140410)

Tian, X. C., Kubota, C., Enright, B. & Yang, X. 2003 Cloning animals by somatic cell nuclear transfer—biological factors. *Reprod. Biol. Endocrinol.* **1**, 98–105. (doi:10.1186/1477-7827-1-98)

Werner, E. 2005 Genome semantics. *In silico* multicellular systems and the central dogma. *FEBS Lett.* **579**, 1779–1782. (doi:10.1016/j.febslet.2005.02.011)

Werner, E. 2007 How central is the genome? *Science* **317**, 753–754. (doi:10.1126/science.1141807)

Molecular Systems Biology 5; Article number 292; doi:10.1038/msb.2009.51
Citation: *Molecular Systems Biology* 5:292

www.molecularsystemsbiology.com

molecular systems biology

EDITORIAL

Systems biology and the virtual physiological human

Molecular Systems Biology **5:** 292; published online 28 July 2009; doi:10.1038/msb.2009.51

Introduction

The virtual physiological human (VPH) initiative is intended to support the development of patient-specific computer models and their application in personalised and predictive healthcare. The VPH, a core target of the European Commission's 7th Framework Programme, will serve as a '*methodological and technological framework that, once established, will enable collaborative investigation of the human body as a single complex system*' (http://www.europhysiome.org/roadmap/). As such, the VPH initiative constitutes an integral part of the international *Physiome Project* (http://www.physiome.org.nz/), a worldwide public domain effort to develop a computational framework for the quantitative description of biological processes in living systems across all relevant levels of structural and functional integration, from molecule to organism, including the human (Kohl *et al*, 2000; Bassingthwaighte *et al*, 2009).

So, what is the connection between this grand challenge and systems biology? To explore this, we must first agree on what we take systems biology to mean.

Systems biology

Description versus definition

Descriptions of systems biology range from the view that it is merely 'new wording, more fashionable, for physiology' (http://is.gd/tQJL), to the all-inclusive 'systems biology involves the application of experimental, theoretical, and computational techniques to the study of biological organisms at all levels, from the molecular, through the cellular, to the organ, organism, and populations. Its aim is to understand biological processes as integrated systems instead of as isolated parts' (http://is.gd/tQK0).

At the same time, attempts to concisely *define* systems biology have not yielded definitive form of words that is acceptable to the majority of researchers engaged in what they consider to be systems biology.

One of the reasons for this situation may be that many different scientific streams have come together in the systems biology pool (see also Bassingthwaighte *et al*, 2009), each with its own conceptual and terminological legacy.

But, another possible explanation for this apparent shortcoming is that systems biology may constitute an *approach* (as detailed below), rather than a discipline (such as biology), or a destination (such as the VPH). Such a scientific approach can be explained *descriptively*, but cannot necessarily be defined *prescriptively*.

In either case, the lack of a generally acceptable definition of systems biology need not be regarded as a surprise, or even as a disadvantage, as the artificial uniformity that could be associated with a definition might exclude important current or future work.

Terminological origins

It may be helpful, at this stage, to step back and consider the etymology of terms, before discussing their possible interrelation.

Biology is contracted from *bios* (Greek for 'life') and *logos* (Greek for 'reasoned account'). It is the science, or the logic, of life (Boyd and Noble, 1993).

A system is 'the object' of the activity *synthithemi* (Greek for 'I put together') and has been defined as follows: '*A system is an entity that maintains its existence through the mutual interaction of its parts*' (von Bertalanffy, 1968). In keeping with this concept (Figure 1), research into systems therefore must combine:

(i) the identification and
(ii) detailed characterisation of the parts, with the
(iii) investigation of their interaction with each other and
(iv) with their wider environment, to
(v) elucidate the maintenance of the entity.

Subject matter

On the basis of the definition of a system, systems biology can be seen as a conceptual approach to biological research that consciously combines 'reductionist' (parts; points i and ii) and 'integrationist' (interactions; points iii and iv) research, to understand the nature and maintenance of entities (point v). In biological systems, preservation of entity includes a broad range of behaviours, including growth and development, adaptation and maladaptation, and progeny, which explains why streams from so many different research directions must be pooled.

In addition, the 'parts' of a biological system (e.g. organs of a body, or tissues within an organ, etc.) can usually be broken

down into smaller biologically relevant entities (such as cells, proteins, amino acids), which—when focussing at a lower level of structural integration—form 'systems' in their own right. This illustrates two further points: first, systems biology as an approach can be applied to research targets independent of their 'scale', that is, their level of structural and functional complexity and second, no particular scale has privileged relevance for systems biology (Noble 2008a, 2008c). From the multi-scale nature of biological systems, it follows further that systems biology inherently involves a multi-scale approach (see below).

So, does this mean that there is nothing special about systems biology? Is it really just another, more fashionable label for good old physiology?

Probably not. Systems biology forms a logical juxtaposition to the recently prevailing 'reductionist' drive, serving as the 'post-genomic' manifestation of the need to balance dissection and synthesis. Certain aspects of systems biology do indeed mirror the 'pre-genomic' approach of subjects such as physiology, but at a higher level. Thus, Claude Bernard showed the way as early as the 19th century and specifically called for the mathematical analysis of biological phenomena (see Noble, 2008a). However, with a few notable exceptions, such as the Hodgkin–Huxley equations for the nerve impulse (Hodgkin and Huxley 1952), their application to the heart (Noble, 1962), or the early ideas of Guyton for a quantitative model of the circulation (Guyton *et al*, 1972), classic physiology largely lacked the ability to pursue the quantitative integration of observed behaviour. This may be one reason why it failed to compete with the rise of molecular biology, which was perceived to be more solidly quantitative. In fact, many academic departments of physiology became molecular or cellular, in focus and in name.

Having turned full circle on what the dialectic method depicts as a three-dimensional spiral of development, we have come 'back to the future', now that bio-science can harness the power of mathematics and computation and apply it to a re-integration of the pieces of the jigsaw—which have been amply provided by reductionist research approaches. Systems biology therefore thrives on the revolutionary improvement of experimental techniques to investigate system components and their interactions, and on significant advances in computational power, tools, and techniques, which allow quantitative modelling and reintegration at hitherto unimaginable detail and breadth. Modern computational models thus address points (i) to (v) above, and project between them, while observing elementary rules such as conservation of mass, energy, and matter and taking into account natural restrictions imposed on parts and interactions by the system's own properties (e.g. a water-based solute system will impose different constraints compared to a hydro-carbon based one; dark-blue background in Figure 1).

So, perhaps this is where the essence of systems biology lies: by providing a framework for the re-unification of biological studies with 'the other' sciences, and their joint application to iterative reduction and synthesis, it forms the approach on which quantitative descriptions of parts (i and ii) and their interactions (iii and iv) give rise to an understanding of the maintenance of biological entities (v) across all relevant levels of structural and functional integration (Figure 2).

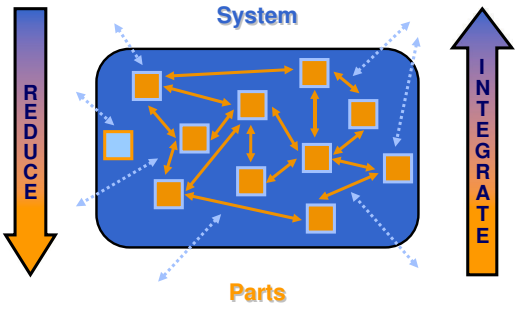


Figure 1 A system as an 'entity that maintains its existence through the mutual interaction of its parts' (von Bertalanffy, 1968). Systems research must combine the (i) identification and (ii) detailed characterisation of parts (orange boxes, as opposed to 'look-alikes', pale blue box, which need to be identified and excluded), with the exploration of their interactions (iii) with each other (orange arrows), and (iv) with the environment (pale blue dashed arrows affecting parts either directly, or indirectly through modulation of internal interactions), to develop a (v) systemic understanding (an important, but often overlooked, aspect is that the system itself not only enables, but also restricts, the type and extent of functions and interactions that may occur; dark-blue box). Systems research therefore requires a combination of reductionist and integrative tools and techniques.

An important aspect of this summary is the plural of 'quantitative description'. Like their experimental counterparts, computational models are—by the very definition of the term 'model'—*simplified* representations of reality. Like tools in a toolbox, models for biomedical research, whether 'wet' or 'dry', have a range of applications for which they are suitable. This suitability is affected by the extent to which models are *representative* of the aspect of reality that they mimic; *relevant* for the question under investigation; *reasonable* in terms of their cost (including not merely financial considerations, but also resources such as time, training requirements, or ethical dimensions); and *reproducible* (a challenge also for computational models, not only when they include descriptions of stochasticity, but also when they exhibit language-, compiler-, or hardware-dependence) (Kohl *et al*, 2006). Thus, the multi-level nature of biological systems must find suitable reflection in an integrated set of multiple models, both experimental and computational. This will be discussed next in the context of the VPH initiative.

Systems biology and the VPH

The VPH initiative

As its name suggests, the VPH initiative targets the whole human body as the system of interest. But, it does not herald a return to classical top-down physiology from entity to parts. The aim is to understand human physiology quantitatively, as a dynamic system, and at all relevant levels between genes and the organism.

Equally, it is not a bottom-up analysis from parts to entities. This would be impossible, both conceptually (as the 'parts' of the whole organism form systemic 'entities' of their own), and practically (as the number of possible combinations of interactions between the products of 25 000 genes is simply too vast (Feytmans *et al*, 2005)).

The approach is better characterised by a term introduced by Sydney Brenner, 'middle-out' (Brenner *et al*, 2001), which is based on conceptualising insight at whichever level there is a good understanding of data and processes, and on then

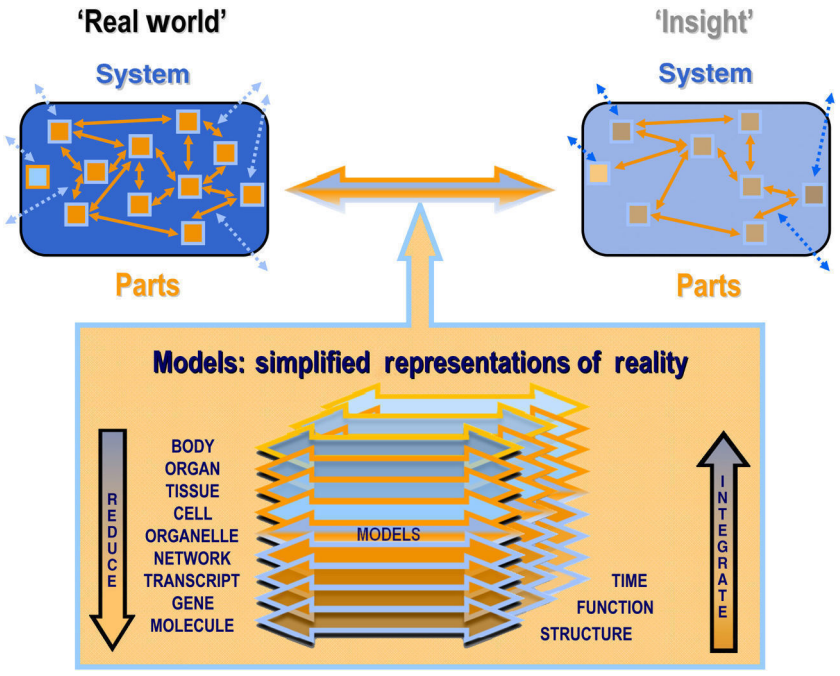


Figure 2 Our understanding of 'real world systems' (top left) usually forms a simplified representation (top right) of that reality, and therefore represents a model in its own right. The progressive development of this understanding is based on the application and analysis of experimental and theoretical models. For biological systems research, these models allow the exploration of partial systems behaviour at all relevant structural levels between body and molecule. 'Wet' experimental models are developed through a broad range of research directions and provide increasingly detailed data on structure–function relations and their change over time. This can be re-integrated using 'dry' conceptual (thought) and formal (computation) models. Many of these developments occur in parallel. Systems biology provides the framework for the targeted interrelation of these different facets of model application to bio-medical research and development. Note that, for simplicity, this diagram depicts models by horizontal arrows, although models can involve multiple scales.

connecting this to higher and lower levels of structural and functional integration. In a system of multi-level interactions that involves both regulatory feedforward and feedback pathways, as well as environmentally prescribed parameter constraints, there is really no alternative to breaking in at one level (the 'middle' part of the metaphor) and then reaching 'out' to neighbouring levels using appropriate, experimentally founded and validated mathematical methods (Bassingthwaighte *et al*, 2009).

Of course, one has to be aware of the possible (and in the present case counterproductive) association of the expressions 'higher' or 'lower' level with 'superior' or 'inferior' in terms of relevance for systems function. Regulatory interactions are, by definition, two-way ('regulatory loop'), and the metaphoric use of high and low is associated here simply with the notion of spatial scale, not relevance. Furthermore, it is important to realize that influences from 'outer' levels to the 'middle' are equally relevant. One might call this an outside-in approach, illustrating the utility and limitations of metaphors, simplified representations of a concept or idea (models!), which are not necessarily of much help when used outside the applicable contextualisation for which they were developed.

A lead example: systems biology of the virtual heart

We will illustrate the ideas discussed above by considering the modelling of cardiac structure and function, partly because that is the area of our own research, but also because, by common consent, it is the most highly developed example of a virtual organ, with applications already within the pharmaceutical industry and in the development of medical devices (Hunter *et al*, 2001; Noble 2008b). There are three reasons for this situation.

First, cardiac cell models have now benefited from a *track record* of nearly 50 years of iterative interaction between modelling and experimentation, with an accumulating body of insights derived as much from the 'failures' as from the 'successes' of theoretical prediction and experimental validation (Noble 2002). In fact, the contradiction of predictions—whether based on hypotheses formed in thought experiments (conceptual models) or quantitative simulation (computer models)—is usually *more* instructive than their confirmation. Although confirmation increases the confidence associated with a particular concept or model, contradiction highlights shortcomings in the quality and/or quantity of data input, processing, or interpretation. This will prompt additional observation, consideration, and conceptualisation, with the potential of advancing models and insight (Kohl *et al*, 2000).

Second, despite its complexity, the heart shows pronounced *spatial regularity* in structural properties (from the tissue level right through to the arrangement of subcellular protein- and membrane-structures), and it is governed by a very high degree of *spatio-temporal coordination* of key functional behaviour (such as the spreading wave of electrical excitation that invokes every single cardiomyocyte during each heartbeat, or the highly orchestrated sequence of ionic fluxes and protein interactions that give rise to remarkably optimised pressure generation some 2.5 billion times in the healthy human heart during a life time).

Third, systems of interaction in the heart show a considerable degree of *modularity*. Basic models of cardiac electrophysiology, for example, do not need to take into account interactions with cardiac mechanics, circulation, metabolism, and so on, to predict important aspects of the interplay between ion distributions, currents, and voltage changes. As they become increasingly detailed, however, wider interactions become more and more relevant, as entities that were classically believed to be linked in a one-directional manner are subject to cross-talk and interaction. Examples include the interdependence of cardiac structure and function (Allessie *et al*, 2002), of ion channels and cell or tissue behaviour (Hodgson *et al*, 2003), or of electrophysiology and mechanics (Kohl *et al*, 2006).

Work on the virtual heart has advanced with progressively increasing complexity. The earliest cell models had just three differential equations that represented the summary kinetics of multiple 'lumped' electrical mechanisms which, by and large, had not yet been identified and were not, therefore, strictly related to individual protein channel subtypes as we know them now. Cell models today may contain 50 or more equations (Ten Tusscher *et al*, 2004), depending on the extent to which individual ion handling mechanisms are represented (e.g. through Markov models of ion channels (Clancy and Rudy, 1999)) and the complexity with which intracellular structural features are simulated (Pásek *et al*, 2008). The insertion of such models into tissue and organ models has also occurred at different levels of tissue size and complexity. Although the goal of reconstructing the whole organ with representative histo-anatomical detail is important for some applications (Burton *et al*, 2006; Plank *et al*, 2009), much insight can be gleaned from multi-cellular simulations using one-dimensional strands of cells, two-dimensional sheets, and three-dimensional simplified tissue geometries (Garny *et al*, 2005). The overall lesson from these simulations has been that theoretical models of biological behaviour are most efficient when they are *as complex as necessary, yet as simple as possible*.

Extension of principles from heart to other systems: opportunities and challenges

We do not have the space here to review the modelling of other organs and systems. Readers can find out more by accessing the websites of the *Physiome Project* (http://www.physiome.org.nz/) and the VPH (http://www.vph-noe.eu/). However, some of the approaches and principles developed for, and applied to, cardiac modelling may be transferrable to other aspects of the VPH initiative. Among the features that are already being tackled with some success by the Physiome community are several general issues related to the various types of modelling approaches and their role in the discovery process (Box 1). These principles have emerged largely from grass-roots development of model systems in the cardiac field. Although instructive, there is of course no reason to regard them as prescriptive indicators of how other VPH-related projects should be pursued.

The reason for this is straightforward and bears relevance for systems biology in general: we simply do not know what approach will eventually succeed. Researchers pursuing a systems approach can be likened more to people finding their way through unchartered territory, than to those walking a

Box 1 General principles learned from the cardiac modelling field

Conceptual Duality: the combined application of reductionist and integrationist tools and concepts lies at the very heart of successful development of a quantitative understanding of systems behaviour. The analysis of heart rhythm resulting from individual protein interactions (reductionist aspect) and their integration through feedback from the overall cell electrical activity (integration) is a good example (Noble, 2006, chapter 5).

Iteration of Theory and Practice: 'wet' experimental and 'dry' theoretical models need to be developed in continuous iteration, where new experimental (or clinical) data feed model development and/or refinement, while computational predictions are used to guide hypothesis formation and experimental design, the outcome of which is the used to validate model predictions. A good example of this approach can be found in the papers of Lei and Kohl (1998) and Cooper *et al* (2000), which used modelling to interpret experiments showing an unexpected effect of cell swelling on pacemaker frequency, leading to work using axial stretch to yield the expected result, also explained by the modelling.

Structure–Function Relationship: biological function cannot be dissociated from underlying structure. This finds a reflection in modelling, whether using 'lumped parameters' to describe general compartmentalisation (Orchard *et al*, 2009) or detailed representations of three-dimensional morphology of proteins (Young *et al*, 2001), cells (Iribe *et al*, 2009), or organs (Zhao *et al*, 2009). Increasingly, this effort benefits from standards, tools, and markup languages, such as SBML (http://sbml.org/Main_Page), CellML (http://www.cellml.org/) and FieldML (http://www.fieldml.org/).

Multi-Scale Modelling: models at different scales of structural integration are required to explore behaviour from molecule to organ or organism. This applies equally to 'wet' and 'dry' research, and involves bridging spatial scales of (at least) nine orders of magnitude (from nm to m) and temporal scales spanning 17 orders of magnitude or more (from nanoseconds for description of molecular motion, to years or decades, for longitudinal assessment of human development in norm and disease (Hunter and Borg, 2003). This requires application of 'new maths' to systems modelling, for example, scale relativity theory (Auffray and Nottale, 2008; Nottale and Auffray, 2008).

Multiplicity of Models (at each individual level): the availability of models of differing levels of complexity, even at the same level of structural integration, allows the treatment of the same biological aspect in different ways, depending on the nature of the question being addressed (for examples see Noble and Rudy, 2001). Although this is common practice in 'wet' studies, it is often questioned in 'dry' research.

Multi-dimensional Modelling: models from 0D to 3D + Time are needed to analyse parts of the system that may, in some situations, be regarded as point-sources (e.g. cell electrophysiology when looking at gross electrical behaviour such as reflected in the electrocardiogram), and in others as complex spatio-temporally structured reaction environments (such as the same cell when considering signal transduction cascades). For an Open Source environment designed to address this see Bernabeu *et al* (2009).

Multi-physics Modelling: addressing questions of varying character, from the stochastic behaviour of ion-channel-interactions to deterministic links between events, or from multiple ODE systems to soft tissue mechanics and fluid dynamics, require different implementations (e.g. finite differences, finite elements, or boundary element methods, Hodgkin–Huxley versus Markov formalisms (see e.g. Fink and Noble, 2009), diffusion reaction versus Monte Carlo approaches, etc).

Modularity of Models: a desirable but thus far ill-implemented need is the definition of model interfaces. These may range from true modularity of components, to translation tools or black-box style parameter inheritance. Likewise, model mapping is an area where much more research into theoretical understanding and practical tools is called for (Terkildsen *et al*, 2008).

High-Speed Simulation: application to real-world scenarios, in particular for time-critical emergency settings, calls for faster-than-real-time simulation. The new generation of supercomputers (e.g. the 10 petaflop machine being constructed for RIKEN in Kobe, Japan) combined with improved algorithms is expected to make this possible (Bordas *et al*, 2009).

Interactivity: interactive assessment of model behaviour is relevant for efficient implementation of 'dry' experiments, as well as for training, education, and interaction between experts from different professional backgrounds (Garny *et al*, 2009).

Box 2 Issues and Challenges

Model Curation and Preservation: the long-term preservation of data and models and the maintained ability to access digital data formats are recognised challenges of modern IT infrastructures. They also present key concerns for the VPH initiative.

Tools, Standards, Ontologies and Access: concerted efforts have been launched to facilitate the identification of suitable tools, standards, and ontologies to support model development, interaction, and access (Hucka *et al*, 2003). This is one of the declared aims of the VPH initiative and requires a willingness to

- contribute to the development of standards;
- adhere to 'good practice', once standards are agreed; and
- share and publish data, metadata, and models in a suitably annotated, re-usable format.

Patient-specific Analysis and Treatment: as non-invasive data-rich imaging methods are becoming increasingly productive in the clinical setting, the goal of incorporating patient-specific data into models for use in diagnosis, treatment planning, and prevention is beginning to become a reality. This goal is desirable for a variety of reasons, ranging from economic (it makes sense to choose treatments that are tailor-made for the patient, rather than block-buster medicines that often miss the target) to ethical (we should look forward to the day when we no longer tolerate disastrous side-effects that could be eliminated by stratifying the patient population) and scientific considerations (prevent, and if that is not possible, treat the patient—not the disease).

path that has already been mapped. Contrary to the Genome Project, we do neither know the smallest part that we need to identify (there is no elementary set of generic building blocks from which we can assemble the jigsaw), nor the extent of the overall entity (in terms of the types and number of interactions that need to be quantified). We have to determine the best approach as we try out various ideas on how to modularise, simplify, connect multiple levels, relate different aspects at the same level, and incorporate increasingly fine-grained structural and functional data. At the same time, we are also seeking mathematical approaches and computational resources that will enable models to be run in a reasonable period of time (Fink and Noble, 2009), while using user interfaces that allow utilisation by non-experts in computational modelling (Garny *et al*, 2003). These considerations are associated with a number of additional challenges that have also been experienced in the cardiac modelling field, but are far from being resolved (some examples are listed in Box 2).

Of particular relevance, in our view, is the need to establish public access to data and models derived from publicly funded work. This could be regarded as a make-or-break issue, as crucial for systems biology as was the decision by a majority of Genome Project investigators to publish and share information on annotated gene sequences, obtained through publicly funded research (rather than patenting them, which would have invoked a whole host of ethical, scientific, and socioeconomic dilemmas).

In this context, a range of ethical issues arise. We will refer briefly to just three of them here. The first is one of scientific integrity and social responsibility (and inherently underlies the drive towards public access to data and models): to serve the usual criteria of scientific scrutiny and public accountability, and to avoid 're-inventing wheels', it is required to enable others to review, (re-)use, develop, and efficiently apply prior work. From this, a second issue arises, related to professional development and career progression: as long as the prevailing approach to assessing 'academic merit' disproportionately rewards 'peer-reviewed' publications as the output of academic endeavour, compared with the (often very time consuming) development of 'peer-used' tools, sharing data and models may end up disadvantaging those professionals who generate them (by relieving them of control over and, conceivably, co-authorship in their follow-on use). A third ethical aspect is the obvious need to protect the privacy of individuals' data (a common challenge to using, re-using, and sharing human data). An international solution to these challenges may be regarded as a second make-or-break issue for systems biology and the VPH.

Conclusions

Systems biology may be interpreted as a *scientific approach* (rather than a subject or destination) that consciously combines 'reductionist' (identification and description of parts) and 'integrationist' (internal and external interactions) research, to foster our understanding of the nature and maintenance of biological entities. During the decade or so in which systems biology has become popular, it has often been interpreted as an extension of molecular biology, here to foster the understanding of subcellular regulation networks and interaction pathways, essentially equating 'system' with 'cell'. While representing an important aspect of the systems approach, there is no *a priori* reason why one level of structural or functional complexity should be more important than any other (Noble, 2008a). Work involving more complex levels of structural and functional integration is essential if systems biology is to deliver in relation to human physiology and health care. In addition to this vertical integration across multiple scales, we also need horizontal integration across boundaries such as between organ systems, and between 'wet' and 'dry' modelling. Often, the best results are obtained when theoretical work is pursued in close and continuous iteration with experimental and/or clinical investigations. An essential task for systems biology is therefore the quantitative integration of *in-silico*, *in-vitro*, and *in-vivo* research. Key make-or-break issues are the extent to which we can harvest the synergies between the multiple international efforts in the field by sharing data and models, and the question of how to address the ethical dimensions of relevant research and development in this area.

Editorial Note

This Guest Editorial was commissioned on the occasion of the *EMBL/EMBO Science & Society Conference* on 'Systems and Synthetic Biology: Scientific and Social Implications', Heidelberg, November 7–8, 2008. Additional contributions from several speakers are available on the *EMBO Reports* website (http://www.nature.com/embor).

Acknowledgements

Work in the authors' laboratory is supported by the European FP6 *BioSim* network and the *normaCOR* grant; by the European FP7 *VPH NoE*, *preDiCT*, and *EU-Heart* projects, as well as by the UK's Biotechnology and Biological Sciences Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council and The Wellcome Trust. PK is a Senior Fellow of the British Heart Foundation.

Conflict of interest

The authors declare that they have no conflict of interest.

Peter Kohl and Denis Noble
Department of Physiology, Anatomy and Genetics, Oxford, UK

References

Allessie M, Ausma J, US (2002) Electrical, contractile and structural remodeling during atrial fibrillation. *Cardiovasc Res* **54:** 230–246

Auffray C, Nottale L (2008) Scale relativity theory and integrative systems biology 1. Founding principles and scale laws. *Progress Biophys Mol Biol* **97:** 79–114

Bassingthwaighte JB, Hunter PJ, Noble D (2009) The Cardiac Physiome: perspectives for the future. *Exp Physiol* **94:** 597–605

Bernabeu MO, Bordas R, Pathmanathan P, Pitt-Francis J, Cooper J, Garny A, Gavaghan DJ, Rodriguez B, Southern JA, Whiteley JP (2009) Chaste: incorporating a novel multiscale spatial and temporal algorithm into a large scale open source library. *Philos Trans R Soc A* **367:** 1907–1930

Bordas R, Carpentieri B, Fotia G, Maggio F, Nobes R, Pitt-Francis J, Southern JA (2009) Simulation of cardiac electrophysiology on next-generation high-performance computers. *Philos Trans R Soc A* **367:** 1951–1970

Boyd CAR, Noble D (ed) (1993) *The Logic of Life*. Oxford: OUP

Brenner S, Noble D, Sejnowski T, Fields RD, Laughlin S, Berridge M, Segel L, Prank K, Dolmetsch RE (2001) Understanding complex systems: top-down, bottom-up or middle-out? In *Novartis Foundation Symposium: Complexity in Biological Information Processing*, Bock G, Goode J (eds), Vol. 239, pp 150–159. Chichester: John Wiley

Burton RAB, Plank G, Schneider JE, Grau V, Ahammer H, Keeling SJ, Lee J, Smith NP, Gavaghan D, Trayanova N, Kohl P (2006) Three-dimensional models of individual cardiac histo-anatomy: tools and challenges. *Ann NY Acad Sci* **1080:** 301–319

Clancy CE, Rudy Y (1999) Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature* **400:** 566–569

Cooper PJ, Lei M, Cheng LX, Kohl P (2000) Axial stretch increases spontaneous pacemaker activity in rabbit isolated sino-atrial node cells. *J Appl Physiol* **89:** 2099–2104

Feytmans E, Noble D, Peitsch M (2005) Genome size and numbers of biological functions. *Trans Comput Syst Biol* **1:** 44–49

Fink M, Noble D (2009) Markov models for ion channels - versatility vs. identifiability and speed. *Philos Transact A Math Phys Eng Sci* **367:** 2161–2179

Garny A, Kohl P, Noble D (2003) Cellular open resource (COR): a public CellML based environment for modelling biological function. *Int J Bifurcat Chaos* **13:** 3579–3590

Garny A, Noble D, Hunter PJ, Kohl P (2009) Cellular open resource (COR): current status and future directions. *Philos Trans R Soc A* **367:** 1885–1905

Garny A, Noble D, Kohl P (2005) Dimensionality in cardiac modelling. *Prog Biophys Mol Biol* **87:** 47–66

Guyton AC, Coleman TG, Granger HJ (1972) Circulation: overall regulation. *Ann Rev Physiol* **34:** 13–46

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* **117:** 500–544

Hodgson DM, Zingman LV, Kane GC, Perez-Terzic C, Bienengraeber M, Ozcan C, Gumina RJ, Pucar D, O'Coclain F, Mann DL, Alekseev AE, Terzic A (2003) Cellular remodeling in heart failure disrupts KATP channel-dependent stress tolerance. *EMBO J* **22:** 1732–1742

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **4:** 524–531

Hunter PJ, Borg TK (2003) Integration from proteins to organs: the physiome project. *Nat Rev Mol Cell Biol* **4:** 237–243

Hunter PJ, Kohl P, Noble D (2001) Integrative models of the heart: achievements and limitations. *Philos Trans R Soc A* **359:** 1049054-1

Iribe G, Ward CW, Camelliti P, Bollensdorff C, Mason F, Burton RA, Garny A, Morphew MK, Hoenger A, Lederer WJ, Kohl P (2009) Axial stretch of rat single ventricular cardiomyocytes causes an acute and transient increase in Ca2 + spark rate. *Circ Res* **104:** 787–895

Kohl P, Bollensdorf C, Garny A (2006) Effects of mechanosensitive ion channels on ventricular electrophysiology: experimental and theoretical models. *Exp Physiol* **91:** 307–321

Kohl P, Noble D, Winslow R, Hunter PJ (2000) Computational modelling of biological systems: tools and visions. *Philos Trans R Soc* **A 358:** 579–610

Lei M, Kohl P (1998) Swelling-induced decrease in spontaneous pacemaker activity of rabbit isolated sino-atrial node cells. *Acta Physiol Scand* **164:** 1–12

Noble D (1962) A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pacemaker potentials. *J Physiol* **160:** 317–352

Noble D (2002) Modelling the heart: insights, failures and progress. *BioEssays* **24:** 1155–1163

Noble D (2006) *The Music of Life*. Oxford: OUP

Noble D (2008a) Claude Bernard, the first Systems Biologist, and the future of Physiology. *Exp Physiol* **93:** 16–26

Noble D (2008b) Computational models of the heart and their use in assessing the actions of drugs. *J Pharmacol Sci* **107:** 107–117

Noble D (2008c) Genes and causation. *Philos Trans R Soc* **A 366:** 3001–3015

Noble D, Rudy Y (2001) Models of cardiac ventricular action potentials: iterative interaction between experiment and simulation. *Philos Trans R Soc A* **359:** 1127–1142

Nottale L, Auffray C (2008) Scale relativity and integrative systems biology 2. Macroscopic quantum-type mechanics. *Prog Biophys Mol Biol* **97:** 115–157

Orchard CH, Pásek M, Brette F (2009) The role of mammalian cardiac t-tubules in excitation-contraction coupling: experimental and computational approaches. *Exp Physiol* **94:** 509–519

Pásek M, Brette F, Nelson A, Pearce C, Qaiser A, Christe G, Orchard CH (2008) Quantification of t-tubule area and protein distribution in rat cardiac ventricular myocytes. *Prog Biophys Mol Biol* **96:** 244–257

Plank G, Burton RAB, Hales P, Bishop M, Mansoori T, Bernabeu M, Garny A, Prassl AJ, Bollensdorf C, Mason F, Mahmood F, Rodriguez B, Grau V, Schneider JE, Gavaghan D, Kohl P (2009) Generation of histo-anatomically representative models of the individual heart: tools and application. *Philoso Trans R Soc A* **367:** 2257–2292

Ten Tusscher KHWJ, Noble D, Noble PJ, Panfilov AV (2004) A model of the human ventricular myocyte. *Am J Physiol* **286:** H1573–H1589

Terkildsen JR, Niederer S, Crampin E, Hunter PJ, Smith NP (2008) Using physiome standards to couple cellular functions for rat cardiac excitation-contraction. *Exp Physiol* **93:** 919–929

von Bertalanffy L (1968) *General System Theory*. New York: George Braziller, Inc

Young HS, Jones LR, Stokes D (2001) Locating phospholamban in co-crystals with Ca(2 +)-ATPase by cryoelectron microscopy. *Biophys J* **81:** 884–894

Zhao J, Trew ML, Legrice IJ, Smaill BH, Pullan AJ (2009) A tissue-specific model of reentry in the right atrial appendage. *J Cardiovasc Electrophysiol* **20:** 675–684

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2010) **368**, 1125–1139
doi:10.1098/rsta.2009.0245

Review

Biophysics and systems biology

By Denis Noble*

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

Biophysics at the systems level, as distinct from molecular biophysics, acquired its most famous paradigm in the work of Hodgkin and Huxley, who integrated their equations for the nerve impulse in 1952. Their approach has since been extended to other organs of the body, notably including the heart. The modern field of computational biology has expanded rapidly during the first decade of the twenty-first century and, through its contribution to what is now called systems biology, it is set to revise many of the fundamental principles of biology, including the relations between genotypes and phenotypes. Evolutionary theory, in particular, will require re-assessment. To succeed in this, computational and systems biology will need to develop the theoretical framework required to deal with multilevel interactions. While computational power is necessary, and is forthcoming, it is not sufficient. We will also require mathematical insight, perhaps of a nature we have not yet identified. This article is therefore also a challenge to mathematicians to develop such insights.

Keywords: cell biophysics; systems biology; computational biology; mathematical biology

1. Introduction: the origins of biophysics and systems biology

As a young PhD student at University College London, I witnessed the celebrations of the 300th anniversary of the Royal Society in 1960. As the magnificent procession of red-gowned Fellows of the Royal Society (FRS) paraded into the Royal Albert Hall, two black gowns suddenly appeared. They were worn by Alan Hodgkin and Andrew Huxley. The founders of the field of cellular biophysics, with their ground-breaking mathematical reconstruction of the nerve impulse (Hodgkin & Huxley 1952), were simply Mr Hodgkin and Mr Huxley—neither had submitted a thesis for a PhD. With 'FRS' to their names, they hardly needed to! A year later, Alan Hodgkin examined my PhD thesis, which applied their ideas to reconstructing the electrical functioning of the heart (Noble 1960, 1962), and 3 years later we were celebrating their Nobel Prize.

It is highly appropriate to recall these events in a volume to celebrate the 350th anniversary, but they also remind us that the field that is now called systems biology has important historical roots. Hodgkin and Huxley themselves were not

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

One contribution of 17 to a Theme Issue 'Personal perspectives in the physical sciences for the Royal Society's 350th anniversary'.

the first. I would nominate Claude Bernard as the first systems biologist (Noble 2008*a*), since in the middle of the nineteenth century he formulated the systems principle of control of the internal environment (Bernard 1865). This is well known and is widely recognized as the homeostatic basis of modern physiological science. It is much less well known that Bernard also presaged the development of mathematical biology when he wrote 'this application of mathematics to natural phenomena is the aim of all science, because the expression of the laws of phenomena should always be mathematical.'[1] Other historical roots can be found in the work of Harvey (Auffray & Noble 2009) and Mendel (Auffray 2005). Despite these strong historical roots, however, the field did not flourish in the second half of the twentieth century. Soon after Hodgkin and Huxley's achievement it was to be swept aside as molecular biology took the centre stage.

2. The achievements and problems of molecular biology

Physicists and mathematicians contributed greatly to the spectacular growth of molecular biology. The double-helical structure of DNA was discovered in the Cavendish laboratory in Cambridge (Watson & Crick 1953*a*,*b*) and in the biophysics laboratory at King's College London (Franklin & Gosling 1953*a*,*b*; Wilkins *et al.* 1953), while some of the seminal ideas of molecular biology were first developed by Schrödinger (1944). In addition to correctly predicting that the genetic material would be found to be an aperiodic crystal, his book, *What is Life?,* followed a proposal by Max Delbrück (see Dronamrajua 1999) that was to prove fundamental in the twentieth century interpretation of molecular biology. This was that physics and biology are essentially different disciplines in that while physics is about the emergence of order from disorder, such as the ordered global behaviour of a gas from the disordered Brownian motion of the individual molecules, biology dealt with order even at the molecular level. The paradigm for this view was the effects of mutations of the genetic material. Even a single switch from one nucleotide to another, corresponding to a single amino acid change in the protein for which the DNA sequence acts as a template, can have dramatic effects on the phenotype at higher levels. A good example in the case of the heart is that of the various sodium channel mutations that can cause arrhythmia (Clancy & Rudy 1999), and there are excellent examples in the processes of embryonic development (Davidson 2006).

The attribution of control to the DNA was strongly reinforced by Monod and Jacob (Jacob *et al.* 1960), who interpreted their work as evidence for the existence of a 'genetic program', an analogy explicitly based on comparison with an electronic computer: 'The programme is a model borrowed from electronic computers. It equates the genetic material with the magnetic tape of a computer' (Jacob 1982), while the rest of the organism, particularly the fertilized egg cell, could be compared with the computer itself. Specific instructions at the level of DNA could then be seen to 'program' or control the development and behaviour of the organism. These ideas married well with the gene-centred theories of evolution and the metaphor of 'selfish' genes (Dawkins 1976, 1982, 2006), which relegated the organism to the role of a disposable transient carrier of its DNA.

[1]Cette application des mathématiques aux phénomènes naturels est le but de toute science, parce que l'expression de la loi des phénomènes doit toujours être mathématique.

It is not surprising therefore that the peak of the achievement of molecular biology, the sequencing of the complete human genome, was widely signalled as finally reading the 'book of life'. However, the main architects of that project are much more circumspect: 'One of the most profound discoveries I have made in all my research is that you cannot define a human life or any life based on DNA alone…'. Why? Because 'An organism's environment is ultimately as unique as its genetic code' (Venter 2007). Sulston is also cautious: 'The complexity of control, overlaid by the unique experience of each individual, means that we must continue to treat every human as unique and special, and not imagine that we can predict the course of a human life other than in broad terms' (Sulston & Ferry 2002). So also is Sydney Brenner, whose work has contributed so much to the field: 'I believe very strongly that the fundamental unit, the correct level of abstraction, is the cell and not the genome' (lecture at Columbia University 2003).

I have briefly summarized some of these aspects of the development of molecular biology because, in fulfilling my brief to look into the crystal ball and give my own perspective on where my subject is heading in the next 50 years, I am going to turn some of the concepts derived from the successes of molecular biology upside down. I suggest that the next stage in the development of biological science will be revolutionary in its conceptual foundations (Shapiro 2005; see also Saks *et al.* 2009) and strongly mathematical in its methods. I also see this as the fulfilment of Claude Bernard's dream of the role of mathematics in his discipline, a dream that certainly could not be achieved in his lifetime.

3. Digital, analogue and stochastic genetic causes

Since the C, G, A, T sequences can be represented digitally (two bits are sufficient to represent four different entities, so the three billion base pairs could be represented by six billion bits), the idea of a determinate genetic program in the DNA, controlling the development and functioning of the organism, rather like the digital code of a computer program, was seductive, but for it to be correct, three conditions need to be satisfied. The first is that the relevant program logic should actually be found in the DNA sequences. The second is that this should control the production of proteins. The third is that this should be a determinate process. It is now known that none of these conditions are fulfilled. Molecular biology itself has revealed these deficiencies in at least six different ways.

(i) The C, G, A, T sequences of nucleotides in the genome do not themselves form a program as normally understood, with complete logic (i.e. one that could be subjected to syntactic analysis) of a kind that could separately run a computer. We cannot therefore predict life using these sequences alone. Instead, the sequences form a large set of templates that the cell uses to make specific proteins, and a smaller bank of switches, the regulatory genes, forming about 10 per cent of human genes, and the regulatory sites on which the regulatory proteins and other molecules act. Impressive switching circuits can be drawn to represent these (Levine & Davidson 2005). But they require much more than the DNA sequences themselves to operate since those switches depend on input from the rest of the organism, and from the environment. Organisms are interaction machines, not Turing machines (Shapiro 2005; Neuman 2008; Noble 2008*c*). There is therefore no

computer into which we could insert the DNA sequences to generate life, other than life itself. Far from being just a transient vehicle, the organism itself contains the key to interpreting its DNA, and so to give it meaning. I will return later to this question (see §7).

(ii) In higher organisms, the sequences are broken into sometimes widely dispersed fragments, the exons, which can be combined in different ways to form templates for many different proteins. Something else must then determine which combination is used, which protein is formed and at which time. The DNA sequences therefore better resemble a database on which the system draws rather than a logical program of instructions (Atlan & Koppel 1990; Shapiro 2005; Noble 2006). For that we must look elsewhere, if indeed it exists at all. The dispersed nature of the exons and the combinatorial way in which they are used also challenges the concept of genes as discrete DNA sequences (Keller 2000*a*; Pearson 2006; Scherrer & Jost 2007).

(iii) What determines which proteins are made and in what quantity is not the DNA alone. Different cells and tissues use precisely the same DNA to produce widely different patterns of gene expression. This is what makes a heart cell different from, say, a bone cell or a pancreatic cell. These instructions come from the cells and tissues themselves, in the form of varying levels of transcription factors and epigenetic marks (Bird 2007) that are specific to the different types of cell. These processes are robust and inherited. Differentiated heart cells always form new heart cells as the heart develops, not new bone cells. They would need to be 'de-differentiated' to form multipotent stem cells in order to give rise to a different differentiated cell. This should not surprise us. Some kinds of cellular inheritance, perhaps starting with the ability of a lipid membrane-enclosed globule to divide, almost certainly predated genome inheritance (Maynard Smith & Szathmáry 1995).

(iv) The resulting patterns of gene expression are not only widely variable from one tissue to another, they themselves are not digital. The expression levels vary continuously in a way that is better described as an analogue. Since we must include these analogue levels in any description of how the process works, any 'program' we might identify is not based on digital coding alone. It is significant therefore that the inclusion of analogue processing is seen by some computer scientists as an important way in which a system can perform beyond the Turing limits (Siegelmann 1995, 1998, 1999). Organisms are, at the least, 'super-Turing' machines in this sense.

(v) Gene expression is a stochastic process (Kaern *et al.* 2005). Even within the same tissue, there are large variations in gene expression levels in different cells. Such stochasticity is incompatible with the operation of a determinate Turing machine (Kupiec 2008; Neuman 2008).

(vi) Finally, there is continuous interaction between DNA and its environment. As Barbara McClintock put it in her Nobel prize lecture (1983) for her work on 'jumping genes', the genome is better viewed as 'a highly sensitive organ of the cell' that can be reorganized in response to challenges (Keller 1983). We now also understand the extent to which organisms can swap DNA between each other, particularly in the world of micro-organisms (Goldenfeld & Woese 2007).

Another way to express the significance of these developments in molecular biology is to say that not much is left of the so-called 'central dogma of biology' (see Shapiro (2009) for more details) other than that part of Crick's original statement of it that is correct, which is that while DNA is a template for amino acid sequences in proteins, proteins do not form a template from which DNA can be produced by a reverse version of the DNA→protein transcription process. But in the extended sense in which it is frequently used in a neo-Darwinist context, as forbidding the passage of information from the organism and environment to DNA, the 'dogma' is seriously incorrect. Information is continually flowing in the opposite direction. I will return later to the significance of this fact for neo-Darwinism itself.

To these facts we must add a few more before we reassess the comparison between physics and biology.

(vii) Many genetic changes, either knockouts or mutations, appear not to have significant phenotypic effects; or rather they have effects that are subtle, often revealed only when the organism is under stress. For example, complete deletion of genes in yeast has no obvious phenotypic effect in 80 per cent of cases. Yet, 97 per cent have an effect on growth during stress (Hillenmeyer *et al.* 2008). The reason is that changes at the level of the genome are frequently buffered, i.e. alternative processes kick in at lower levels (such as gene–protein networks) to ensure continued functionality at higher levels (such as cells, tissues and organs). And even when a phenotype change does occur there is no guarantee that its magnitude reveals the full quantitative contribution of that particular gene since the magnitude of the effect may also be buffered. This is a problem I have recently referred to as the 'genetic differential effect problem' (Noble 2008*c*) and it has of course been known for many years. There is nothing new about the existence of the problem. What is new is that gene knockouts have revealed how extensive the problem is. Moreover, there is a possible solution to the problem to which I will return later.

(viii) The existence of stochastic gene expression allows some form of selection operating at the level of tissues and organs (Laforge *et al.* 2004; Kaern *et al.* 2005; Kupiec 2008, 2009). In fact, such selection may be a prerequisite of successful living systems which can use only those variations that are fit for purpose. As Kupiec has noted, Darwinian selection could also be very effective within the individual organism, as well as between organisms.

(ix) Not only is gene expression stochastic, the products of gene expression, the proteins, each have many interactions (at least dozens) with other elements in the organism. Proteins are not as highly specific as was once anticipated. Bray (Bray & Lay 1994; Bray 2009) has highlighted the role of multiple interactions in comparing the evolution of protein networks with that of neural networks.

4. The multifactorial nature of biological functions

So, while it is true to say that changes at the molecular level can *sometimes* have large effects at the higher phenotype levels, these effects are frequently buffered. Even the sodium channel mutations I referred to earlier do not, by themselves,

trigger cardiac arrhythmia. The picture that emerges is that of a multifactorial system. Biology, it turns out, must also create order from stochastic processes at the lower level (Auffray *et al.* 2003). Physics and biology do not after all differ in quite the way that Schrödinger thought. This is a point that has been forcibly argued recently by Kupiec (2008, 2009). There is absolutely no way in which biological systems could be immune from the stochasticity that is inherent in Brownian motion itself. It is essential therefore that biological theory, like physical theory, should take this into account.

The systems approach has already pointed the way to achieve this. The massively combinatorial nature of biological interactions could have evolved precisely to overcome stochastic effects at the molecular level (Shapiro 2009). As Bray (2009) notes, protein networks have many features in common with the neural networks developed by artificial intelligence researchers. They can 'evolve' effective behaviour strategies from networks initialized with purely random connections, and once they have 'evolved' they show a high degree of tolerance when individual components are 'knocked out'. There is then what Bray calls 'graceful degradation', which can take various forms (not necessarily requiring random connectivity). This provides an insight into the nature of the robustness of biological systems. Far from stochasticity being a problem, it is actually an advantage as the system evolves. 'Graceful degradation' is also a good description of what happens in knockout organisms. All may appear to be well when the organism is well-fed and protected. The deficiency may reveal itself only when the conditions are hostile.

I suspect that more relevant insights will come from analysis of such artificial networks and even more so from the modelling of real biological networks. Note that such networks do not require a separate 'program' to operate. The learning process in the case of artificial networks, and evolutionary interaction with the environment in the case of biological networks, is the 'programming' of the system. So, if we still wish to use the program metaphor, it is important to recognize that the program is the system itself (Noble 2008*c*). The plant geneticist Enrico Coen expressed this point well when he wrote 'Organisms are not simply manufactured according to a set of instructions. There is no easy way to separate instructions from the process of carrying them out, to distinguish plan from execution' (Coen 1999). This is another version of the points made earlier about the limitations of regarding the DNA sequences as a program.

5. The multilevel nature of biological functions

This takes me to the question of multilevel analysis. Organisms are not simply protein soups. Biological functions are integrated at many different levels. Thus, pacemaker rhythm in the heart is integrated at the level of the cell. There is no oscillator at the biochemical level of subcellular protein networks (Noble 2006). Tempting though it may be to think so, there is therefore no 'gene for' pacemaker rhythm. A set of genes, or more correctly the proteins formed from their templates, is involved, together with the cellular architecture—and which set we choose to represent depends on the nature of the questions we are asking. But that does not prevent us from building computer programs that mimic pacemaker rhythm. Simulation of cardiac activity has been developed over

a period of nearly five decades and is now sufficiently highly developed that it can be used in the pharmaceutical industry to clarify the actions of drugs (Noble 2008*b*).

Does not the fact that we can succeed in doing this prove that, after all, there are genetic programs? Well no, for two reasons. First the logic represented by such computer simulation programs is certainly not to be found simply in the DNA sequences. The programs are representations of the processes involved at *all* the relevant biological levels, right up to and including the intricate architecture of the cell itself. And when even higher levels are modelled, the structural biology included is that of tissues or the entire organ (Hunter *et al.* 2003; Garny *et al.* 2005). In the case of the heart, the three-dimensional imaging technology to achieve this has now advanced to paracellular or even subcellular levels (Plank *et al.* 2009).

Second, reflecting Coen's point above, the processes represented in our modelling programs *are* the functionality itself. To the extent that the program succeeds in reproducing the behaviour of the biological system it reveals the *processes* involved, not a separate set of *instructions*.

Multilevel simulation will be a major development in biology as the project known as the Human Physiome Project develops. Recent issues of this journal have been devoted to one of its components, the Virtual Physiological Human (VPH) project (Clapworthy *et al.* 2008; Fenner *et al.* 2008) and some of the achievements and future challenges of the Physiome Project (Bassingthwaighte *et al.* 2009) and its relation to systems biology (Kohl & Noble 2009) have recently been reviewed.

6. A theory of biological relativity?

One of the major theoretical outcomes of multilevel modelling is that causation in biological systems runs in both directions: upwards from the genome and downwards from all other levels.[2] There are feedforward and feedback loops between the different levels. Developing the mathematical and computational tools to deal with these multiple causation loops is itself a major challenge. The mathematics that naturally suits one level may be very different from that for another level. Connecting levels is not therefore trivial. Nor are the problems simply mathematical and computational. They also require biological insight to determine how much detail at one level is relevant to functionality at other levels. These problems are now exercising the minds of interdisciplinary teams of researchers involved in the Physiome Project and they offer great opportunities for physical and mathematical scientists in the future. They have also led some physicists and biologists to develop what might be called theories of biological relativity. My own version of this idea is that, in multilevel systems, there is no privileged level of causation (Noble 2008*a*,*c*). Others have also pointed out that such a principle need not be restricted to biological systems. It could become a

[2]'Upwards' and 'downwards' in this context are metaphorical. A more neutral terminology would refer to different (larger and smaller) scales. But the concept of level is strongly entrenched in biological science so I have continued to use it here. There is also possible confusion with 'scale' as used in scale relativity, though I believe that one of the key questions for the future is that of relating the ideas of scale relativity to multilevel systems biology.

general theory of relativity of levels. Such a theory, called scale relativity (Nottale 1993, 2000), already exists in physics and its possible applications to biological systems have been the subject of major recent reviews (Auffray & Nottale 2008; Nottale & Auffray 2008).

I will not review these theories in detail here. I wish rather to draw attention to a related general question. Is multilevel analysis simply a matter of including downward causation (Noble 2006)? And what exactly do we mean by that term?

In my own field the paradigm example originated with Alan Hodgkin. The proteins that form ion channels in excitable cells generate electric current that charges or discharges the cell capacitance. That can be seen as upward causation. But the electrical potential of the cell also controls the gating of the ion channel proteins. This downward causation closes the loop of the 'Hodgkin cycle'.

Is downward causation always discrete feedback or feedforward? The answer is no and the basis for that answer is profound, forming one of the reasons why I think that systems biology is revolutionary. A feedback loop can be closed. Feedback loops could exist between the levels of an organism, while the organism itself could still be modelled as a closed system. Yet, we know that organisms are not closed systems. Firstly they exchange energy and matter with the environment, including particularly other organisms whose existence forms a major part of the selection pressure. That is well recognized as a reason for regarding organisms as open systems. But there are other reasons also. I think that the best way to explain that is mathematical.

We model many biological processes as systems of differential equations. These equations describe the rates at which those processes occur. The number of such equations depends on the kind of question we are asking. At a cellular or subcellular (protein network) level, there may be a few dozen equations for the protein and other chemical entities involved. When we include structural details at the tissue or organ level, we may be dealing with millions of equations. Whatever the number, there is an inescapable requirement before we can begin to solve the equations. We must know or make plausible guesses for the initial and boundary conditions. They are not set by the differential equations themselves. These conditions restrain the solutions that are possible. In fact, beyond a certain level of complexity, the more interesting question becomes the explanation of that restraining set of conditions, not just the behaviour of the system, since the restraints may completely change the behaviour of the system. A restraint, therefore, is not necessarily a feedback. Restraints can be simply the background set of conditions within which the system operates, i.e. its environment. Through these interactions organisms can adapt to many different conditions. Their robustness in doing so distinguishes them from complex nonlinear systems that are highly sensitive to initial conditions or which end up unable to escape attractors.

7. 'Genetic programs'

This is a suitable point at which to return to the question of 'genetic programs'. As we have seen, DNA sequences act as templates for proteins and as switches for turning genes on and off when they are in an organism, starting with the

fertilized egg cell and maternal environment in the case of higher animals. A possible objection to my conclusion that the DNA sequences are better viewed as a database rather than as a program is that all programs require a computer to implement them. It was part of Monod and Jacob's idea that, if DNA is the program, the organism is equivalent to the computer. Programs also do nothing outside the context of a computer. Could we somehow update this approach to save the 'program' metaphor? It is so ingrained into modern thought, among laypeople as well as most scientists, that it may now be difficult to convince people to abandon it. It is therefore worth spelling out, once again, what the difficulties are.

DNA sequences alone are not capable of being parsed as the complete logic of a program. Whenever we talk of a genetic program we must also include steps that involve the rest of the organism (e.g. my discussion of the 'circadian rhythm' program in Noble (2006, pp. 69–73), and this is certainly true for the analysis of cardiac rhythm (Noble 2006, pp. 56–65)). Much of the logic of living systems lies beyond DNA. To save the program metaphor therefore we would have to say that the 'program' is distributed between the tape and the machine. This would, incidentally, explain an important fact. Virtually all attempts at cross-species cloning fail to develop to the adult (Chung *et al.* 2009). A possible explanation is that the egg cell information is too specific (Chen *et al.* 2006). In fact, in the only case so far, that of a carp nucleus and goldfish egg, the egg cytoplasm clearly influences the phenotype (Sun *et al.* 2005). Strathmann (1993) also refers to the influence of the egg cytoplasm on gene expression during early development as one of the impediments to hybridization in an evolutionary context. There is no good reason why cells themselves should have ceased to evolve once genomes arose. But if we need a specific (special purpose) 'computer' for each 'program', the program concept loses much of its attraction.

The way to save the genetic program idea would therefore be to abandon the identification of genes with specific sequences of DNA alone and return to the original idea of genes as the causes of particular phenotypes (Kitcher 1982; Mayr 1982; Dupré 1993; Pichot 1999; Keller 2000*b*; Noble 2008*c*) by including other relevant processes in the organism. The problem with this approach is that the closer we get to characterizing the 'program' for a particular phenotype, the more it looks like the functionality itself. Thus, the process of cardiac rhythm can be represented as such a 'program' (indeed, modellers write computer programs to reproduce the process), but it is not a sequence of instructions separate from the functionality itself. This is another way to understand the quotation from Coen referred to earlier. The clear distinction between the replicator and the vehicle disappears and, with it, a fundamental aspect of the 'selfish gene' view.

If we do wish to retain the idea of a program, for example in talking about embryonic development where the concept of a 'developmental program' has its best applications (Keller 2000*a*), it might be better to think in the same terms in which we talk of neural nets being programmed. They are programmed by the initial setting up of their connections and then by the learning process, the set of restraints that allows them to 'home in' to a particular functionality. Those open-ended restraints are as much a part of the 'program' as the initial setting up of the system. The analogy with organisms as interaction machines is obvious. I am not proposing that organisms function as neural nets; only that the example

of neural nets expands our concept of the word 'program' in a relevant way. The program is a distributed one (Siegelmann 1998) involving much more than DNA sequences, and is therefore far removed from Monod and Jacob's original concept of a genetic program.

8. Systems biology and evolution

Where do the restraints come from in biological systems? Clearly, the immediate environment of the system is one source of restraint. Proteins are restrained by the cellular architecture (where they are found in or between the membrane and filament systems), cells are restrained by the tissues and organs they find themselves in (by the structure of the tissues and organs and by the intercellular signalling) and all levels are restrained by the external environment. Even these restraints though would not exhaust the list. Organisms are also a product of their evolutionary history, i.e. the interactions with past environments. These restraints are stored in two forms of inheritance—DNA and cellular. The DNA sequences restrict which amino acid sequences can be present in proteins, while the inherited cellular architecture restricts their locations, movements and reactions.

This is one of the reasons why systems biology cannot be restricted to the analysis of protein and gene circuits. The structural information is also crucial. Much of its evolution may have been independent of the cell's own DNA since the early evolution of the eukaryotic cell involved many forms of symbiosis. The best known example is the mitochondria, which are now accepted to have originally been invading (or should we say 'captured'?) bacteria, as were chloroplasts (Cavalier-Smith 2000, 2004). They even retain some of the original DNA, though some also migrated to the nucleus. There are other examples of symbiosis (Margulis 1981; Margulis & Sagan 2002; Williamson 2003, 2006; Williamson & Vickers 2007). Cooperativity may have been quite as important as competition in evolution (see also Goldenfeld & Woese 2007).

Cavalier-Smith has described some of these inherited features of animal and plant cells as the 'membranome', an important concept since lipids are not formed from DNA templates. An organism needs to inherit the membranome, which it does of course—it comes complete with the fertilized egg cell—yet another reason why it does not make sense to describe the organism as merely a vehicle for DNA. As I have argued elsewhere (Noble 2008*c*), the relative contributions of DNA and non-DNA inheritance are difficult to estimate (one is largely digital and so easy to calculate, whereas the other is analogue and hard to calculate), but the non-DNA inheritance is very substantial. It also contains many historical restraints of evolution.

This is the point at which I should attempt to explain the neo-Darwinian model and the modern synthesis and what is wrong with them from a systems viewpoint.

Neo-Darwinism brings together natural selection and nineteenth century genetics, while the modern synthesis (Huxley 1942) fuses Darwinism with twentieth century genetics. 'Neo-Darwinism' is the term often used for both of these syntheses. Darwin knew nothing of Mendel's work on genetics. Moreover, he also accepted the idea of the inheritance of acquired characteristics, as did Lamarck (Lamarck 1809; Corsi 2001), who is incorrectly represented in many

texts as inventing the idea. Darwin's disagreements with Lamarck were not over the mechanisms of inheritance. Both were ignorant of those mechanisms. Their disagreement was more over the question of whether evolution had a direction or whether variation was random. Historically, we would do better to recognize Lamarck as the inventor of the term 'biology' as a separate science, and as championing the idea that species change (transformationism). Darwin can then be seen as discovering one of the mechanisms in his theory of natural selection, involved not only in transformations but also in the origin of species.

The problem with both revisions of Darwinism is that they involve a version of genetics that we need to revise. This version was one in which the central dogma of biology was taken to mean that the genetic material is never modified by the rest of the organism and the environment. Francis Crick's original statements of the 'central dogma of molecular biology' (Crick 1958, 1970) do not in fact make such a strong claim. He stated a more limited chemical fact: that DNA sequences are used as templates to make proteins, but proteins are not used as reverse templates to make DNA. So, even if its proteins were to become modified during the lifetime of an individual, that modification cannot be inherited. The 'dogma' was then interpreted by many biologists to mean that information flows only one way. As we have seen, it does not. The *quantities* of proteins synthesized count as relevant information just as much as their amino acid sequences. But those quantities are most certainly dependent on signals from the rest of the system through the levels of transcription factors (including proteins and RNA) and the epigenetic marking of DNA itself and of the histone tails. All of this is open to the rest of the organism and to the environment to degrees we have yet to fully determine.

I will give just one example here to illustrate the potential significance of this openness. More examples can be found elsewhere (Jablonka & Lamb 1995, 2005). Neuroscientists have recently studied the epigenetic factors involved in maternal grooming behaviour in colonies of rats. Grooming depends on the environment. Colonies that are safe groom their young a lot. Colonies that are fighting off predators do not. This behaviour is inherited. The mechanisms are a fascinating example of epigenetic effects. The genome in the hippocampal region of the brain is epigenetically marked by the grooming behaviour and this predisposes the young to show that behaviour (Weaver *et al.* 2004, 2007). This is an important development, but as Weaver himself points out (Weaver 2009) it is currently restricted to one gene and one region of the brain. That underlines the importance of further research in this area. The implications of this form of epigenetic influence, however, are profound since it can transmit patterns of epigenetic marking through the generations even though they are not transmitted via the germline. This constitutes another form of inheritance of acquired characteristics to add to those reviewed by Jablonka and Lamb.

There is a tendency to dismiss such challenges to extensions of the central dogma as merely examples of cultural evolution. They seem to show rather that the boundaries between the different evolutionary processes are fuzzy. Once such interactions between behaviour and epigenetics are established and transmitted through the generations they can favour genetic combinations that lock them into the genome (Jablonka & Lamb 2005, pp. 260–270). This mechanism was originally

described by Waddington (1942, 1957, 1959; Bard 2008), who demonstrated that, in fruitflies, just 14 generations of induced phenotype change could be assimilated into the genome. Mutations and genetic recombinations themselves are not random (Shapiro 2005). Moreover, they do not occur in a random context. They occur in the context of all the restraints exerted on the organism, including those of the environment. In such a process, it is the phenotype, not individual genes, that are the targets of selection (Keller 1999). Central building blocks of the neo-Darwinian synthesis are now known to be incompatible with the most recent discoveries in molecular biology.

9. Reverse engineering in systems biology

I referred earlier to the 'genetic differential effect problem'. In a previous article in this journal I have proposed that computational systems biology could provide a solution (Noble 2008*c*). The idea is basically simple. If our understanding and simulations are good enough they should include the robustness of biological systems, including their resistance to damage from mutations and knockouts. Moreover, if the models include representations of specific gene products (i.e. they extend down to the protein level) then it should be possible to reverse engineer to arrive at *quantitative* estimates of the contribution of each gene product to the functionality represented. That may be possible even if the system completely buffers the mutation or knockout so that no effect is observed in the phenotype. I give an example of this in the previous article from work on the heart (Noble 2008*c*). However, I would readily agree that, in its present state of development, computational systems biology is a long way from being able to do this in general. But it is worth bearing this in mind as an important long-term goal.

Work in the author's laboratory is funded by the EU (Framework 6 and Framework 7), The British Heart Foundation, EPSRC and BBSRC. I acknowledge valuable criticisms from Charles Auffray, Jonathan Bard, Evelyn Fox Keller, Peter Kohl, Jean-Jacques Kupiec, Lynn Margulis, Laurent Nottale, James Shapiro, Hava Siegelmann, Eric Werner and Michael Yudkin.

References

Atlan, H. & Koppel, M. 1990 The cellular computer DNA: program or data? *Bull. Math. Biol.* **52**, 335–348.

Auffray, C. 2005 Aux sources de la biologie des systèmes et de la génétique: la pertinence des expérimentations de Gregor Mendel sur le développement des plantes hybrides (2e volet). *L'Observatoire de la génétique* **21**.

Auffray, C. & Noble, D. 2009 Conceptual and experimental origins of integrative systems biology in William Harvey's masterpiece on the movement of the heart and the blood in animals. *Int. J. Mol. Sci.* **10**, 1658–1669. (doi:10.3390/ijms10041658)

Auffray, C. & Nottale, L. 2008 Scale relativity theory and integrative systems biology. I. Founding principles and scale laws. *Prog. Biophys. Mol. Biol.* **97**, 79–114. (doi:10.1016/j.pbiomolbio.2007.09.002)

Auffray, C., Imbeaud, S., Roux-Rouquie, M. & Hood, L. 2003 Self-organized living systems: conjunction of a stable organization with chaotic fluctuations in biological space-time. *Phil. Trans. R. Soc. Lond. A* **361**, 1125–1139. (doi:10.1098/rsta.2003.1188)

Bard, J. B. L. 2008 Waddington's legacy to developmental and theoretical biology. *Biol. Theory* **3**, 188–197. (doi:10.1162/biot.2008.3.3.188)

Bassingthwaighte, J. B., Hunter, P. J. & Noble, D. 2009 The cardiac physiome: perspectives for the future. *Exp. Physiol.* **94**, 597–605. (doi:10.1113/expphysiol.2008.044099)

Bernard, C. 1865 *Introduction à l'étude de la médecine expérimentale*. Paris, France: J. B. Baillière. (Reprinted by Flammarion 1984.)

Bird, A. 2007 Perceptions of epigenetics. *Nature* **447**, 396–398. (doi:10.1038/nature05913)

Bray, D. 2009 *Wetware. A computer in every cell*. New Haven, CT: Yale University Press.

Bray, D. & Lay, S. 1994 Computer simulated evolution of a network of cell-signalling molecules. *Biophys. J.* **66**, 972–977. (doi:10.1016/S0006-3495(94)80878-1)

Cavalier-Smith, T. 2000 Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**, 174–182. (doi:10.1016/S1360-1385(00)01598-3)

Cavalier-Smith, T. 2004 The membranome and membrane heredity in development and evolution. In *Organelles, genomes and eukaryote phylogeny: an evolutionary synthesis in the age of genomics* (eds R. P. Hirt & D. S. Horner), pp. 335–351. Boca Raton, FL: CRC Press.

Chen, T., Zhang, Y.-L., Jiang, Y., Liu, J.-H., Schatten, H., Chen, D.-Y. & Sun, Q.-Y. 2006 Interspecies nuclear transfer reveals that demethylation of specific repetitive sequences is determined by recipient ooplasm but not by donor intrinsic property in cloned embryos. *Mol. Reprod. Dev.* **73**, 313–317. (doi:10.1002/mrd.20421)

Chung, Y. *et al.* 2009 Reprogramming of human somatic cells using human and animal oocytes. *Cloning Stem Cells* **11**, 1–11. (doi:10.1089/clo.2009.0004)

Clancy, C. E. & Rudy, Y. 1999 Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature* **400**, 566–569. (doi:10.1038/23034)

Clapworthy, G., Viceconti, M., Coveney, P. & Kohl, P. (eds) 2008 Editorial. *Phil. Trans. R. Soc. A* **366**, 2975–2978. (doi:10.1098/rsta.2008.0103)

Coen, E. 1999 *The art of genes*. Oxford, UK: Oxford University Press.

Corsi, P. 2001 *Lamarck. Genèse et enjeux du transformisme*. Paris, France: CNRS Editions.

Crick, F. H. C. 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **XII**, 138–163.

Crick, F. H. C. 1970 Central dogma of molecular biology. *Nature* **227**, 561–563. (doi:10.1038/227561a0)

Davidson, E. H. 2006 *The regulatory genome: gene regulatory networks in development and evolution*. New York, NY: Academic Press.

Dawkins, R. 1976, 2006 *The selfish gene*. Oxford, UK: Oxford University Press.

Dawkins, R. 1982 *The extended phenotype*. London, UK: Freeman.

Dawkins, R. 2006 *The selfish gene* (revised edn). Oxford, UK: Oxford University Press.

Dronamrajua, K. R. 1999 Erwin Schrödinger and the origins of molecular biology. *Genetics* **153**, 1071–1076.

Dupré, J. 1993 *The disorder of things*. Cambridge, MA: Harvard University Press.

Fenner, J. W. *et al.* 2008 The EuroPhysiome, STEP and a roadmap for the virtual physiological human. *Phil. Trans. R. Soc. A* **366**, 2979–2999. (doi:10.1098/rsta.2008.0089)

Franklin, R. E. & Gosling, R. G. 1953*a* Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **172**, 156–157. (doi:10.1038/172156a0)

Franklin, R. E. & Gosling, R. G. 1953*b* Molecular configuration in sodium thymonucleate. *Nature* **171**, 740–741. (doi:10.1038/171740a0)

Garny, A., Noble, D. & Kohl, P. 2005 Dimensionality in cardiac modelling. *Prog. Biophys. Mol. Biol.* **87**, 47–66. (doi:10.1016/j.pbiomolbio.2004.06.006)

Goldenfeld, N. & Woese, C. 2007 Biology's next revolution. *Nature* **445**, 369. (doi:10.1038/445369a)

Hillenmeyer, M. E. *et al.* 2008 The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365. (doi:10.1126/science.1150021)

Hodgkin, A. L. & Huxley, A. F. 1952 A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* **117**, 500–544.

Hunter, P. J., Pullan, A. J. & Smaill, B. H. 2003 Modelling total heart function. *Rev. Biomed. Eng.* **5**, 147–177. (doi:10.1146/annurev.bioeng.5.040202.121537)

Huxley, J. S. 1942 *Evolution: the modern synthesis*. London, UK: Allen & Unwin.

Jablonka, E. & Lamb, M. 1995 *Epigenetic inheritance and evolution. The Lamarckian dimension*. Oxford, UK: Oxford University Press.

Jablonka, E. & Lamb, M. 2005 *Evolution in four dimensions*. Boston, MA: MIT Press.
Jacob, F. 1982 *The possible and the actual*. New York, NY: Pantheon Books.
Jacob, F., Perrin, D., Sanchez, C., Monod, J. & Edelstein, S. 1960 The operon: a group of genes with expression coordinated by an operator. *C. R. Acad. Sci. Paris* **250**, 1727–1729.
Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. 2005 Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464. (doi:10.1038/nrg1615)
Keller, E. F. 1983 *A feeling for the organism: the life and work of Barbara McClintock*. New York, NY: W.H. Freeman.
Keller, E. F. 2000*a* *The century of the gene*. Cambridge, MA: Harvard University Press.
Keller, E. F. 2000*b* Is there an organism in this text? In *Controlling our destinies. historical, philosophical, ethical and theological perspectives on the human genome project* (ed. P. R. Sloan), pp. 273–288. Notre Dame, IN: University of Notre Dame Press.
Keller, L. 1999 *Levels of selection in evolution*. Princeton, NJ: Princeton University Press.
Kitcher, P. 1982 Genes. *Br. J. Phil. Sci.* **33**, 337–359. (doi:10.1093/bjps/33.4.337)
Kohl, P. & Noble, D. 2009 Systems biology and the virtual physiological human. *Mol. Syst. Biol.* **5**. (doi:10.1038/msb.2009.51)
Kupiec, J.-J. 2008 *L'origine des individus*. Paris, France: Fayard.
Kupiec, J.-J. 2009 *The origin of individuals: a Darwinian approach to developmental biology*. London, UK: World Scientific Publishing Company.
Laforge, B., Guez, D., Martinez, M. & Kupiec, J.-J. 2004 Modeling embryogenesis and cancer: an approach based on an equilibrium between the autostabilization of stochastic gene expression and the interdependence of cells for proliferation. *Prog. Biophys. Mol. Biol.* **89**, 93–120. (doi:10.1016/j.pbiomolbio.2004.11.004)
Lamarck, J.-B. 1809 *Philosophie Zoologique*. Paris, France: Dentu. (Reprinted by Flammarion 1994 as original edition with introduction by André Pichot.)
Levine, M. & Davidson, E. H. 2005 Gene regulatory networks for development. *Proc. Natl Acad. Sci. USA* **102**, 4936–4942. (doi:10.1073/pnas.0408031102)
Margulis, L. 1981 *Symbiosis in cell evolution*. London, UK: W.H. Freeman Co.
Margulis, L. & Sagan, D. 2002 *Acquiring genomes*. New York, NY: Basic Books.
Maynard Smith, J. & Szathmáry, E. 1995 *The major transitions in evolution*. Oxford, UK: Oxford University Press.
Mayr, E. 1982 *The growth of biological thought*. Cambridge, MA: Harvard University Press.
Neuman, Y. 2008 *Reviving the living: meaning making in living systems*. Amsterdam, The Netherlands: Elsevier.
Noble, D. 1960 Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations. *Nature* **188**, 495–497. (doi:10.1038/188495b0)
Noble, D. 1962 A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pacemaker potentials. *J. Physiol.* **160**, 317–352.
Noble, D. 2006 *The music of life*. Oxford, UK: Oxford University Press.
Noble, D. 2008*a* Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* **93**, 16–26. (doi:10.1113/expphysiol.2007.038695)
Noble, D. 2008*b* Computational models of the heart and their use in assessing the actions of drugs. *J. Pharmacol. Sci.* **107**, 107–117. (doi:10.1254/jphs.CR0070042)
Noble, D. 2008*c* Genes and causation. *Phil. Trans. R. Soc. A* **366**, 3001–3015. (doi:10.1098/rsta.2008.0086)
Nottale, L. 1993 *Fractal space-time and microphysics: towards a theory of scale relativity*. Singapore: World Scientific.
Nottale, L. 2000 *La relativité dans tous ses états. Du mouvements aux changements d'échelle*. Paris, France: Hachette.
Nottale, L. & Auffray, C. 2008 Scale relativity and integrative systems biology. II. Macroscopic quantum-type mechanics. *Prog. Biophys. Mol. Biol.* **97**, 115–157. (doi:10.1016/j.pbiomolbio.2007.09.001)
Pearson, H. 2006 What is a gene? *Nature* **441**, 399–401. (doi:10.1038/441398a)
Pichot, A. 1999 *Histoire de la notion de gène*. Paris, France: Flammarion.

Plank, G. *et al.* 2009 Generation of histo-anatomically representative models of the individual heart: tools and application. *Phil. Trans. R. Soc. A* **367**, 2257–2292. (doi:10.1098/rsta.2009.0056)

Saks, V., Monge, C. & Guzun, R. 2009 Philosophical basis and some historical aspects of systems biology: from Hegel to Noble—applications for bioenergetic research. *Int. J. Mol. Sci.* **10**, 1161–1192. (doi:10.3390/ijms10031161)

Scherrer, K. & Jost, J. 2007 Gene and genome concept. Coding versus regulation. *Theory Biosci.* **126**, 65–113. (doi:10.1007/s12064-007-0012-x)

Schrödinger, E. 1944 *What is life?* Cambridge, UK: Cambridge University Press.

Shapiro, J. A. 2005 A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* **345**, 91–100. (doi:10.1016/j.gene.2004.11.020)

Shapiro, J. A. 2009 Revisiting the central dogma in the 21st century. *Ann. N Y Acad. Sci.* **1178**, 6–28. (doi:10.1111/j.1749-6632.2009.04990.x)

Siegelmann, H. T. 1995 Computation beyond the Turing Limit. *Science* **268**, 545–548. (doi:10.1126/science.268.5210.545)

Siegelmann, H. T. 1998 *Neural networks and analog computation: beyond the Turing limit*. Boston, MA: Birkhauser.

Siegelmann, H. T. 1999 Stochastic analog networks and computational complexity. *J. Complexity* **15**, 451–475. (doi:10.1006/jcom.1999.0505)

Strathmann, R. R. 1993 Larvae and evolution: towards a new zoology. *Q. Rev. Biol.* **68**, 280–282. (doi:10.1086/418103)

Sulston, J. & Ferry, G. 2002 *The common thread*. London, UK: Bantam Press.

Sun, Y. H., Chen, S. P., Wang, Y. P., Hu, W. & Zhu, Z. Y. 2005 Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. *Biol. Reprod.* **72**, 510–515. (doi:10.1095/biolreprod.104.031302)

Venter, C. 2007 *A life decoded*. London, UK: Allen Lane.

Waddington, C. H. 1942 Canalization of development and the inheritance of acquired characteristics. *Nature* **150**, 563–565. (doi:10.1038/150563a0)

Waddington, C. H. 1957 *The strategy of the genes*. London, UK: Allen and Unwin.

Waddington, C. H. 1959 Canalization of development and genetic assimilation of acquired characteristics. *Nature* **183**, 1654–1655. (doi:10.1038/1831654a0)

Watson, J. D. & Crick, F. H. C. 1953*a* Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967. (doi:10.1038/171964b0)

Watson, J. D. & Crick, F. H. C. 1953*b* Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738. (doi:10.1038/171737a0)

Weaver, I. C. G. 2009 Life at the interface between a dynamic environment and a fixed genome. In *Mammalian brain development* (ed. D. Janigro), pp. 17–40. Totowa, NJ: Humana Press.

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Sekl, J. R., Dymov, S., Szyf, M. & Meaney, M. J. 2004 Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7**, 847–854. (doi:10.1038/nn1276)

Weaver, I. C. G., D'Alessio, A. C., Brown, S. E., Hellstrom, I. C., Dymov, S., Sharma, S., Szyf, M. & Meaney, M. J. 2007 The transcription factor nerve growth factor-inducible protein a mediates epigenetic programming: altering epigenetic marks by immediate-early genes. *J. Neurosci.* **27**, 1756–1768. (doi:10.1523/JNEUROSCI.4164-06.2007)

Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. 1953 Molecular structure of deoxypentose nucleic acids. *Nature* **171**, 738–740. (doi:10.1038/171738a0)

Williamson, D. I. 2003 *The origins of larvae*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Williamson, D. I. 2006 Hybridization in the evolution of animal form and life cycle. *Zool. J. Linn. Soc.* **148**, 585–602. (doi:10.1111/j.1096-3642.2006.00236.x)

Williamson, D. I. & Vickers, S. E. 2007 The origins of larvae. *Am. Sci.* **95**, 509–517.

On reading the amusing article 'Why I hate epigenetics' (*Physiology News* **77**, Winter 2009, p. 43) Denis Noble dreamt that he was the Editor and had received the following letter, which he then translated into English for the benefit of readers of *Physiology News*:

Jardin des Plantes,
Paris, le 21 novembre 2009

Monsieur l'éditeur

I had no idea that my scientific ideas were to become so politically sensitive, though I have been told that the distinguished Edinburgh professor of genetics and developmental biology, Conrad Waddington, was ignored by his fellow American scientists during the McCarthy inquisitions of the mid-20th century because of possible association with something called Communism, largely because he invented the term 'epigenetics' and claimed to have shown that it confirmed my ideas on inheritance. He called those ideas 'lamarckism' and was certainly not the first to do so. That damnable giraffe's neck (!) keeps returning to haunt me, whereas I thought I would be remembered for having introduced a new scientific subject, which I called biology (I was the first to do so), and for demonstrating the transformation of species and, hence, the basic truth of evolution.

I am deeply puzzled by the term 'lamarckism' for another reason also. Your brilliant Honorary Member, Charles Darwin, elected to that position on the foundation of your esteemed Society in 1876, also espoused the idea that acquired characteristics could be inherited [DN: see note 1]. In fact, like all biologists of our time, and even earlier, we absorbed this idea from our predecessors. I am amused that an idea for which I was not the inventor should have become so strongly associated with my name. I may be a 'demented gloating little troll' – in fact, I died so poor that they had to throw my body into a common lime-pit – but I can't quite see why I am associated with the idea any more than Mr Darwin. He never disagreed with me on this issue, since neither of us knew anything about the later discoveries of genetics that seemed to exclude it. He even introduced the idea of gemmules, particles that he imagined to flow through the blood stream to communicate acquired characteristics to the reproductive organs. Incidentally, your modern ideas on micro-chimerism are not so far from his idea of gemmules. It isn't just epigenetics that is resurrecting the idea of the inheritance of acquired characteristics, nor would Mr Darwin be surprised. I have it on good authority that he was uncomfortable with the dogmatism of those who usurped his name by calling themselves neo-darwinists. [DN: see note 2]

No, the main issue on which Mr Darwin and I disagreed was whether there was a direction to evolution, what I called 'le Pouvoir de la Vie'. This was not a mystical concept. In fact, I thought of it as derivable from basic physical principles, and so a perfectly natural phenomenon. Some of your modern ideas on complexity are not far removed from what I was thinking. Wouldn't it be better therefore for me to be seen as having laid the firm foundations of evidence for the transformation of species on which your Mr Darwin was to build? I argued the case for evolution with all the powerful skeptics of my day. The highly influential Georges Cuvier [DN: see note 3] ridiculed me mercilessly, even to the extent of gloating over my body in its pauper's grave. The so-called eulogy that he delivered on my death was described by your distinguished evolutionary theorist, Mr Stephen Jay Gould, as 'one of the most deprecatory and chillingly partisan biographies I have ever read.'

The fact is that I was reviled and died desperately poor (for which my family had to pay a heavy price) precisely because I had established the truth of, and argued strongly for, the idea of evolution. In this year of 2009, when you are rightly celebrating the bicentenary of Mr Darwin's birth, it would be nice if people might pause a little and recognize that it is also the bicentenary of my main work, *Philosophie Zoologique*. [DN: see note 4]

Veuillez accepter, cher Monsieur l'éditeur, l'expression de mes sentiments les plus distingués,

Jean-Baptiste Pierre Antoine de Monet, Chevalier de la Marck

Notes by Denis Noble

1. In his introduction to Harvard's republication in 1964 of *The Origin of Species*, Ernst Mayr wrote (pp. xxv–xxvi) "Curiously few evolutionists have noted that, in addition to natural selection, Darwin admits use and disuse as an important evolutionary mechanism. In this he is perfectly clear. For instance,... on page 137 he says that the reduced size of the eyes in moles and other burrowing mammals is 'probably due to gradual reduction from disuse, but aided perhaps by natural selection'. In the case of cave animals, when speaking of the loss of eyes he says, 'I attribute their loss wholly to disuse' (p. 137). On page 455 he begins unequivocally, 'At whatever period of life disuse or selection reduces an organ...' The importance he gives to use or disuse is indicated by the frequency with which he invokes this agent of evolution in the Origin. I find references on pages 11, 43, 134, 135, 136, 137, 447, 454, 455, 472, 479, and 480."

2. See Gabriel Dover's book *Dear Mr. Darwin: Letters on the Evolution of Life and Human Nature* (Phoenix books, 2001).

3. Cuvier argued that the fossil record showed sudden, not gradual, changes – an idea that Stephen Jay Gould later espoused in his theory of punctuated equilibrium. Despite the similarity of his ideas with those of Cuvier, he was shocked by the dismissive tone of Cuvier's 'eulogy' of Lamarck.

4. *Philosophie Zoologique* is a much better book than one might imagine, given the low esteem in which Lamarck is held today. He really did establish the transformation of species and, although he was not the first to develop the idea of evolution, he was an indefatigable proponent of the idea at a time when it was even more ridiculed than in Darwin's day – recall that Lamarck died (1829) long before publication of *The Origin of Species* (1859).

Systems Biology: An Approach

P Kohl[1], EJ Crampin[2], TA Quinn[1] and D Noble[1]

In just over a decade, Systems Biology has moved from being an idea, or rather a disparate set of ideas, to a mainstream feature of research and funding priorities. Institutes, departments, and centers of various flavors of Systems Biology have sprung up all over the world. An Internet search now produces more than 2 million hits. Of the 2,800 entries in PubMed with "Systems Biology" in either the title or the abstract, only two papers were published before 2000, and >90% were published in the past five years. In this article, we interpret Systems Biology as an approach rather than as a field or a destination of research. We illustrate that this approach is productive for the exploration of systems behavior, or "phenotypes," at all levels of structural and functional complexity, explicitly including the supracellular domain, and suggest how this may be related conceptually to genomes and biochemical networks. We discuss the role of models in Systems Biology and conclude with a consideration of their utility in biomedical research and development.

SYSTEMS BIOLOGY AS AN APPROACH

Origins

The use of Systems Biology approaches in analyzing biochemical networks is well established,[1] and it is now also gaining ground in explorations of higher levels of physiological function, as exemplified by the Physiome[2] and Virtual Physiological Human[3,4] projects. However, the use of the term "system" in the field of biology long predates "Systems Biology."

Throughout its existence as a discipline, physiology has concerned itself with the systems of the body (circulatory, nervous, immune, and so on). Back in 1542, Jean Fernel wrote, "So, if the parts of a complete Medicine are set in order, physiology will be the first of all; it concerns itself with the nature of the wholly healthy human being, all the powers and functions."[5] Claude Bernard is widely credited with introducing one of the key biological concepts—control of the internal environment—and he may therefore be viewed as the first "systems biologist,"[6] although good claims can also be made for William Harvey,[7] Gregor Mendel,[8] and others.

Essence

In order to explore the essence of Systems Biology—a notion that, in spite of its broad appeal, is still lacking a definition—it may be helpful to start by considering the meaning of each of the two words. "Biology" is easy to define: it is the science (Greek λόγος; "reason[ed] account") that is concerned with living matter (Greek βίος; "life").

Although perhaps less well appreciated in the biological field, the term "system" is equally well defined, as "an entity that maintains its existence through the mutual interaction of its parts."[9] Systems research, therefore, necessarily involves the combined application of "reductionist" and "integrationist" research techniques, to allow identification and detailed characterization of the parts, investigation of their interaction with one another and with their wider environment, and elucidation of how parts and interactions give rise to maintenance of the entity[10] (**Figure 1**).

Systems Biology, therefore, can be seen to stand for an *approach* to bioresearch, rather than a field or a destination.

This approach consciously combines reduction and integration from the outset of research and development activities, and it necessarily involves going across spatial scales of structural and functional integration (i.e., between the parts and the entity). There is no inherent restriction on the level at which "the system" may be defined. In fact, there is no such thing as *the* system because structures that are parts of one system (say, a mitochondrion in a cell) may form systems in their own right at a different level of integration (for example, in the contexts of electron transport chains and ATP synthesis). The focus of Systems Biology can be, but is not required to be, at the single-cell level (a predominant target so far). As an approach, Systems Biology is equally applicable to small or large biological entities.

In addition to addressing the relationship between structure and function from the nano- to the macroscale, Systems Biology interprets biological phenomena as dynamic processes whose inherent time resolution depends on the behavior studied. This range extends from submicroseconds for molecular-level

[1]Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK; [2]Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand. Correspondence: P Kohl (peter.kohl@dpag.ox.ac.uk)

Received 25 March 2010; accepted 20 April 2010; advance online publication 9 June 2010. doi:10.1038/clpt.2010.92

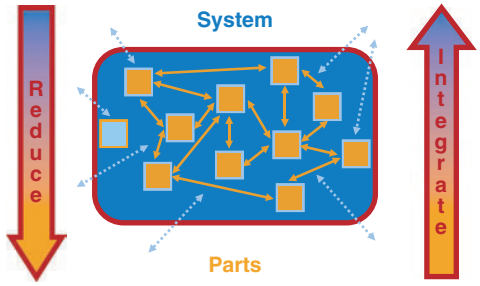


Figure 1 A system as an "entity that maintains its existence through the mutual interaction of its parts."[9] Systems research must combine (i) the identification and (ii) the detailed characterization of parts (orange boxes; as opposed to "lookalikes" (pale blue box), which need to be identified and excluded), with the exploration of their interactions (iii) with each other (orange arrows) and (iv) with the environment (pale blue dashed arrows) affecting parts either directly or indirectly, via modulation of internal interactions, to develop (v) a systemic understanding of the entity. An important, but often overlooked, aspect is that the system itself not only enables but also restricts the type and extent of functions and interactions that may occur (dark blue box). Systems research therefore requires a combination of tools and techniques for reduction and integration. Reprinted from ref. 10.

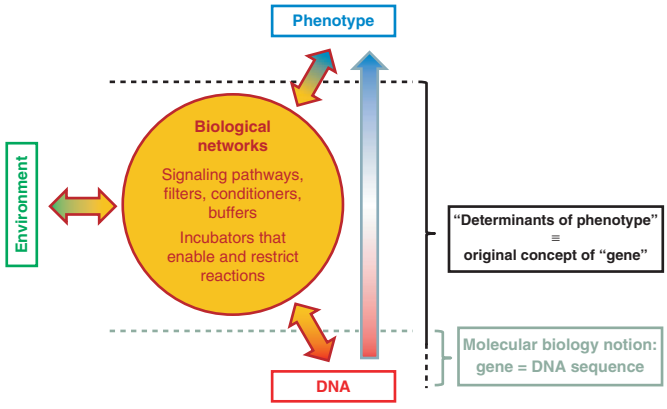


Figure 2 General relationships between genes, environment, and phenotype characters according to current physiological and biochemical understanding. The division of the conceptual entities—environment, phenotype, DNA, and biological networks—is neither strict nor mutually exclusive (and it does not specifically address the presence of any epigenomic information processing). Depending on the point of view, DNA, for example, is part of biological network activity (when you look "down" from the phenotype level), whereas biological networks are part of the environment (if you look "up" from DNA). It is hoped that this scheme will help to emphasize the complexity of interactions mediated by biological networks, which perform a whole host of key functions, such as enabling, filtering, conditioning, and buffering of the interplay between environment, phenotype, and DNA sequences. As shown on the right, the "determinants of a phenotype" (the original concept of genes) include much more than DNA sequences (the currently prevailing concept).

interactions to days, months, and years, e.g., for the development of a disease in humans.

Thus, Systems Biology explores how parts of biological entities function and interact to give rise to the behavior of the system as a whole. It is important to realize that "the entity," for example a cell, enables and restricts the range of components and interactions that are conceivable (e.g., a saline-based solute environment affects lipid bilayers in ways that are principally different from those of an alcohol-based solvent system, prescribing functional properties that need not be "encoded" other than in the basic biochemical and biophysical properties of the matter involved). However, the interrelation between genomic code and phenotypic representation deserves consideration in this context.

THE CONNECTION BETWEEN GENOMES AND PHENOTYPES

In order to understand biological systems, it is necessary to understand the relationship between the genome and the phenotype. When the concept of a gene was first introduced more than a century ago (see p.124 in Johannsen, 1909, where the term was derived from Greek γίνομαι; "to become"),[11] the relationship was thought to be simple. For each inheritable character, there was postulated to be a "gene" transmitting that character through the generations. This seemed to be the best interpretation of Mendel's experiments, implying discrete genetic elements that were responsible for phenotype characters. Later, even after this broad concept of genes was replaced by one focusing on DNA sequences as an equivalent information carrier, this idea persisted in the "one gene, one protein" hypothesis, even though proteins themselves are not the same as phenotype characters of complex organisms. Incidentally, this hypothesis is generally, but falsely, attributed to a 1941 PNAS paper by George W Beadle and Edward L Tatum.[12] In that paper, the authors show an example in fungi of "one gene, one enzyme" control of a step in vitamin B6 synthesis, but they highlight in the introduction "...it would appear that there must exist orders of directness of gene control ranging from simple one-to-one relations to relations of great complexity." The "one gene, one protein" hypothesis was developed over the following decade, and earned Beadle and Tatum the Nobel Prize in 1958, 5 years after the structural description of DNA by James D Watson and Francis Crick.

We now know that the relationships between "genotype" and "phenotype" are even more complex. Protein-coding DNA is assumed to form only 1% of metazoan genomes. It is controlled through multiple mechanisms involving DNA that is stably transcribed (i.e., functional) yet not protein-coding. The proportion of functional, non-protein-coding DNA is understood to be an order of magnitude larger than that of protein-coding DNA; however total functional DNA represents only ~10% of overall DNA content.[13] Many questions regarding the spatio-temporal organization of the regulatory genome remain to be resolved.[14] Also, whether the other 90% of DNA really has no function at all is an interesting question, particularly if one allows the notion of functionality to extend beyond its use as an RNA template (such as for scaffolding). Complete removal of the "junk DNA" is experimentally difficult (it does not form a coherent set of large segments). Interestingly, one study that removed two very large blocks of non-coding DNA (2.3 Mb) in mice found no significant changes in phenotype.[15] However, this is equivalent to just under 0.1% of the mouse genome (which would make it feasible, at least, to assume that structural effects of such deletion would have been minor or absent). It should also be recalled that many deletions, even of protein-coding regions, do not necessarily manifest themselves as a phenotypic change, unless the system is stressed.[16] Further complexity arises from the fact that multiple

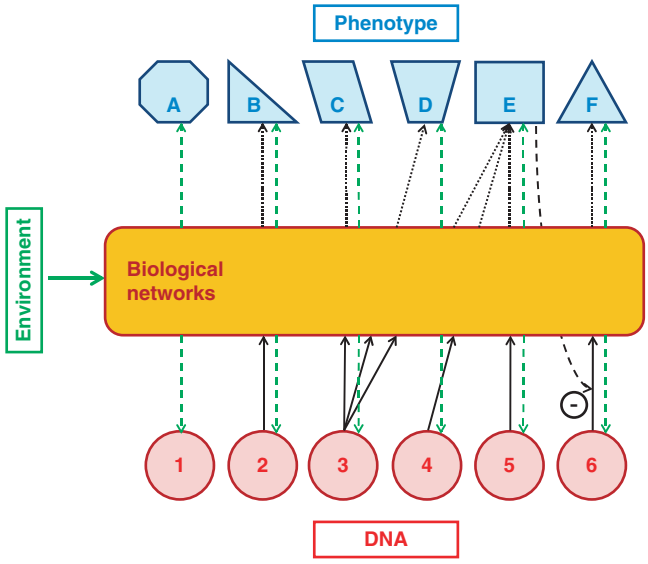


Figure 3 Simplified examples of interrelation between genes, environment, and phenotype characters according to current physiological and biochemical understanding. Interactions between particular DNA sequences and particular phenotype characters are mediated by biological networks. There is therefore no reason to assume direct causal relations between particular DNA sequences and particular phenotype characters in complex biological systems. To emphasize this, we have drawn each arrow of causation between a DNA sequence and a character as changing (from continuous to dotted) as it is transmitted through, and modified by, the biological interaction networks. Strictly speaking, not only do the causal arrows change, they interact within the network. The dotted arrows should therefore not be seen as mere continuations of the solid-line arrows. Green arrows highlight the fact that environmental influences (whether "external" or "internal" to the biological networks in this scheme) affect DNA sequences, their expression, and the shaping of phenotypic traits. Any diagram of these complex relationships is limited in what it can show. For details, see the text.

splice variants, even of the same DNA sequence, can give rise to alternative proteins. These effects are open to influences by the environment (here broadly defined as what is external to the system in question), and actual "DNA sequences" may not be as compact or uniquely defined as was initially assumed.[17]

There is therefore a (at least) three-way interaction between DNA, the environment, and the phenotype. **Figure 2** is an attempt to represent this interaction in a simplified scheme. Interactions are mediated through the networks within and between cells, including subcellular components such as proteins and organelles. These networks not only provide signaling pathways but also filter and condition the transmission of signals between environment, DNA, and phenotype. This is the basic explanation for the finding that interventions at the level of functional DNA (knockouts, insertions, and mutations) do not necessarily show a phenotypic effect. They are buffered by the networks, so that, even when changes at the level of proteins occur, there may be alternative (and normally redundant or quiescent) ways to ensure the retention of phenotype characters.

The influences of the phenotype and the environment on DNA are mediated by various mechanisms. DNA itself is chemically marked, e.g., by methylation of cytosines,[18,19] and control of expression is affected by interactions with histones (the histone code[20]). Together, these form part of the epigenome (http://www.epigenome.org) that constitutes a cellular memory, which can be transmitted to the subsequent generation(s). Longer-term effects include many forms of modification of the DNA itself through environment-induced genome rearrangement, nonrandom mutations, and gene transfer.[21] These have played a major role in the evolution of eukaryotic cells,[22] as have "gene" and "genome" duplication.[23] Similar mechanisms also play a major role in the immune system, in which targeted hypermutation in B cells can generate changes in the genome that are as much as 10^6 times greater than the normal mutation rates in the genome as a whole. This effectively extends the already huge range of antibodies that can be produced to an infinite one. Whereas the exact mechanism by which the recognition of a foreign antigen triggers or selects such DNA changes is not known, the existence of the process is well established.[24] This behavior is entirely somatic (restricted to the cells of the immune system) and is therefore not transmitted through the germline. It was originally thought that epigenetic marking was also restricted to somatic processes. There is, however, increasing evidence to show that some epigenetic marking can be transmitted via the germline[25] or via behavioral re-marking in each generation.[26]

The existence of these mechanisms makes the definition of a gene even more problematic. The horizontal lines in **Figure 2** indicate the difference between the original concept of genes and the modern definition. The original notion of a gene as the sufficient determinant of a phenotype includes everything below the black dashed line in **Figure 2** (although those who introduced the concept, such as Johannsen,[11] would not have known that). A "gene" in this sense is now understood to be a distributed cause, all of which is inherited (i.e., inheritance includes both DNA and other cellular components; here, conceptually separated—although they are, of course, usually combined). The modern molecular-biology definition of a gene is DNA alone (below the gray broken line in **Figure 2**) and is therefore very different from the original meaning, also from a causal viewpoint. This confusion in terminology lies at the heart of many arguments over the role of genes in physiological function, with an extremely simplified variant represented by the vertical arrow on the right in **Figure 2**. Genes, defined as DNA sequences, may form necessary but not sufficient causes of phenotype characters.

Figure 3 elaborates on this by depicting the relationships between individual DNA sequences and phenotype characters. To simplify what would otherwise be an illegible tangle of connections, we show just six DNA sequences and six phenotype characters and indicate only some of the connections that could exist between these 12 elements.

DNA sequence 1 does not contribute to any of the given phenotype characters, and its modification may give rise to irrelevant data and interpretations. Similarly (but unrelatedly), phenotype A is not affected by any of the given DNA sequences, and therefore assessment of causal relationships between the six DNA sequences shown and "A" may lead to false-negative conclusions (as DNA sequences outside the given range may be relevant). These two will be the most frequently encountered "causal" relations.

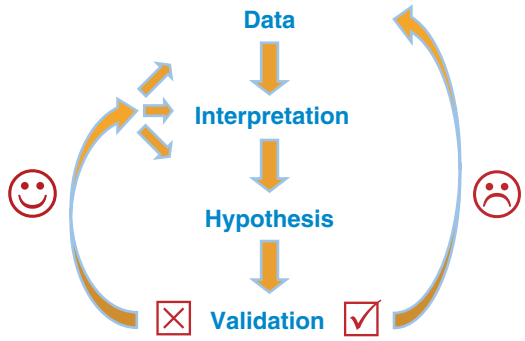


Figure 4 Schematic illustration of the scientific process and the role of validation. Emphasis is placed on the fact that, contrary to the common perception, the intellectual benefit of hypothesis rejection (left) may exceed that of confirmation (right). The value of successful hypothesis validation lies in increasing the level of confidence in a particular conceptual approach. Rejection highlights shortcomings in the approach and can be productive in guiding improved data acquisition, interpretation, and hypothesis formation.

DNA sequence 2 directly, and solely, contributes to phenotype characteristic B. This is the "ideal" scenario, which was once thought to be generally applicable. It is, in fact, either extremely rare or simply does not occur, except at the level of proteins in lower organisms such as prokaryotes.

DNA sequence 3 contributes to multiple phenotype characters (C, D, and E), whereas phenotype character E depends on DNA sequences 3–5. Such multiple connections are now known to be universal. The DNA–phenotype effects are, therefore, conditional. For example, a change in sequence 3 may not be translated into character E unless sequences 4 and 5 are knocked out as well; this again may contribute to potentially false-negative findings.

In addition, DNA–phenotype effects may affect other links, such as the one depicted by the dashed-line black arrow from phenotype characteristic E to DNA sequence 6 and, consequently, to characteristic F (this is merely one example and does not even begin to address the complexity of feedback from phenotype characteristics to underlying genetic determinants); this type of interaction may give rise to false-positive interpretations of data.

Each phenotype character also depends on cellular inheritance and on the influence of the environment via epigenetic and/or acute effects (see green arrows in **Figure 3**). All these influences are mediated by networks within cells and tissues. The traditional, "differential" view of genetics avoids acknowledging this mediation by focusing on a single change (usually a mutation, addition, or deletion) in a DNA sequence and the observed net change in phenotype. It then defines this as "the gene for" that characteristic (or, more precisely, the observed "difference" in characteristics). Clearly, this ignores the great majority of the components that, in combination, give rise to a phenotype character.

The logic of these conditional effects may be very complex, with various combinations forming a sufficient set of parameters that may give rise to similar or identical phenotypes. The major goal of a Systems Biology approach to genome–phenotype relations is to work out this logic. An "integral" view of genetics, which takes these complexities into account, is therefore essential to the success of Systems Biology.[10,27,28]

ROLE OF MODELS FOR SYSTEMS RESEARCH

Conceivably, if biology had turned out to be as simple as early geneticists envisaged, it could have continued to be an essentially descriptive subject. Identifying functions and their genetic causes could have been viewed as simply linking the two together, bit by bit, a function or a gene at a time. The complexity represented (albeit only partially and simplistically) in **Figures 2** and **3** shows that this is far from being the case. Beyond a certain degree of complexity, descriptive intuition often fails. When large numbers of genes and proteins are involved, the combinatorial problems become seriously challenging.[29] This is one of the reasons for another major characteristic of the Systems Biology approach: it makes extensive use of mathematical modeling in order to represent and understand complex interactions of parts and biological entities.

Mathematical models, however, need to be used with care. They are aids to thought, not a replacement for it. The only serious difference between a biologist who uses mathematical modeling and one who does not is that the former explores the consequences of his ideas quantitatively, including implementation of computational experiments to assess the plausibility of those ideas. The potential benefits of doing so are obvious because quantitatively plausible predictions improve subsequent hypothesis-driven experimental research. William Harvey[30] used this approach in his convincing arguments for the circulation of blood, when he calculated how quickly the blood in the body would run out if it did not recirculate (see also ref. 7). Using mathematics for quantitative prediction, Harvey arrived at an assessment of the plausibility of a certain hypothesis (or lack thereof, as the case may be).

Modeling of the electrophysiology of the heart, in particular, has repeatedly been used to direct new experimental approaches. In this process, the "failures" (predictions that were shown wrong in subsequent experimental assessment) have been as important as the "successes,"[31] as **Figure 4** illustrates. Let us assume, for a moment, that we all agree that proper scientific process is based on review of the available data and knowledge, followed by interpretation to form a falsifiable hypothesis, which is then subjected to validation.[32] Falsifiability of a theory as a virtue has been highlighted before, for example, by leading philosopher of science, Sir Karl Popper, who stated: "A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice."[32]

This view holds for the exploration of biological behavior. For the purpose of this argument, it does not matter whether this process is aided by formalized theoretical models (e.g., computer simulations) or is based entirely on conceptualization by an individual or group. If the validation shows agreement with the hypothesis, all it does is reconfirm what has been anticipated. Thus, arguably, no new insight is generated, although the data that emerge from the validation can be fed back into the scientific process (see **Figure 4**, right), and the same models (or concepts) will be applied in the future with a higher degree of confidence. Compare that to rejection of a hypothesis (**Figure 4**, left). Often seen as a less desirable outcome, it is when we show our best-conceived predictions to be wrong that we

learn something about shortcomings in input data, their interpretation (including any formalisms applied to aid this process), and/or the ensuing hypothesis (assuming that the approach to validation was suitable and applied correctly). This is the stage of the scientific process in which new insight is generated and the seeds for further progress are laid.[33]

Therefore, experimental information is the key to proper model development and validation, suggesting that "dry" computational modeling should not be pursued in isolation from "wet" lab or clinical studies. Incidentally, the reverse statement is prudent, too. Studies involving biological samples benefit from theoretical assessment of most likely outcomes, helping in the selection of promising approaches, supporting experimental design, and avoiding ill-conceived studies.[34] In other words, the cycle of "wet" data generation, "dry" interpretation and hypothesis formation, "wet" validation, and so on, should be seen as a continuous chain. Theoretical and practical research approaches do not thrive in isolation from each other.

The main limitations of mathematical modeling in biology arise from the very complexity that makes such modeling necessary.[35] By definition (model = simplified representation of reality), all models are partial descriptions of the original, whether they are conceptual (to think is to model!), mathematical/computational, or experimental/clinical. Of note, even an individual human would not be a perfect model system for the entire species, calling for patient-specific tools (including models) for prevention, diagnosis, and treatment.

Of course, a full representation of all aspects of a given reality in a "model" would render it a copy (or a clone). This would suffer exactly the same shortcomings with regard to the insight generated, ranging from complexity-related difficulty in identifying causal interrelations to ethico-legal boundaries on permissible interventions and data-gathering approaches. By the very definition of the term, an "all-inclusive" model would cease to be a model. The attempt to make such a model would strip it of all its advantages. It would be overburdened by what stands in need of simplification or explanation and offer no advantages for targeted assessment of hypotheses.

Like tools in a toolbox, each model has its inherent limitations and its specific utility. As an illustration, let us consider models of a train. Depending on purpose (toddler's toy, collector's replica, miniature railway), emphasis may be on simplicity, mechanical sturdiness, and color; on "to-scale" representation of appearance; or on mechanical function and ride comfort. An "all-inclusive model" of a train that captures every aspect, however, would be another train (and, as in patients, there are no two truly identical ones either). The copy train would not be suitable for application to the aforementioned model purposes, whether for the toddler, for the collector's display cabinet, or for your local landscaped gardens. Therefore, models can be good or bad only with respect to a particular purpose (in fact, well-suited or ill-suited would be more appropriate categories), but modeling *per se*—the utilization of simplified representations of reality—is neither: it is simply necessary. We all do it, in one way or another.

The difficulty in the case of complex biological systems (as opposed to man-made items) is that, on the basis of our present level of understanding, models remain very partial indeed. Therefore, for some time to come, there will be a place for both negative and positive validation to drive model improvement and to calibrate confidence. A problem to be wary of, not only in the context of formalized (mathematical) modeling, is what we can call the plausibility trap—just because a model reproduces observed behavior does not mean that implicated mechanisms are major contributors or even that they are involved at all. All that such models can do is to illustrate quantitative plausibility (which, in its own right, is certainly a major achievement). Even established theoretical models, therefore, require continual validation of predictions against the above described outcome-dependent consequences.

SYSTEMS BIOLOGY APPLICATION

If Systems Biology is accepted as an approach to biomedical research and development that, from the outset, consciously combines reduction and integration across a wide range of spatio-temporal scales, then one can explore different starting points for this systematic exploration of biological function.

Bottom–up

This is the classic molecular biology approach and can also be termed the "forward approach." It starts with "bottom" elements of the organism—genes and proteins—and represents these by equations that describe their known interactions. "Bottom" here is, of course, metaphorical. Genes and proteins are everywhere, in all cells of the body. It is a conceptual convenience to place them at the bottom of any multiscale representation, that is, with structures of low spatial dimensionality. From these components and their interactions, the modeler aims to reconstruct the system, including multiple feed-forward properties. It is conceivable that this might work in the case of the simplest organisms, such as prokaryotes, which can be represented as a relatively formless set of molecules with their networks surrounded by a lipid cell membrane. In the case of eukaryotes, many of the interactions between the components are restricted by the complex cell structure, including organelles. The forward approach would necessarily include these structures, in which case it is no longer purely bottom–up because, as we have already noted, many of these structural features are inherited independently of DNA sequences. Levels higher than DNA and proteins would be necessary for successful modeling. This does not imply that a bottom–up approach is of no value. It simply means that this approach, and the vast databanks that are being developed through genomics, proteomics, and bioinformatics, need to be complemented by other approaches. This need is underlined by studies showing that the great majority of DNA knockouts do not afford any insight into normal physiological function (for an example, see ref. 16).

Top–down

This may be regarded as the classic physiology approach, somewhat akin to reverse engineering. First, study the system at a high level, then burrow down to lower levels in an attempt to arrive at an inverse solution. In this case, we start with the system and try to infer its parts and their functionality. This

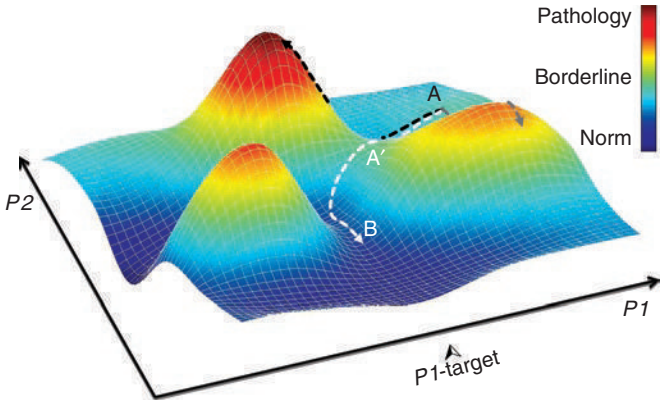


Figure 5 Schematic illustration of the landscape concept in parameter space. The value of a hypothetical biological function (color-coded, *z* axis) varies as a function of multiple parameters, including *P1* and *P2*. Assume a patient whose biological profile places him in position *A*, where the desired action (or a "side effect" associated with another treatment) is a reduction in the *P1* value toward a new target level. Direct reduction in *P1* (black trajectory) leads to severe negative consequences. Covariation in both *P1* and *P2* (white trajectory) allows transition toward the desired *P1* levels without detrimental changes. An isolated reduction in *P2* to the same extent (gray trajectory) would also be detrimental, showing that the combined action (passage from *A* to *B*) would not have been an intuitively obvious path to take.

approach has succeeded in some cases. The study of circulatory and respiratory physiology started off with the successful identification and characterization of a system (closed circulatory loop, pump function of the heart, gas exchange in lungs and tissues), leading eventually to identification of cells (red blood cells) and molecules (such as hemoglobin) that transport oxygen, and so on. It must be admitted, of course, that this approach has had its failures. High in the list of these failures is the classic view of genetics. Burrowing down to the level of DNA using differences in the phenotype to infer the existence of particular genes and then identifying individual properties from these DNA sequences can be seen as one of the great success stories of twentieth-century biology. Unfortunately, however, it works in only a small proportion of cases. The reasons are explained in **Figure 2**. There is no basis for supposing that we can always correctly infer the existence of particular DNA sequences from observations based on the phenotype because the relations between genotypes and phenotypes are massively multifactorial (**Figure 3**). In cross-species cloning, for example, cytoplasmic networks can even influence phenotypes (such as numbers of vertebrae), contradicting the expected genome influence.[36] In this case, the "gene" (in the classic sense of the term) is in the egg cytoplasm networks!

Middle–out

The limitations of the bottom–up and top–down approaches used in isolation have led to the adoption of the middle–out approach in a major proportion of work in Systems Biology at higher levels.[37] It can be represented as locally combining the bottom–up and top–down approaches, but that is only part of the story. Its success in the Physiome Project was possible precisely because it is pragmatic. Modeling begins at any level of the organization at which there are sufficient reliable data to build a model. This is the starting point of the middle–out approach. It involves exploration of parameter spaces at the chosen level. The next step is to reach toward both higher and lower levels of structural complexity (the "out" part of the metaphor). A good example of this approach is the modeling of the heart, which started at the level of the cell by representing processes and components that contribute to electrical, mechanical, or metabolic functions (see refs. 38, 39). It then reached upward to tissue and organ levels by incorporating the cell models into detailed models of higher-level tissue and organ structure (see refs. 40, 41) and downward to the genome by representing the effects of known genetic changes on the proteins represented in the model (see refs. 42, 43).

Whichever approach is adopted, successful models span different levels of organization. Causes of particular phenotype characteristics are unraveled as multidimensional interactions—the networks depicted in **Figure 2**. This leads us to a discussion of a very important conceptual tool: the multidimensionality of the many complex interactions in biological systems can be represented by what can be termed "landscape diagrams."

The landscape concept

Appreciation of the complexity and multidimensionality of the relationships between the components of organisms is not new. The idea of representing these relationships in the form of landscapes was introduced by Wright[44] and Waddington[45,46] (for a review, see ref. 47). When Waddington introduced his landscape metaphor, he used it to depict the rearrangements of genes in the gene pool that trigger the expression of different combinations of pre-existing alleles in response to environmental stress, a process he called epigenetics (note that the modern definition of epigenetics is different—it usually refers to chemical marking of the DNA). However, the landscape concept can usefully be applied much more broadly, relating the function of the biological system (or phenotype) to properties that we may seek to vary clinically (such as by pharmacological or device-based interventions) in order to manipulate the system toward a state of stability, safety, or health. Because of its focus on interactions, the landscape approach is already being used in Systems Biology.[48]

The underlying concept is that networks of interactions in a biological system can be represented as a multidimensional space in which variations in any of the parameters can be seen to correspond to perturbations in one (or more) of the dimensions. These effects find representation as changes, either in the landscape itself, as a translocation of functional states from one point to another within a given landscape, or a combination of both. **Figure 5** illustrates a conceptual example of state translocation to show how covariation of two parameters (P1, P2) may give rise to principally different effects on systems behavior (see the color scale) than one would have predicted from changing either of these parameters in isolation.

The importance of parameter interaction in complex systems has long been appreciated by engineers, and, correspondingly, mathematical theories to deal with this issue have been

developed. In one such approach, parameter interactions can be explored using "response surface methodology,"[49] a subset of "design of experiments" theory.[50] This collection of statistical techniques is tailored for parameter space exploration, with the aim of identifying maximally effective target combinations with the minimal number of experiments. Initially applied to optimization of production processes in various industries, the potential of these techniques for parameter optimization in drug- and device-based diagnosis and therapy has begun to be explored.[51,52]

The landscape approach aims to proceed beyond parameter optimization, to identify trajectories for dynamic parameter variation while keeping responses within a certain range. In **Figure 5**, for example, a straight connection from A to B would involve transition via a response range that, depending on dynamics (e.g., dwell times along parts of the trajectory), could be detrimental. This is avoided by moving through the intermediate target A′. Trajectory identification can be conducted in multiple ways. One option is to acquire a thorough knowledge of the entire landscape. This can be done using brute-force multidimensional parameter space exploration or with the guidance of coarse (or even adaptive) grid-point characterization, followed by detailed mapping of regions of interest (e.g., areas of steep changes in biological function or regions near known sites of desirable/undesirable functional behavior). Alternatively, one can conduct neighborhood mapping from (multiple) known source or target locations and try to interrelate identified fragments.

This is not a mere conceptual pastime; it is relevant to the development of therapeutic interventions. Early forays include the mid-nineteenth-century studies of Fraser, who noted the "hyperantagonistic" effect of two drugs: the herbal poison "physostigma" (a cholinesterase inhibitor) and "atropia" (atropine, a competitive antagonist for the muscarinic acetylcholine receptor that can act as a therapeutic antidote, unless given in excess).[53] Today, multi-drug combinations are common in medical treatments, and the effects of drugs can be additive, synergistic, antagonistic, or give rise to qualitatively different side effects (for example, via changes in compound metabolism). A good practical example is the evolution of knowledge concerning the actions of ranolazine (CV Therapeutics, now Gilead, Palo Alto, CA). This compound blocks the hERG channel (human Ether-à-go-go Related Gene, underlying the rapid delayed rectifying potassium current, I_{Kr}) and thereby prolongs the action potential in cardiac muscle cells. This type of response can be associated with an increased likelihood of heart rhythm disturbances. This is not the case here, however, because ranolazine also partially blocks the persistent sodium current ($i_{Na,p}$).[54] This combined action has two beneficial effects: it suppresses the development of so-called "early after-depolarizations" (which can cause acute initiation of heart rhythm disturbances), and it reduces sodium loading of the cell (which is a risk factor in the longer-term development of arrhythmias[55,56]). The blocking of $i_{Na,p}$ in isolation can also have negative side effects, in that this channel subtype is important for the initiation and conduction of the heart's electrical activation. Therefore, similar to what is shown in **Figure 5**, the combination of two wrongs can actually make a right. To date, ranolazine has been given US Food and Drug Administration approval for use in chest pain of cardiac origin (*angina pectoris*); further studies evaluate whether it is also an effective antiarrhythmic drug.

Similarly, the landscape concept can be productive in the development and application of medical devices. An example comes from the study of biventricular pacing optimization. Initial multiparameter pacing studies relied largely on varying one pacing parameter at a time, neglecting possible parameter interdependence that may give rise to nonlinear or cumulative effects. The advantage of exploring multiple variables simultaneously has been demonstrated in studies of simultaneous optimization of left ventricular pacing site and interventricular[57,58] or atrioventricular[59,60] pacing delay. Here, independent variation of single parameters may cause hemodynamic deterioration, whereas covariation improves patient status. The best trajectory of parameter variation for biventricular pacing optimization, for example, has been identified using a gradient method for targeted neighborhood mapping to guide the user through optimal parameter combinations.[61]

There are also many physiological examples of similar relationships in the heart. For example, hyperkalemia on its own can be fatal, as can be an excess of adrenaline. But when the two increase together, such as in exercise, the result is "safe."[62] The covariance of parameters can also go in opposite directions. For instance, when the background sodium current $i_{Na,b}$ is progressively reduced in a sinus-node pacemaker model, the hyperpolarization-activated "funny" current, i_f, automatically increases. The net result of this is a minimal change in beating rate.[63] This kind of reciprocal variation must be a basis for the robustness that biological systems display in response to interventions such as gene knockouts, many of which appear to have no phenotypic effect. Hillenmeyer *et al.*[16] studied this phenomenon in yeast and found that 80% of knockouts had no effect on the phenotype, as measured by cell growth and division, in a normal physiological environment. But when the organisms were metabolically stressed, 97% of the same knockouts did affect growth. In this example, the phenotypic expression of any given gene was therefore conditional on what the metabolic networks were experiencing. When backup networks are called into play because a particular metabolite is in short supply, the deficiency at the level of DNA may be revealed.

In mathematical models, robustness—that is, lack of significant changes in systems behavior despite significant parameter variation (for an example, see ref. 64)—is also referred to as "parameter sloppiness."[65] Determining safe areas in a functional landscape (**Figure 5**) is therefore equivalent to identifying regions of sloppiness. This is done by systematically exploring the range of parameter changes to which critical behavior of the system is insensitive. Such "insensitivity analysis" can be conducted either locally or in global parameter space. Estimates of global parameter sensitivity are typically based on sampling local sensitivities over multiple regions of a landscape (for example, by using the Morris method, see ref. 66). This requires close iteration between experimental data input and theoretical modeling

and is somewhat akin to the daunting task of drawing a map of a city by taking underground train transportation and characterizing the landscapes that present themselves at each overground exit without knowing the precise spatial interrelation among the stations.

What helps is that "sloppiness" is thought to be a universal property of Systems Biology models (much as "robustness" is common among biological systems). If this is true, it will be of great importance, for both the development of mathematical models and their practical application. Knowledge of critical parameter ranges is essential for producing reliable and predictive models, while insight into "uncritical" aspects will allow parameter reduction and model simplification. In the ideal scenario, models will be as complex as necessary yet as simple as possible to address a given problem.[67]

CONCLUSIONS

Systems Biology is an approach to biomedical research that consciously combines reduction and integration of information across multiple spatial scales to identify and characterize parts and explore the ways in which their interaction with one another and with the environment results in the maintenance of the entire system. In this effort, it faces the difficult task of connecting genomes and phenotypes, which are linked in a bidirectional manner and through complex networks of interaction, including modulation by the environment of the system itself. This process would be impossible without the use of advanced computational modeling techniques to explore the landscapes that are constituted by mutually interacting and highly dynamic parameters. The challenge for Systems Biology is to use multiparameter perturbations to identify the safe areas, in which covariation of multiple processes supports the maintenance of stability. Valleys in the landscape interconnect such areas, and their topography can guide the selection of patient-specific and safe treatment options.

This approach can be of use to the pharmaceutical industry in three ways. First, we may identify multitarget drug profiles that would be beneficial for a given purpose or condition. In fact, there may well be multiple solutions to the same problem, thereby expanding the range of available options for individual patients. Second, we should be able to predict tectonic changes, which involve the landscape itself being altered in such a way that the system shifts to a principally different, perhaps unstable, state outside the normal physiological range. Characterizing the factors that determine a switch from normal, or even disturbed, cardiac rhythms with a regular pattern (e.g., bradycardias or tachycardias) to chaotic behavior (e.g., fibrillation) is a good example. Achieving this, and then relating it to known properties of drug compounds, would greatly help the pharmaceutical discovery process (see ref. 68 for a comprehensive account of why this shift toward virtual R&D strategies will be vital for the industry as a whole). Third, if we have identified one (or several) safe combination(s) of background activity and intervention profiles, we may be able to map out isolines that demarcate the safe from the unsafe directions ("map out the valleys"). Patient-specific insensitivity analysis in particular could hold the key to identifying and eliminating the main obstacle to many otherwise efficient pharmacological treatments—drug side effects.

ACKNOWLEDGMENTS

Our work was supported by the European FP6 grants BioSim and normaCOR; by FP7 grants VPH NoE, preDiCT, and euHeart; and by the UK Biotechnology and Biological Sciences Research Council, the UK Medical Research Council, and the Wellcome Trust. P.K. is a senior fellow of the British Heart Foundation. E.J.C. acknowledges support from KAUST through a visiting fellowship held at the Oxford Centre for Collaborative Applied Mathematics. T.A.Q. is a postdoctoral research fellow of the UK Engineering and Physical Sciences Research Council.

CONFLICT OF INTEREST

The authors declared no conflict of interest.

1. Kitano, H. *Systems biology: towards systems-level understanding of biological systems*. In *Foundations of Systems Biology* (ed. Kitano, H.) (MIT Press, Cambridge, MA, 2002).
2. Bassingthwaighte, J., Hunter, P. & Noble, D. The Cardiac Physiome: perspectives for the future. *Exp. Physiol.* **94**, 597–605 (2009).
3. Fenner, J.W. *et al*. The EuroPhysiome, STEP and a roadmap for the virtual physiological human. *Philos. Transact. A. Math. Phys. Eng. Sci.* **366**, 2979–2999 (2008).
4. Hunter, P. *et al*. A vision and strategy for the VPH in 2010 and beyond. *Philos. Trans. R. Soc. A* **368**, 2595–2614 (2010).
5. Fernel, J. *Physiologia* (1542). Translated and annotated by Forrester, J.M. *Trans. Am. Philos. Soc.* **93/1**: 636 pp. (2003).
6. Noble, D. Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* **93**, 16–26 (2008).
7. Auffray, C. & Noble, D. Conceptual and experimental origins of integrative systems biology in William Harvey's masterpiece on the movement of the heart and the blood in animals. *Int. J. Mol. Sci.* **10**, 1658–1669 (2009).
8. Auffray, C. & Nottale, L. Scale relativity theory and integrative systems biology: I founding principles and scale laws. *Prog. Biophys. Mol. Biol.* **97**, 79–114 (2008).
9. von Bertalanffy, L. *General System Theory* (George Braziller, Inc, New York, 1968).
10. Kohl, P. & Noble, D. Systems biology and the virtual physiological human. *Mol. Syst. Biol.* **5**, 292 (2009).
11. Johannsen, W. *Elemente der Exakten Erblichkeitslehre* (Gustav Fischer, Jena, Germany, 1909).
12. Beadle, G.W. & Tatum, E.L. Genetic control of biochemical reactions in neurospora. *Proc. Natl. Acad. Sci. U.S.A.* **27**, 499–506 (1941).
13. Ponting, C.P. The functional repertoires of metazoan genomes. *Nat. Rev. Genet.* **9**, 689–698 (2008).
14. Alonso, M.E., Pernaute, B., Crespo, M., Gómez-Skarmeta, J.L. & Manzanares, M. Understanding the regulatory genome. *Int. J. Dev. Biol.* **53**, 1367–1378 (2009).
15. Nóbrega, M.A., Zhu, Y., Plajzer-Frick, I., Afzal, V. & Rubin, E.M. Megabase deletions of gene deserts result in viable mice. *Nature* **431**, 988–993 (2004).
16. Hillenmeyer, M.E. *et al*. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365 (2008).
17. Pennisi, E. Genomics. DNA study forces rethink of what it means to be a gene. *Science* **316**, 1556–1557 (2007).
18. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
19. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
20. Turner, B.M. Cellular memory and the histone code. *Cell* **111**, 285–291 (2002).
21. Shapiro, J.A. A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* **345**, 91–100 (2005).
22. Embley, T.M. & Martin, W. Eukaryotic evolution, changes and challenges. *Nature* **440**, 623–630 (2006).
23. Veron, A.S., Kaufmann, K. & Bornberg-Bauer, E. Evidence of interaction network evolution by whole-genome duplications: a case study in MADS-box proteins. *Mol. Biol. Evol.* **24**, 670–678 (2007).
24. Li, Z., Woo, C.J., Iglesias-Ussel, M.D., Ronai, D. & Scharff, M.D. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* **18**, 1–11 (2004).
25. Anway, M.D., Memon, M.A., Uzumcu, M. & Skinner, M.K. Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *J. Androl.* **27**, 868–879 (2006).

26. Weaver, I.C.G. *Life at the interface between a dynamic environment and a fixed genome.* In *Mammalian Brain Development* (ed. Janigrom, D.) 17–40 (Humana Press, Springer, New York, 2009).
27. Noble, D. Genes and causation. *Philos. Transact. A. Math. Phys. Eng. Sci.* **366**, 3001–3015 (2008).
28. Noble, D. Biophysics and systems biology. *Philos. Transact. A. Math. Phys. Eng. Sci.* **368**, 1125–1139 (2010).
29. Feytmans, E., Noble, D. & Peitsch, M. Genome size and numbers of biological functions. *Trans. Comput. Syst. Biol.* **1**, 44–49 (2005).
30. Harvey, W. *An Anatomical Disputation Concerning the Movement of the Heart and Blood in Living Creatures.* (Blackwell, Oxford, UK, 1627).
31. Noble, D. Modelling the heart: insights, failures and progress. *Bioessays* **24**, 1155–1163 (2002).
32. Popper, K. *Conjectures and Refutations* 33–39 (Routledge and Keagan Paul, London, 1963).
33. Kohl, P., Noble, D., Winslow, R. & Hunter, P.J. Computational modelling of biological systems: tools and visions. *Philos. Trans. R. Soc. A* **358**, 579–610 (2000).
34. Fink, M., Noble, P.J. & Noble, D. Mathematical models in cardiac electrophysiology research can help the 3Rs. NC3R <http://www.nc3rs.org.uk/news.asp?id=1162> (2008).
35. Hunter, P.J., Kohl, P. & Noble, D. Integrative models of the heart: achievements and limitations. *Philos. Trans. R. Soc. A* **359**, 1049–1054 (2001).
36. Sun, Y.H., Chen, S.P., Wang, Y.P., Hu, W. & Zhu, Z.Y. Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (Cyprinus carpio) nuclei and goldfish (Carassius auratus) enucleated eggs. *Biol. Reprod.* **72**, 510–515 (2005).
37. Brenner, S. *et al. Understanding complex systems: top-down, bottom-up or middle-out?* In *Novartis Foundation Symposium: Complexity in Biological Information Processing*, Vol. 239, 150–159 (Wiley, Chichester, UK, 2001).
38. Luo, C.H. & Rudy, Y. A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes. *Circ. Res.* **74**, 1071–1096 (1994).
39. Ten Tusscher, K.H.W.J., Noble, D., Noble, P.J. & Panfilov, A.V. A model of the human ventricular myocyte. *Am. J. Physiol.* **286**, H1573–H1589 (2004).
40. Plank, G. *et al*. Generation of histo-anatomically representative models of the individual heart: tools and application. *Philos. Transact. A. Math. Phys. Eng. Sci.* **367**, 2257–2292 (2009).
41. Vetter, F.J. & McCulloch, A.D. Three-dimensional analysis of regional cardiac function: a model of rabbit ventricular anatomy. *Prog. Biophys. Mol. Biol.* **69**, 157–183 (1998).
42. Noble, D., Sarai, N., Noble, P.J., Kobayashi, T., Matsuoka, S. & Noma, A. Resistance of cardiac cells to NCX knockout: a model study. *Ann. NY Acad. Sci.* **1099**, 306–309 (2007).
43. Sung, R.J., Wu, S.N., Wu, J.S., Chang, H.D. & Luo, C.H. Electrophysiological mechanisms of ventricular arrhythmias in relation to Andersen-Tawil syndrome under conditions of reduced IK1: a simulation study. *Am. J. Physiol. Heart Circ. Physiol.* **291**, H2597–H2605 (2006).
44. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Congr. Genet.* **1**, 356–366 (1932).
45. Waddington, C.H. Canalization of development and the inheritance of acquired characteristics. *Nature* **150**, 563–565 (1942).
46. Waddington, C.H. Canalization of development and genetic assimilation of acquired characters. *Nature* **183**, 1654–1655 (1959).
47. Bard, J.B.L. Waddington's legacy to developmental and theoretical biology. *Biol. Theory* **3**, 188–197 (2008).
48. Ao, P. Global view of bionetwork dynamics: adaptive landscape. *J. Genet. Genomics* **36**, 63–73 (2009).
49. Myers, R.H. & Montgomery, D.C. *Response Surface Methodology* (Wiley, New York, 2002).
50. Montgomery, D.C. *Design and Analysis of Experiments* (Wiley, New York, 1984).
51. Carter, W.H. Jr & Wampler, G.L. Review of the application of response surface methodology in the combination therapy of cancer. *Cancer Treat. Rep.* **70**, 133–140 (1986).
52. Tirand, L. *et al*. Response surface methodology: an extensive potential to optimize *in vivo* photodynamic therapy conditions. *Int. J. Radiat. Oncol. Biol. Phys.* **75**, 244–252 (2009).
53. Fraser, T.R. The antagonism between the actions of active substances. *Br Med J* **2**, 485–487 (1871).
54. Sakmann, B.F., Spindler, A.J., Bryant, S.M., Linz, K.W. & Noble, D. Distribution of a persistent sodium current across the ventricular wall in guinea pigs. *Circ. Res.* **87**, 910–914 (2000).
55. Noble, D. Computational models of the heart and their use in assessing the actions of drugs. *J. Pharmacol. Sci.* **107**, 107–117 (2008).
56. Noble, D. & Noble, P.J. Late sodium current in the pathophysiology of cardiovascular disease: consequences of sodium-calcium overload. *Heart* **92** (suppl. 4), iv1–iv5 (2006).
57. Berberian, G., Cabreriza, S.E., Quinn, T.A., Garofalo, C.A. & Spotnitz, H.M. Left ventricular pacing site-timing optimization during biventricular pacing using a multi-electrode patch. *Ann. Thorac. Surg.* **82**, 2292–2294 (2006).
58. Quinn, T.A., Cabreriza, S.E., Richmond, M.E., Weinberg, A.D., Holmes, J.W. & Spotnitz, H.M. Simultaneous variation of ventricular pacing site and timing with biventricular pacing in acute ventricular failure improves function by interventricular assist. *Am. J. Physiol. Heart Circ. Physiol.* **297**, H2220–H2226 (2009).
59. Whinnett, Z.I. *et al*. Haemodynamic effects of changes in atrioventricular and interventricular delay in cardiac resynchronisation therapy show a consistent pattern: analysis of shape, magnitude and relative importance of atrioventricular and interventricular delay. *Heart* **92**, 1628–1634 (2006).
60. Zuber, M., Toggweiler, S., Roos, M., Kobza, R., Jamshidi, P. & Erne, P. Comparison of different approaches for optimization of atrioventricular and interventricular delay in biventricular pacing. *Europace* **10**, 367–373 (2008).
61. Quinn, T.A. Optimization of biventricular pacing for the treatment of acute ventricular dysfunction. PhD thesis, Columbia University (2008).
62. Sears, C.E., Noble, P., Noble, D. & Paterson, D.J. Vagal control of heart rate is modulated by extracellular potassium. *J. Auton. Nerv. Syst.* **77**, 164–171 (1999).
63. Noble, D., Denyer, J.C., Brown, H.F. & DiFrancesco, D. Reciprocal role of the inward currents ib,Na and if in controlling and stabilizing pacemaker frequency of rabbit sino-atrial node cells. *Proc. R. Soc. B* **250**, 199–207 (1992).
64. Dassow, G.v., Meir, E., Munro, E.M. & Odell, G.M. The segment polarity network is a robust developmental module. *Nature* **406**, 188–192 (2000).
65. Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R. & Sethna, J.P. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, 1871–1878 (2007).
66. Cooling, M., Hunter, P. & Crampin, E.J. Modeling hypertrophic IP3 transients in the cardiac myocyte. *Biophys. J.* **93**, 3421–3433 (2007).
67. Garny, A., Noble, D. & Kohl, P. Dimensionality in cardiac modelling. *Prog. Biophys. Mol. Biol.* **87**, 47–66 (2005).
68. PricewaterhouseCoopers. Pharma 2020: Virtual R&D—which path will you take? <http://www.pwc.com/gx/en/pharma-life-sciences/pharma-2020/pharma2020-virtual-rd-which-path-will-you-take.jhtml> (2008).

Interface Focus

Interface Focus (2011) **1**, 7–15
doi:10.1098/rsfs.2010.0444
Published online 17 November 2010

Differential and integral views of genetics in computational systems biology

Denis Noble*

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

This article uses an integrative systems biological view of the relationship between genotypes and phenotypes to clarify some conceptual problems in biological debates about causality. The differential (gene-centric) view is incomplete in a sense analogous to using differentiation without integration in mathematics. Differences in genotype are frequently not reflected in significant differences in phenotype as they are buffered by networks of molecular interactions capable of substituting an alternative pathway to achieve a given phenotype characteristic when one pathway is removed. Those networks integrate the influences of many genes on each phenotype so that the effect of a modification in DNA depends on the context in which it occurs. Mathematical modelling of these interactions can help to understand the mechanisms of buffering and the contextual-dependence of phenotypic outcome, and so to represent correctly and quantitatively the relations between genomes and phenotypes. By incorporating all the causal factors in generating a phenotype, this approach also highlights the role of non-DNA forms of inheritance, and of the interactions at multiple levels.

Keywords: genotype; phenotype; computational systems biology

1. INTRODUCTION

Are organisms encoded as molecular descriptions in their genes? By analysing the genome, could we solve the forward problem of computing the behaviour of the system from this information, as was implied by the original idea of the 'genetic programme' [1] and the more modern representation of the genome as the 'book of life'? In this article, I will argue that this is both impossible and incorrect. We therefore need to replace the gene-centric 'differential' view of the relation between genotype and phenotype with an integrative view.

2. IMPOSSIBILITY

Current estimates of the number of genes in the human genome range up to 25 000, though the number would be even larger if we included regions of the genome forming templates for non-protein coding RNAs and as yet unknown numbers of microRNAs [2]. With no further information to restrict them, the number of conceivable interactions between 25 000 components is approximately 10^{70000} [3]. Many more proteins are formed than the number of genes, depending on the number of splice variants and post-transcriptional modifications. Proteins are the real workhorses of the organism so the calculation should really be based on this number, which may be in excess of 100 000, and further increased by a wide variety of post-translational modifications that influence their function.

Of course, such calculations are not realistic. In practice, the great majority of the conceivable interactions cannot occur. Compartmentalization ensures that some components never interact directly with each other, and proteins certainly do not interact with everything they encounter. Nevertheless, we cannot rely on specificity of interactions to reduce the number by as much as was once thought. Most proteins are not very specific [4,5]. Each has many interactions (with central hubs having dozens) with other elements in the organism [6], and many (around 30%) are unstructured in the sense that they lack a unique three-dimensional structure and so can change to react in variable ways in protein and metabolic networks [7].

In figure 1, I show the calculations for a more reasonable range of possible interactions by calculating the results for between 0 and 100 gene products for each biological function (phenotype characteristic) for genomes up to 30 000 in size. At 100 gene products per function, we calculate around 10^{300} possible interactions. Even when we reduce the number of genes involved in each function to 25 we still calculate a figure, 10^{80}, which is as large as the estimated number of elementary particles in the universe. These are therefore literally 'astronomic' numbers. We do not yet have any way of exploring interaction spaces of this degree of

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

One contribution of 16 to a Theme Issue 'Advancing systems medicine and therapeutics through biosimulation'.

Received 17 August 2010
Accepted 25 October 2010

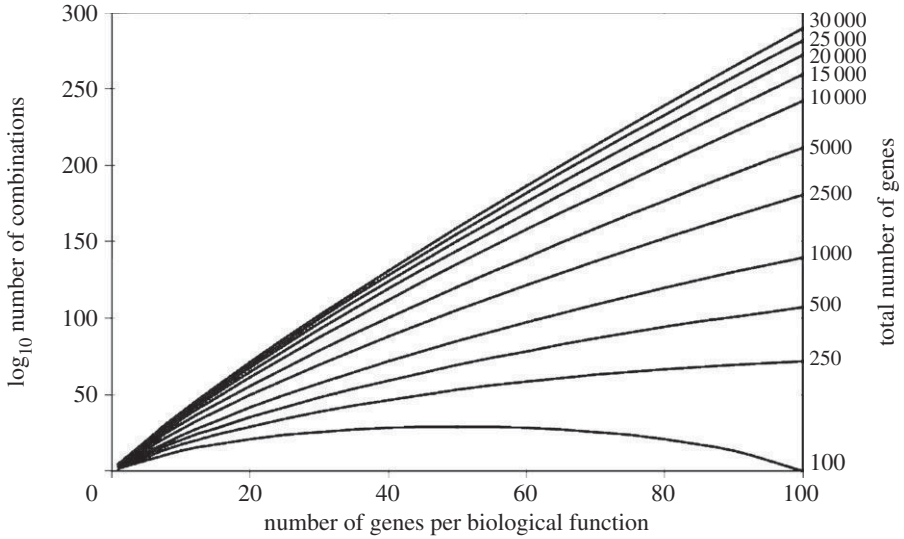


Figure 1. Genetic combinatorial explosion. Solutions of the equation $n\mathrm{Pr} = n(n-1)(n-2)\ldots\ldots(n-r+1) = n!/(n-r)!$, where n denotes number of genes in the genome, r is the number assumed to be involved in each function. Ordinate: number of possible combinations (potential biological functions). Abscissa: Number of genes required in each function. The curves show results for genomes of various sizes between 100 and 30 000 genes and for up to 100 genes involved in each function (adapted from Feytmans *et al.* [3]).

multi-dimensionality without insight into how the interactions are restricted. Computational biology has serious difficulties with the problem of combinatorial explosion even when we deal with just 100 elements, let alone tens of thousands.

Given these estimates of the scale of the forward problem, no-one should contemplate calculating the interactions in this massively 'blind' bottom-up fashion. That is the reason why the middle-out approach has been proposed [8]. This was originally a suggestion made by Brenner *et al.* [9]. The quotations from that Novartis Foundation discussion are interesting in the present context. Brenner wrote 'I know one approach that will fail, which is to start with genes, make proteins from them and to try to build things bottom-up' ([9], p. 51) and, then later, 'Middle-out. The bottom-up approach has very grave difficulties to go all the way' ([9], p. 154). My interpretation of the 'middle-out' approach is that you start calculating at the level at which you have the relevant data. In my work, this is at the level of cells, where we calculate the interactions between the protein and other components that generate cardiac rhythm, then we reach 'out' to go down towards the level of genes [10] and upwards towards the level of the whole organ [11,12].[1] By starting, in our case, at the level of the cell, we focus on the data relevant to that level and to a particular function at that level in order to reduce the number of components we must take into account. Other computational biologists choose other levels as their middle.

In practice, therefore, even a dedicated bottom-up computational biologist would look for ways in which nature itself has restricted the interactions that are theoretically possible. Organisms evolve step by step, with each step changing the options subsequently possible. I will argue that much of this restriction is embodied in the structural detail of the cells, tissues and organs of the body, as well as in its DNA. To take this route is therefore already to abandon the idea that the reconstruction can be based on DNA sequences alone.

3. INCORRECT

One possible answer to the argument so far could be that while we may not be able, in practice, to calculate all the possible interactions, nevertheless it may be true that the essence of all biological systems is that they are encoded as molecular descriptions in their genes. An argument from impossibility of computation is not, in itself, an argument against the truth of a hypothesis. In the pre-relativity and pre-quantum mechanical world of physics (a world of Laplacian billiard balls), many people considered determinate behaviour of the universe to be obviously correct even though they would readily have admitted the practical impossibility of doing the calculations.

To the problem of computability therefore we must add that it is clearly incorrect to suppose that all biological systems are encoded in DNA alone. An organism inherits not just its DNA. It also inherits the complete fertilized egg cell and any non-DNA components that come via sperm. With the DNA alone, the development process cannot even get started, as DNA itself is inert until triggered by transcription factors (various proteins and RNAs). These initially come from the mother [13] and from the father, possibly through RNAs carried in the sperm [14–16]. It is only through an interaction between DNA and its environment, mediated by these triggering molecules, that

[1]Note that the terms 'bottom', 'up', 'middle' and 'out' are conveying the sense of a hierarchy between levels of organization in biological systems that tends to ignore interactions that take place between levels in all directions. So very much as 'bottom-up' and 'top-down' approaches are arguably complementary, we should consider 'out-in' as well as 'middle-out' approaches in our attempts to integrate upward and downward causation chains.

development begins. The centriole also is inherited via sperm [17], while maternal transfer of antibodies and other factors has also been identified as a major source of transgenerational phenotype plasticity [18–20].

4. COMPARING THE DIFFERENT FORMS OF INHERITANCE

How does non-DNA inheritance compare with that through DNA? The eukaryotic cell is an unbelievably complex structure. It is not simply a bag formed by a cell membrane enclosing a protein soup. Even prokaryotes, formerly thought to fit that description, are structured [21] and some are also compartmentalized [22]. But the eukaryotic cell is divided up into many more compartments formed by the membranous organelles and other structures. The nucleus is also highly structured. It is not simply a container for naked DNA, which is why nuclear transfer experiments are not strict tests for excluding non-DNA inheritance.

If we wished to represent these structures as digital information to enable computation, we would need to convert the three-dimensional images of the cell at a level of resolution that would capture the way in which these structures restrict the molecular interactions. This would require a resolution of around 10 nm to give at least 10 image points across an organelle of around 100 nm diameter. To represent the three-dimensional structure of a cell around 100 μm across would require a grid of 10 000 image points across. Each gridpoint (or group of points forming a compartment) would need data on the proteins and other molecules that could be present and at what level. Assuming the cell has a similar size in all directions (i.e. is approximately a cube), we would require 10^{12} gridpoints, i.e. 1000 billion points. Even a cell as small as 10 μm across would require a billion grid points. Recall that the genome is about three billion base pairs. It is therefore easy to represent the three-dimensional image structure of a cell as containing as much information as the genome, or even more since there are only four possible nucleotides at each position in the genome sequence, whereas each grid point of the cellular structure representation is associated with digital or analogue information on a large number of features that are present or absent locally.

There are many qualifications to be put on these calculations and comparisons. Many of the cell structures are repetitive. This is what enables cell modellers to lump together compartments like mitochondria, endoplasmic reticulum, ribosomes, filaments, and other organelles and structures, though we are also beginning to understand that, sometimes, this is an oversimplification. A good example is the calcium signalling system in muscles, where the tiny spaces in which calcium signalling occurs, that couples excitation to contraction have to be represented at ever finer detail to capture what the experimental information tells us. Current estimates of the number of calcium ions in a single dyad (the space across which calcium signalling occurs) is only between 10 and 100 [23], too small for the laws of mass action to be valid.

Nevertheless, there is extensive repetition. One mitochondrion is basically similar to another, as are ribosomes and all the other organelles. But then, extensive repetition is also characteristic of the genome. A large fraction of the three billion base pairs forms repetitive sequences. Protein template regions of the human genome are estimated to be less than 1.5 per cent. Even if 99 per cent of the structural information from a cell image were to be redundant because of repetition, we would still arrive at figures comparable to the effective information content of the genome. And, for the arguments in this paper to be valid, it does not really matter whether the information is strictly comparable, nor whether one is greater than the other. Significance of information matters as much as its quantity. All I need to establish at this point is that, in a bottom-up reconstruction—or indeed in any other kind of reconstruction—it would be courting failure to ignore the structural detail. That is precisely what restricts the combinations of interactions (a protein in one compartment cannot interact directly with one in another, and proteins floating in lipid bilayer membranes have their parts exposed to different sets of molecules) and may therefore make the computations possible. Successful systems biology has to combine reduction and integration [24,25]. There is no alternative. Electrophysiological cell modellers are familiar with this necessity since the electrochemical potential gradients across membranes are central to function. The influence of these gradients on the gating of ion channel proteins is a fundamental feature of models of the Hodgkin–Huxley type. Only by integrating the equations for the kinetics of these channels with the electrochemical properties of the whole cell can the analysis be successful. As such models have been extended from nerve to cardiac and other kinds of muscle the incorporation of ever finer detail of cell structure has become increasingly important.

5. THE DIFFERENTIAL VIEW OF GENETICS

These points are so obvious, and have been so ever since electron microscopes first revealed the fine details of those intricate sub-cellular structures around 50 years ago, that one has to ask how mainstream genetics came to ignore the problem. The answer lies in what I will call the differential view of genetics.

At this point, a little history of genetics is relevant. The original concept of a gene was whatever is the inheritable cause of a particular characteristic in the phenotype, such as eye colour, number of limbs/digits, and so on. For each identifiable phenotype characteristic, there would be a gene (actually an allele—a particular variant of a gene) responsible for that characteristic. A gene could be defined therefore as something whose presence or absence makes a difference to the phenotype. When genetics was combined with natural selection to produce the modern synthesis [26], which is usually called neo-Darwinism, the idea took hold that only those differences were relevant to evolutionary success and all that mattered in relating

genetics to phenotypes was to identify the genetic causes of those differences. Since each phenotype must have such a cause (on this view at least) then selection of phenotypes amounts, in effect, to selection of individual genes. It does not really matter which way one looks at it. They are effectively equivalent [27]. The gene's-eye view then relegates the organism itself to the role of disposable carrier of its genes [28]. To this view we can add the idea that, in any case, only differences of genetic make-up can be observed. The procedure is simply to alter the genes, by mutation, deletion, addition and observe the effect on the phenotype.

I will call this gene-centric approach the 'differential view' of genetics to distinguish it from the 'integral view' I will propose later. To the differential view, we must add an implicit assumption. Since, on this view, no differences in the phenotype that are not caused by a genetic difference can be inherited, the fertilized egg cell (or just the cell itself in the case of unicellular organisms) does not evolve other than by mutations and other forms of evolution of its genes. The inherited information in the rest of the egg cell is ignored because (i) it is thought to be equivalent in different species (the prediction being that a cross-species clone will always show the phenotype of whichever species provides the genes), and (ii) it does not evolve or, if it does through the acquisition of new characteristics, these differences are not passed on to subsequent generations, which amounts to the same thing. Evolution requires inheritance. A temporary change does not matter.

At this stage in the argument, I will divide the holders of the differential view into two categories. The 'strong' version is that, while it is correct to say that the intricate structure of the egg cell is inherited as well as the genes, in principle that structure can be deduced from the genome information. On this view, a complete bottom-up reconstruction might still be possible even without the non-genetic information. This is a version of an old idea, that the complete organism is somehow represented in the genetic information. It just needs to be unfolded during development, like a building emerging from its blueprint.

The 'weak' version is one that does not make this assumption but still supposes that the genetic information carries all the differences that make one species different from another.

The weak version is easier to deal with, so I will start with that. In fact, it is remarkably easy to deal with. Only by restricting ourselves to the differential view of genetics it is possible to ignore the non-genetic structural information. But Nature does not play just with differences when it develops an organism. The organism develops only because the non-genetic structural information is also inherited *and is used to develop the organism*. When we try to solve the forward problem, we will be compelled to take that structural information into account *even if it were to be identical in different species*. To use a computer analogy, we need not only the 'programme' of life, we also need the 'computer' of life, the interpreter of the genome, i.e. the highly complex egg cell. In other words, we have to take the context of the cell into account, not only its genome. There is a question remaining, which is whether the

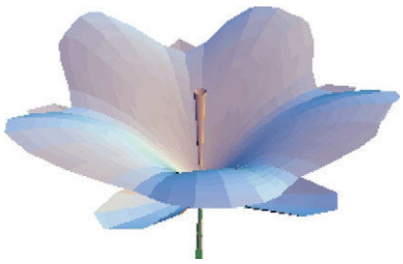


Figure 2. Solutions of a generalized Schrödinger equation for diffusive spheric growth from a centre (adapted from Nottale & Auffray [32]).

weak version is correct in assuming the identity of egg cell information between species. I will deal with that question later. The important point at this stage is that, even with that assumption, the forward problem cannot be solved on the basis of genetic information alone. Recall that genes need to be activated to do anything at all.

Proponents of the strong version would probably also take this route in solving the forward problem, but only as a temporary measure. They would argue that, when we have gained sufficient experience in solving this problem, we will come to see how the structural information is somehow also encoded in the genetic information.

This is an article of faith, not a proven hypothesis. As I have argued elsewhere [29,30], the DNA sequences do not form a 'programme' that could be described as complete in the sense that it can be parsed and analysed to reveal its logic. What we have found in the genome is better described as a database of templates [31] to enable a cell to make proteins and RNA. Unless that complete 'programme' can be found (which I would now regard as highly implausible given what we already know of the structure of the genome), I do not think the strong version is worth considering further. It is also implausible from an evolutionary viewpoint. Cells must have evolved before genomes. Why on earth would nature bother to 'code' for detail which is inherited anyway in the complete cell? This would be as unnecessary as attempting to 'code for' the properties of water or of lipids. Those properties are essential for life (they are what allow cells to form), but they do not require genes. Mother Nature would have learnt fairly quickly how to be parsimonious in creating genetic information: do not code for what happens naturally in the physico-chemical universe. Many wonderful things can be constructed on the basis of relatively little transmitted information, relying simply on physico-chemical processes, and these include what seem at first sight to be highly complex structures like that of a flower (see, for example, [32]; figures 2 and 3).

The point here is not that a flower can be made without genes (clearly, the image in figure 2 is *not* a flower—it does not have the biochemistry of a flower, for example), but rather that genes do not need to code for everything. Nature can, as it were, get 'free rides' from the physics of structure: the attractors towards which systems move naturally. Such physical structures do not require detailed templates

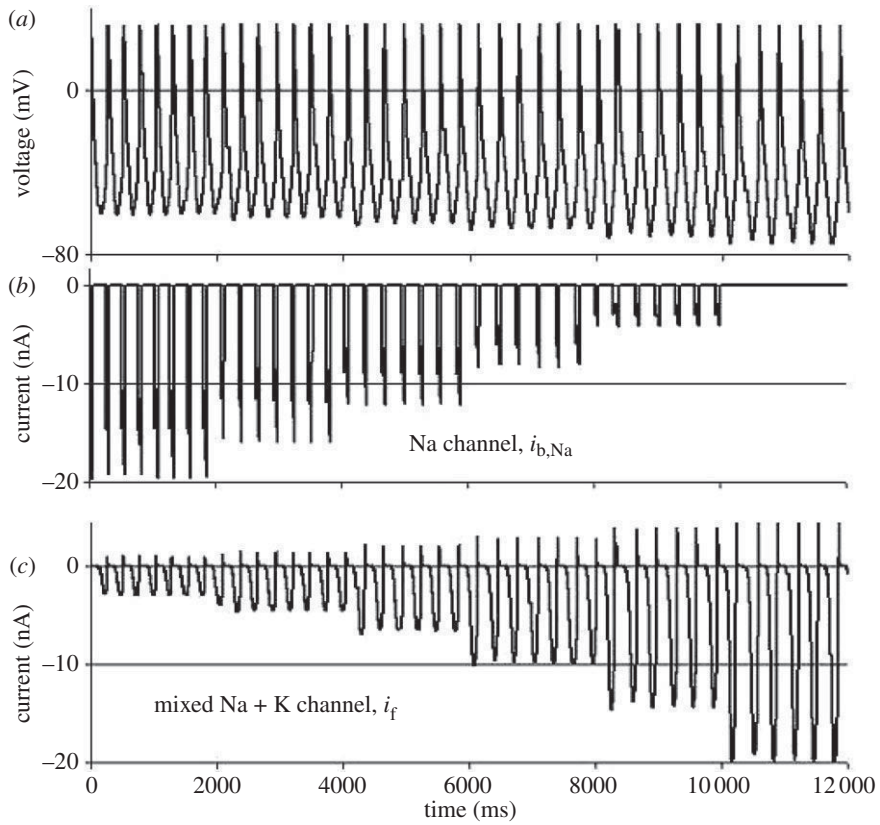


Figure 3. Example of the use of computational systems biology to model a genetic buffering mechanism. (*a*) Membrane potential variations in a model of the sinus node pacemaker of the heart. (*b*) The background sodium channel, $i_{b,Na}$, is progressively reduced until it is eventually 'knocked out'. (*c*) The mixed (sodium and potassium) cation current channel, i_f, progressively takes over the function, and so ensures that the change in frequency is minimized (adapted from Noble *et al.* [61]), recomputed using COR: http://cor.physiol.ox.ac.uk/. Coordinates: membrane potential in millivolt, current in nanoampere, time (abscissa) in milliseconds.

in the DNA sequences, they appear as the natural expression of the underlying physics. The structures can then act as templates for the self-organization of the protein networks, thus making self-organization a process depending both on the genome and the inherited structure.

6. IS THE DIFFERENTIAL VIEW CORRECT?

Both the strong and weak versions exclude the possibility of inheritance of changes in the non-DNA structural information. Indications that this may not be entirely correct have existed for many years. Over 50 years ago, McLaren & Michie [33] showed that the skeletal morphology (number of tail vertebrae) of different strains of mice depended on that of the mother into which the fertilized egg cell was implanted, and cannot therefore be entirely determined by the genome. Many other maternal effects have since been found in mammals [13,34]. We can now begin to understand how these effects may occur. The genome is marked epigenetically in various ways that modify gene-expression patterns. These markings can also be transmitted from one generation to another, either via the germline or via behavioural marking of the relevant genes [14,35,36].

Transmission of changes in structural information also occurs in unicellular animals. Again, this has been known for many years. Surgical modification of the direction of cilia patterns in paramecium, produced by cutting a pole of the animal and reinserting it the wrong way round, are robustly inherited by the daughter cells down many generations [37,38].

Interest in this kind of phenomenon has returned, perhaps in the wake of discoveries in epigenetics that make the phenomena explicable. A good example is the work of Sun *et al.* [39] on cross-species cloning of fish from different genera. They enucleated fertilized goldfish eggs and then inserted a carp nucleus. The overall body structure of the resulting adult fish is intermediate. Some features are clearly inherited from the goldfish egg. Intriguingly, in the light of McLaren and Michie's work, this included the number of vertebrae. The goldfish has fewer than the carp. So does the cross-species clone.[2]

Sun *et al.*'s [39] work is remarkable for another reason also. Success in creating adult cross-species clones is very rare. Virtually all other attempts at

[2]Note also that cross-species clones are not a full test of the differential view, since what is transferred between the species is not just DNA. The whole nucleus is transferred. All epigenetic marking that is determined by nuclear material would go with it. Cytoplasmic factors from the egg would have to compete with the nuclear factors to exert their effects.

cross-species cloning failed to develop to the adult [40]. An obvious possible explanation is that the egg cell information is too specific [41] as it has also evolved to become usually incompatible between different species. Strathmann [42] also refers to the influence of the egg cytoplasm on gene expression during early development as one of the impediments to hybridization in an evolutionary context. There is no good reason why cells themselves should have ceased to evolve once genomes arose. But if we need a specific (special purpose) 'computer' for each 'programme', the programme concept loses much of its attraction. The programming of living systems is distributed. Organisms are systems in continuous interaction with their environment. They are not Turing machines.

Contrary to the differential view, therefore, inheritance involves much more than nuclear DNA (see also [43]). It is simply incorrect to assume that all inherited differences are attributable to DNA [44,45].

7. THE INTEGRAL VIEW OF GENETICS

The alternative to the differential view is the integral approach. It is best defined as the complement to the differential approach. We study the contributions of a gene to all the functions in which its products take part. This is the approach of integrative biology, and here I am using 'integral' and 'integrative' in much the same sense. Integrative biology does not always or necessarily use mathematics of course, but even when it does not, the analogy with mathematical integration is still appropriate, precisely because it is not limited to investigating differences, and the additional information taken into account is analogous to the initial (= initial states of the networks of interactions) and boundary (= structural) conditions of mathematics. Indeed, they are exactly analogous when the mathematical modelling uses differential equations (as in figure 3 above). The middle-out approach is necessarily integrative. It must address the complexities arising from taking these conditions into account. The argument for the integrative approach is not that it is somehow easier or eliminates the complexity. On the contrary, the complexity is a major challenge. So, we need strong arguments for adopting this approach.

One such argument is that, most often, the differential approach does not work in revealing gene functions. Many interventions, such as knockouts, at the level of the genome are effectively buffered by the organism. In yeast, for example, 80 per cent of knockouts are normally 'silent' [46]. While there must be underlying effects in the protein networks, these are clearly hidden by the buffering at the higher levels. In fact, the failure of knockouts to systematically and reliably reveal gene functions is one of the great (and expensive) disappointments of recent biology. Note however that the disappointment exists only in the differential genetic view. By contrast, it is an exciting challenge from the integrative systems perspective. This very effective 'buffering' of genetic change is itself an important integrative property of cells and organisms. It is part of the robustness of organisms.

Moreover, even when a difference in the phenotype is manifest, it may not reveal the function(s) of the gene. In fact, it cannot do so, since *all* the functions shared between the original and the mutated gene are necessarily hidden from view. This is clearly evident when we talk of oncogenes [47]. What we mean is that a particular change in DNA sequence predisposes to cancer. But this does not tell us the function(s) of the unmutated gene, which would be better characterized as a cell cycle gene, an apoptosis gene, etc. Only a full physiological analysis of the roles of the proteins, for which the DNA sequence forms templates, in higher level functions can reveal that. That will include identifying the real biological regulators as systems properties. Knockout experiments by themselves do not identify regulators [48]. Moreover, those gene changes that do yield a simple phenotype change are the few that happen to reflect the final output of the networks of interactions.

So, the view that we can only observe *differences* in phenotype correlated with *differences* in genotype leads both to incorrect labelling of gene functions, and it falls into the fallacy of confusing the tip with the whole iceberg. We want to know what the relevant gene products do in the organism as a physiological whole, not simply by observing differences. Most genes and their products, RNA and proteins, have multiple functions.

My point here is not that we should abandon knockouts and other interventions at the genome level. It is rather that this approach needs to be complemented by an integrative one. In contrast to the days when genes were hypothetical entities—postulated as hidden causes (postulated alleles—gene variants) of particular phenotypes—we now identify genes as particular sequences of DNA. These are far from being hypothetical hidden entities. It now makes sense to ask: what are all the phenotypic functions in which they (or rather their products, the RNAs and proteins) are involved.

Restricting ourselves to the differential view of genetics is rather like working only at the level of differential equations in mathematics, as though the integral sign had never been invented. This is a good analogy since the constants of integration, the initial and boundary conditions, restrain the solutions possible in a way comparable to that by which the cell and tissue structures restrain whatever molecular interactions are possible. Modelling of biological functions should follow the lead of modellers in the engineering sciences. Engineering models are constructed to represent the integrative activity of all the components in the system. Good models of this kind in biology can even succeed in explaining the buffering process and why particular knockouts and other interventions at the DNA level do not reveal the function (figure 3 and [8], pp. 106–108).

An example of this approach is shown in figure 3. A computational model of rhythmic activity in the sino-atrial node of the heart was used to investigate the effect of progressive reduction in one of the ion channel proteins contributing current, $i_{b,Na}$, that determines the pacemaker frequency. In normal circumstances, 80 per cent of the depolarizing current

is carried by this channel. One might therefore expect a very large influence on frequency as the channel activity is reduced and finally knocked-out. In fact, the computed change in frequency is surprisingly small. The model reveals the mechanism of this very powerful buffering. As $i_{b,Na}$ is reduced, there is a small shift of the waveform in a negative direction: the amplitude of the negative phase of the voltage wave increases. This small voltage change is sufficient to increase the activation of a different ion channel current, i_f, to replace $i_{b,Na}$, so maintaining the frequency. The rest of the heart receives the signal corresponding to the frequency, but the change in amplitude is not transmitted. It is 'hidden'. This is how effective buffering systems work. Moreover, via the modelling we can achieve quantitative estimates of the absolute contribution of each protein channel to the rhythm, whereas simply recording the overall effect of the 'knockout' would hide those contributions; we would conclude that the contribution is very small. The integral approach succeeds, by estimating 80 per cent as the normal contribution of the sodium channel protein, where the differential approach fails by estimating only 10 per cent.

Finally, the integral view helps to resolve two related problems in heredity and evolutionary theory. The first is the question of the concept of a gene [49,50]. The existence of multiple splice variants of many genes, and the possibility even of splicing exons from different gene sequences, has led some biologists to propose that we should redefine the 'gene', for example as the completed mRNA [51]. An obvious difficulty with this approach is why should we stop at the mRNA stage? Why not go further and redefine the gene in terms of the proteins for which DNA sequences act as the templates, or even higher (see commentary by Noble [52])? The distinction between genotype and phenotype would then be less clear-cut and could even disappear. Something therefore seems wrong in this approach, at least if we wish to maintain the difference, and surely it does make sense to distinguish between what is inherited and what is produced as a consequence of that inheritance.

But perhaps we do not need to redefine genes at all. Why not just let the concept of individual genes be recognized as a *partial* truth, with reference to the genome as a whole, and specifically its organization, providing the more complete view? There could be different ways in which we can divide the genome up, only some of which would correspond to the current concept of a gene. Viewing the genome as an 'organ of the cell' [53] fits more naturally with the idea that the genome is a read-write memory [54], which is formatted in various ways to suit the organism, not to suit our need to categorize it. We certainly should not restrict our understanding of the way in which genomes can evolve by our imperfect definitions of a gene.

The second problem that this view helps to resolve is the vexed question of inheritance of acquired characteristics and how to fit it into modern evolutionary theory. Such inheritance is a problem for the neo-Darwinian synthesis precisely because it was formulated to exclude it. Too many exceptions now exist for that to be any longer tenable ([45]; see also the examples discussed previously).

In fact, the need to extend the synthesis has been evident for a long time. Consider, for example, the experiments of Waddington [55], who introduced the original idea of epigenetics. His definition was the rearrangement of gene alleles in response to environmental stress. His experiments on *Drosophila* showed that stress conditions could favour unusual forms of development, and that, after selection for these forms over a certain number of generations, the stress condition was no longer required (see discussion in Bard [56]). The new form had become permanently inheritable. We might argue over whether this should be called Lamarckism (see [57] for historical reasons why this term may be incorrect), but it is clearly an inherited acquired characteristic. Yet no mutations need occur to make this possible. All the gene alleles required for the new phenotype were already in the *population* but not in the right combinations in most, or even any, *individuals* to produce the new phenotype without the environmental stress. Those that did produce the new phenotype on being stressed had combinations that were at least partly correct. Selection among these could then improve the chances of individuals occurring for which the combinations were entirely correct so that the new phenotype could now be inherited even without the environmental stress. Waddington called this process an 'assimilation' of the acquired characteristic. There is nothing mysterious in the process of assimilation. Artificial selection has been used countless times to create new strains of animals and plants, and it has been used recently in biological research to create different colonies of high- and low-performing rats for studying disease states [58]. The main genetic loci involved can now be identified by whole genome studies (see, for example, [59]). The essential difference is that Waddington used an environmental stress that altered gene expression and revealed cryptic genetic variation and selected for this stress-induced response, rather than just selecting for the response from within an unstressed population The implication is obvious: in an environment in which the new phenotype was an advantage, natural selection could itself produce the assimilation. Natural selection is not incompatible with inheritance of acquired characteristics. As Darwin himself realized (for details, see Mayr [60]), the processes are complementary.

Neo-Darwinists dismissed Waddington's work largely because it did not involve the environment actually changing *individual* DNA gene sequences. But this is to restrict acquisition of evolutionarily significant change to individual DNA sequences (the gene's-eye view). On an integrative view, a new *combination* of alleles is just as significant from an evolutionary point of view. Speciation (defined, e.g., as failure of interbreeding) could occur just as readily from this process—and, as we now know, many other processes, such as gene transfer, genome duplication, symbiogenesis—as it might through the accumulation of mutations. What is the difference, from the organism's point of view, between a mutation in a particular DNA sequence that enables a particular phenotype to be displayed and a new combination of alleles that achieves the same result? There is an inherited change at the global genome level, even if no mutations

in individual genes were involved. Sequences change, even if they do not occur within what we characterize as genes. Taking the integrative view naturally leads to a more inclusive view of the mechanisms of evolutionary change. Focusing on individual genes obscures this view.

In this article, I have been strongly critical of the gene-centred differential view. Let me end on a more positive note. The integral view does not exclude the differential view any more than integration excludes differentiation in mathematics. They complement each other. Genome sequencing, epigenomics, metabolomics, proteomics, transcriptomics are all contributing basic information that is of great value. We have only to think of how much genome sequencing of different species has contributed to evolutionary theory to recognize that the huge investment involved was well worth the effort. As integrative computational biology advances, it will be using this massive data collection, and it will be doing so in a meaningful way. The 'meaning' of a biological function lies at the level at which it is integrated, often enough at the level of a whole cell (a point frequently emphasized by Sydney Brenner), but in principle, the integration can be at any level in the organism. It is through identifying that level and the meaning to the whole organism of the function concerned that we acquire the spectacles required to interpret the data at other levels.

Work in the author's laboratory is funded by the EU (the Biosim network of excellence under Framework 6 and the PreDiCT project under Framework 7) and the British Heart Foundation. I would like to thank the participants of the seminars on Conceptual Foundations of Systems Biology at Balliol College, particularly Jonathan Bard, Tom Melham and Eric Werner, and Peter Kohl for the context of discussions in which some of the ideas for this article were developed. I thank Charles Auffray and the journal referees for many valuable suggestions on the manuscript.

REFERENCES

1 Jacob, F. & Monod, J. 1961 Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356. (doi:10.1016/S0022-2836(61)80072-7)
2 Baulcombe, D. 2002 DNA events. An RNA microcosm. *Science* **297**, 2002–2003. (doi:10.1126/science.1077906)
3 Feytmans, E., Noble, D. & Peitsch, M. 2005 Genome size and numbers of biological functions. *Trans. Comput. Syst. Biol.* **1**, 44–49. (doi:10.1007/978-3-540-32126-2_4)
4 Bray, D. 2009 *Wetware. A computer in every cell.* New Haven, CT: Yale University Press.
5 Kupiec, J. 2009 *The origin of individuals: a Darwinian approach to developmental biology.* London, UK: World Scientific Publishing Company.
6 Bork, P., Jensen, L. J., von Mering, C., Ramani, A. K., Lee, I.-S. & Marcotte, E. M. 2004 Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.* **14**, 292–299. (doi:10.1016/j.sbi.2004.05.003)
7 Gsponer, J. & Babu, M. M. 2009 The rules of disorder or why disorder rules. *Progr. Biophys. Mol. Biol.* **99**, 94–103. (doi:10.1016/j.pbiomolbio.2009.03.001)
8 Noble, D. 2006 *The music of life.* Oxford, UK: Oxford University Press.
9 Brenner, S., Noble, D., Sejnowski, T., Fields, R. D., Laughlin, S., Berridge, M., Segel, L., Prank, K. & Dolmetsch, R. E. 2001 Understanding complex systems: top-down, bottom-up or middle-out? In *Novartis Foundation Symposium: Complexity in biological information processing*, vol. 239, pp. 150–159. Chichester, UK: John Wiley.
10 Clancy, C. E. & Rudy, Y. 1999 Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature* **400**, 566–569. (doi:10.1038/23034)
11 Bassingthwaighte, J. B., Hunter, P. J. & Noble, D. 2009 The Cardiac Physiome: perspectives for the future. *Exp. Physiol.* **94**, 597–605. (doi:10.1113/expphysiol.2008.044099)
12 Noble, D. 2007 From the Hodgkin–Huxley axon to the virtual heart. *J. Physiol.* **580**, 15–22. (doi:10.1113/jphysiol.2006.119370)
13 Gluckman, P. & Hanson, M. 2004 *The fetal matrix. Evolution, development and disease.* Cambridge, UK: Cambridge University Press.
14 Anway, M. D., Memon, M. A., Uzumcu, M. & Skinner, M. K. 2006 Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *J. Androl.* **27**, 868–879. (doi:10.2164/jandrol.106.000349)
15 Barroso, G., Valdespin, C., Vega, E., Kershenovich, R., Avila, R., Avendano, C. & Oehninger, S. 2009 Developmental sperm contributions: fertilization and beyond. *Fertil. Steril.* **92**, 835–848. (doi:10.1016/j.fertnstert.2009.06.030)
16 Pembrey, M. E., Bygren, L. O., Kaati, G., Edvinsson, S., Northstone, K., Sjostrom, M., Golding, J. & ALSPAC study team 2006 Sex-specific, male-line transgenerational responses in humans. *Eur. J. Hum. Genet.* **14**, 159–166. (doi:10.1038/sj.ejhg.5201538)
17 Sathananthan, A. H. 2009 Editorial. Human centriole: origin, and how it impacts fertilization, embryogenesis, infertility and cloning. *Ind. J. Med. Res.* **129**, 348–350.
18 Agrawal, A. A., Laforsch, C. & Tollrian, R. 1999 Transgenerational induction of defences in animals and plants. *Nature* **401**, 60–63. (doi:10.1038/43425)
19 Boulinier, T. & Staszewski, V. 2008 Maternal transfer of antibodies: raising immuno-ecology issues. *Trends Ecol. Evol.* **23**, 282–288. (doi:10.1016/j.tree.2007.12.006)
20 Hasselquist, D. & Nilsson, J. A. 2009 Maternal transfer of antibodies in vertebrates: trans-generational effects on offspring immunity. *Phil. Trans. R. Soc. B* **364**, 51–60. (doi:10.1098/rstb.2008.0137)
21 Michie, K. A. & Lowe, J. 2006 Dynamic filaments of the bacterial cytoskeleton. *Ann. Rev. Biochem.* **75**, 467–492. (doi:10.1146/annurev.biochem.75.103004.142452)
22 Fuerst, J. 2005 Intracellular compartmentation in planctomycetes. *Ann. Rev. Microbiol.* **59**, 299–328. (doi:10.1146/annurev.micro.59.030804.121258)
23 Tanskanen, A. J., Greenstein, J. L., Chen, A., Sun, S. X. & Winslow, R. L. 2007 Protein geometry and placement in the cardiac dyad influence macroscopic properties of calcium-induced calcium release. *Biophys. J.* **92**, 3379–3396. (doi:10.1529/biophysj.106.089425)
24 Kohl, P., Crampin, E., Quinn, T. A. & Noble, D. 2010 Systems biology: an approach. *Clin. Pharmacol. Ther.* **88**, 25–33. (doi:10.1038/clpt.2010.92)
25 Kohl, P. & Noble, D. 2009 Systems biology and the Virtual Physiological Human. *Mol. Syst. Biol.* **5**, 292, 1–6.
26 Huxley, J. S. 1942 *Evolution: the modern synthesis.* London, UK: Allen & Unwin.
27 Dawkins, R. 1982 *The extended phenotype.* London, UK: Freeman.
28 Dawkins, R. 1976 *The selfish gene.* Oxford, UK: OUP.
29 Noble, D. 2008 Genes and causation. *Phil. Trans. R. Soc. A* **366**, 3001–3015. (doi:10.1098/rsta.2008.0086)

30 Noble, D. 2010 Biophysics and systems biology. *Phil. Trans. R. Soc. A* **368**, 1125–1139. (doi:10.1098/rsta.2009.0245)
31 Atlan, H. & Koppel, M. 1990 The cellular computer DNA: program or data? *Bull. Math. Biol.* **52**, 335–348.
32 Nottale, L. & Auffray, C. 2008 Scale relativity and integrative systems biology 2. Macroscopic quantum-type mechanics. *Progr. Biophys. Mol. Biol.* **97**, 115–157. (doi:10.1016/j.pbiomolbio.2007.09.001)
33 McLaren, A. & Michie, D. 1958 An effect of uterine environment upon skeletal morphology of the mouse. *Nature* **181**, 1147–1148. (doi:10.1038/1811147a0)
34 Mousseau, T. A. & Fox, C. W. 1998 *Maternal effects as adaptations*. Oxford, UK: Oxford University Press.
35 Weaver, I. C. G. 2009 Life at the interface between a dynamic environment and a fixed genome. In *Mammalian brain development* (ed. D. Janigro), pp. 17–40. New York, NY: Humana Press, Springer.
36 Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Sekl, J. R., Dymov, S., Szyf, M. & Meaney, M. J. 2004 Epigenetic programming by maternal behavior. *Nat. Neurosci.* **7**, 847–854. (doi:10.1038/nn1276)
37 Beisson, J. & Sonneborn, T. M. 1965 Cytoplasmic inheritance of the organization of the cell cortex in paramecium Aurelia. *Proc. Natl Acad. Sci. USA* **53**, 275–282. (doi:10.1073/pnas.53.2.275)
38 Sonneborn, T. M. 1970 Gene action on development. *Proc. R. Soc. Lond. B* **176**, 347–366. (doi:10.1098/rspb.1970.0054)
39 Sun, Y. H., Chen, S. P., Wang, Y. P., Hu, W. & Zhu, Z. Y. 2005 Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. *Biol. Reprod.* **72**, 510–515. (doi:10.1095/biolreprod.104.031302)
40 Chung, Y. *et al.* 2009 Reprogramming of human somatic cells using human and animal oocytes. *Cloning Stem Cells* **11**, 1–11. (doi:10.1089/clo.2009.0004)
41 Chen, T., Zhang, Y.-L., Jiang, Y., Liu, J.-H., Schatten, H., Chen, D.-Y. & Sun, Y. 2006 Interspecies nuclear transfer reveals that demethylation of specific repetitive sequences is determined by recipient ooplasm but not by donor intrinsic property in cloned embryos. *Mol. Reprod. Dev.* **73**, 313–317. (doi:10.1002/mrd.20421)
42 Strathmann, R. R. 1993 Larvae and evolution: towards a new zoology (book review). *Q. Rev. Biol.* **68**, 280–282. (doi:10.1086/418103)
43 Maurel, M.-C. & Kanellopoulos-Langevin, C. 2008 Heredity: venturing beyond genetics. *Biol. Reprod.* **79**, 2–8. (doi:10.1095/biolreprod.107.065607)
44 Jablonka, E. & Lamb, M. 1995 *Epigenetic inheritance and evolution. The Lamarckian dimension*. Oxford, UK: Oxford University Press.
45 Jablonka, E. & Lamb, M. 2005 *Evolution in four dimensions*. Boston, MA: MIT Press.
46 Hillenmeyer, M. E. *et al.* 2008 The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365. (doi:10.1126/science.1150021)
47 Weinberg, R. A. 1996 How cancer arises. *Scient. Am.* **275**, 62–70. (doi:10.1038/scientificamerican0996-62)
48 Davies, J. 2009 Regulation, necessity, and the misinterpretation of knockouts. *Bioessays* **31**, 826–830. (doi:10.1002/bies.200900044)
49 Pearson, H. 2006 What is a gene? *Nature* **441**, 399–401. (doi:10.1038/441398a)
50 Pennisi, E. 2007 DNA study forces rethink of what it means to be a gene. *Science* **316**, 1556–1557. (doi:10.1126/science.316.5831.1556)
51 Scherrer, K. & Jost, J. 2007 Gene and genon concept. Coding versus regulation. *Theory Biosci.* **126**, 65–113. (doi:10.1007/s12064-007-0012-x)
52 Noble, D. 2009 Commentary on Scherrer & Jost (2007) Gene and genon concept: coding versus regulation. *Theory Biosci.* **128**, 153. (doi:10.1007/s12064-009-0073-0)
53 McClintock, B. 1984 The significance of responses of the genome to challenge. *Science* **226**, 792–801. (doi:10.1126/science.15739260)
54 Shapiro, J. A. 2009 Letting *E. coli* teach me about genome engineering. *Genetics* **183**, 1205–1214. (doi:10.1534/genetics.109.110007)
55 Waddington, C. H. 1959 Canalization of development and genetic assimilation of acquired characteristics. *Nature* **183**, 1654–1655. (doi:10.1038/1831654a0)
56 Bard, J. B. L. 2008 Waddington's legacy to developmental and theoretical biology. *Biol. Theory* **3**, 188–197. (doi:10.1162/biot.2008.3.3.188)
57 Noble, D. 2010 Letter from Lamarck. *Physiol. News* **78**, 31.
58 Koch, L. G. & Britton, S. L. 2001 Artificial selection for intrinsic aerobic endurance running capacity in rats. *Physiol. Genom.* **5**, 45–52.
59 Rubin, C.-J. *et al.* 2010 Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591. (doi:10.1038/nature08832)
60 Mayr, E. 1964 Introduction. In *The origin of species*. Cambridge, MA: Harvard.
61 Noble, D., Denyer, J. C., Brown, H. F. & DiFrancesco, D. 1992 Reciprocal role of the inward currents $i_{b,Na}$ and i_f in controlling and stabilizing pacemaker frequency of rabbit sino-atrial node cells. *Proc. R. Soc. B* **250**, 199–207. (doi:10.1098/rspb.1992.0150)

J Physiol 589.5 (2011) pp 1007–1015

The Journal of Physiology

TOPICAL REVIEW

Neo-Darwinism, the Modern Synthesis and selfish genes: are they of use in physiology?

Denis Noble

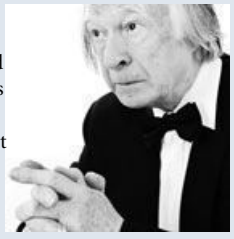
Department of Physiology, Anatomy and Genetics, Parks Road, Oxford OX1 3PT, UK

This article argues that the gene-centric interpretations of evolution, and more particularly the selfish gene expression of those interpretations, form barriers to the integration of physiological science with evolutionary theory. A gene-centred approach analyses the relationships between genotypes and phenotypes in terms of *differences* (change the genotype and observe changes in phenotype). We now know that, most frequently, this does not correctly reveal the relationships because of extensive buffering by robust networks of interactions. By contrast, understanding biological function through physiological analysis requires an *integrative* approach in which the activity of the proteins and RNAs formed from each DNA template is analysed in networks of interactions. These networks also include components that are not specified by nuclear DNA. Inheritance is not through DNA sequences alone. The selfish gene idea is not useful in the physiological sciences, since selfishness cannot be defined as an intrinsic property of nucleotide sequences independently of gene frequency, i.e. the 'success' in the gene pool that is supposed to be attributable to the 'selfish' property. It is not a physiologically testable hypothesis.

(Received 24 October 2010; accepted after revision 29 November 2010; first published online 6 December 2010)
Corresponding author D. Noble: Department of Physiology, Anatomy and Genetics, Parks Road, Oxford OX1 3PT, UK. Email: denis.noble@dpag.ox.ac.uk

Introduction

Interpreting molecular genetic information in terms of higher level functions in the organism is a major current goal in the physiological sciences, as is the reverse strategy of bottom-up reconstruction: they complement each other. Computational systems biology is one of the tools being used (Kohl & Noble, 2009; Hunter *et al.* 2011). Achieving this goal could also be a route through which physiology can reconnect with developmental and evolutionary biology. I will explain why some central aspects of neo-Darwinism (or the Modern Synthesis – in this article I am not always distinguishing between them), and their most popular expression in *The Selfish Gene* (Dawkins, 1976, 2006), form a barrier to the new synthesis required between physiology and evolutionary theory. The barrier can be removed by taking an integrative, multilevel approach in which genes and many other components of organisms that are inherited are viewed as co-operating in networks to express what we call the phenotype (Kohl *et al.* 2010 Fig. 2, reproduced here as Fig. 1 below). In this paper, 'co-operative genes' carries this sense, which should be clearly distinguished from the idea of genes 'for' co-operative behaviour used widely in ecology, animal behaviour and economics. Attributes like 'selfish' and 'cooperative' have different meanings when applied to objects or ensembles at different levels. Cooperation at the level of protein networks, for example, may occur even if the organism in which they cooperate is 'selfish' at the level of the phenotype, and vice versa. The concept of level in evolutionary theory requires careful analysis



Denis Noble is Emeritus Professor of Cardiovascular Physiology in the Department of Physiology, Anatomy and Genetics at Oxford University. Fifty years ago he published the first mathematical model of the electrical activity of the heart based on experimental measurements of ion channels. This has since been developed into the virtual heart project within the Human Physiome Project of the International Union of Physiological Sciences (IUPS). He is currently the President of IUPS. He is author of *The Music of Life* (Oxford University Press, 2006), the first popular book on systems biology, now translated into seven foreign languages.

DOI: 10.1113/jphysiol.2010.201384

(Gould, 2002; Okasha, 2006). Concepts and mechanisms do not necessarily carry through from one level to another – an important point to bear in mind also in multi-level physiology.

I start with a clarification of the relationship between neo-Darwinism, the Modern Synthesis and the selfish gene idea. Neo-Darwinism (a term introduced by the physiologist Georges Romanes (1883)) and its development (see Pigliucci & Muller, 2010*a* for the relevant history) into the Modern Synthesis (Huxley, 1942) as a gene-centred view of evolution can of course be stated without reference to the selfish gene idea. Neo-Darwinism is the term popularly used, even today, for the synthesis between Darwin's theory of evolution by natural selection and the assumption that the variations on which selection acts are produced solely or primarily by gene mutations, though the term Modern Synthesis is more correct since Romanes coined the term neo-Darwinism before Mendel's work on genetics was rediscovered. The Modern Synthesis adds discrete (Mendelian) inheritance to neo-Darwinism. Alternatives to the Modern Synthesis include: symbiogenesis, the idea that major steps in evolution, such as the formation of eukaryotes and multicellular organisms, resulted from cooperation and/or fusion between different organisms; horizontal gene transfer within and between organisms (Woese & Goldenfeld, 2009; Goldenfeld & Woese, 2011), a process now known to extend beyond prokaryotes (Keeling & Palmer, 2008); and the inheritance of acquired characteristics, commonly but mistakenly (Noble, 2010*b*) called 'Lamarckism'. For further examples see Pigliucci & Muller (2010*a*, particularly their Fig. 1.1; 2010*b*) and Jablonka & Lamb (2005).

In the rest of this article reference to neo-Darwinism should be taken to include the Modern Synthesis. The selfish gene idea (Dawkins, 1976, 2006) is a popularization of neo-Darwinism which goes beyond it to characterise genes as elements in organisms with specific (selfish) behaviour. As we will see later, it was originally formulated as a literal scientific hypothesis. The question of its status is a major focus of this paper.

Another way of stating the claims of this article is that they are twofold: first, that neo-Darwinism is, at the least, incomplete as a theory of evolution. Second, that the selfish gene idea adds nothing since it is essentially empty. These are separate claims, even though in the minds of many biologists neo-Darwinism and the selfish gene idea are not always clearly distinguished. Neo-Darwinism is capable of falsification. Indeed, in its original form as a *complete* theory, it has already been falsified. We now need to admit processes outside its remit, so that it needs to be extended (Woese & Goldenfeld, 2009; Pigliucci & Muller, 2010*b*).

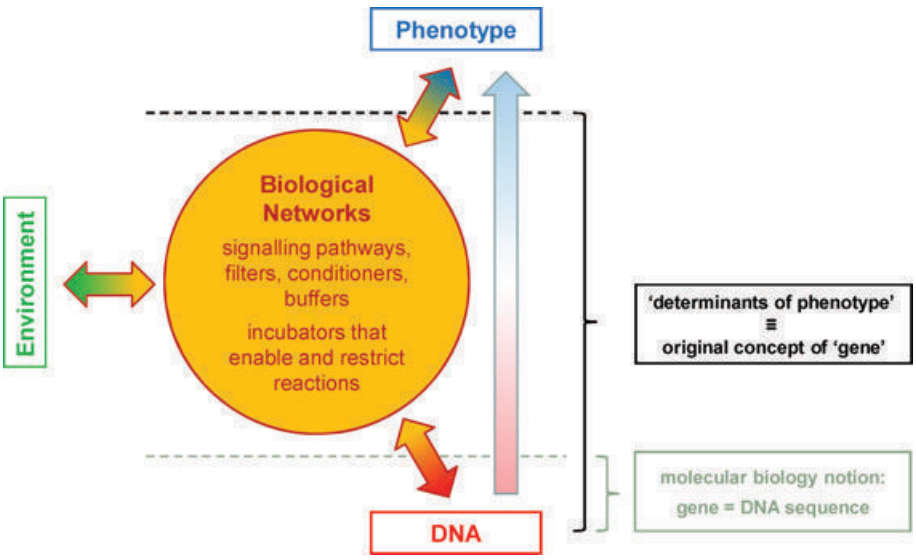


Figure 1. Relations between genes, environment and phenotype characters according to current physiological and biochemical understanding
This diagram represents the interaction between genes (DNA sequences), environment and phenotype as occurring through biological networks. The causation occurs in both directions between all three influences on the networks. This view is very different from the idea that genes 'cause' the phenotype (right hand arrow). This diagram also helps to explain the difference between the original concept of a gene as the cause of a particular phenotype and the modern definition as a DNA sequence. For further description and analysis of the ideas behind this diagram see Kohl *et al.* (2010) from which the diagram is reproduced. Reprinted by permission from Macmillan Publishers Ltd: *Clinical Pharmacology and Therapeutics* 88, 25–33; ©2010 .

As I will show in this paper, the selfish gene idea is not even capable of direct empirical falsification; it has to be judged by different criteria.

The concept of a gene has changed, and is still changing, so what version do we use?

A serious problem in assessing the nature and utility of the selfish gene story in physiological research is that the concept of a gene has changed (see Fig. 1) in fundamental ways (Pichot, 1999; Keller, 2000; Beurton *et al.* 2008). We are dealing with a moving target. From being the (hypothetical allelic) cause of each phenotype character, such as eye colour or number of limbs, the developments in molecular biology have led to its being defined more narrowly and specifically as a DNA sequence that is used by the cell as a template for the synthesis of a protein or RNA. These are not at all the same thing when it comes to questions like 'what do genes do?' and 'what kind of causation is involved?' When Johannsen (1909) introduced the term 'gene' it was *defined* as the (necessary) cause of a phenotype, since it was defined as an inherited phenotype that could be attributed to an allele. But now it has to be *shown* to be a cause, and the nature of that causation needs clarification. The full implications of this difference are explained elsewhere (Noble, 2008). They are reinforced by the fact that most changes at the level of DNA do not have a measurable phenotypic effect under normal physiological conditions (see, for example, Hillenmeyer *et al.* 2008). By the original definition, these would not even have been identified as genes, since a gene was an entity that necessarily had a phenotypic manifestation.

In this article, I frequently refer to the selfish gene idea as a story since one of the questions I am addressing is whether it is more than a story or viewpoint. Colourful metaphorical stories can be highly influential: no-one can deny that the selfish gene idea has had a huge impact on the way in which both lay people and scientists view genetics, including the social implications (Midgley, 2010). Most of the time, people accept its implied scientific basis. It is important therefore to ask whether the idea could be interpreted as an empirical scientific hypothesis, particularly since Dawkins's own initial interpretation was that it was not metaphorical; in reply to Midgley (1979) he wrote: 'that was no metaphor. I believe it is the literal truth, provided certain key words are defined in the particular ways favoured by biologists' (Dawkins, 1981). But a metaphor does not cease to be a metaphor simply because one defines a word to mean something other than its normal meaning. Indeed, it is the function of metaphor to do *precisely* this. So, we must first clarify what the idea means.

Is the 'selfish gene' story metaphor or empirical science or both?

Genes, as DNA sequences, do not of course form selves in any ordinary sense. The DNA molecule on its own does absolutely nothing since it reacts biochemically only to triggering signals. It cannot even initiate its own transcription or replication. It cannot therefore be characterised as selfish in any plausible sense of the word. If we extract DNA and put it in a Petri dish with nutrients, it will do nothing. The cell from which we extracted it would, however, continue to function until it needs to make more proteins, just as red cells function for a hundred days or more without a nucleus. It would therefore be more correct to say that genes are not active causes; they are, rather, caused to give their information by and to the system that activates them. The only kind of causation that can be attributed to them is passive, much in the way a computer program reads and uses databases. The selfish gene idea therefore has to be interpreted not only as a metaphor, but as one that struggles to chime with modern biology. That is where the difficulties begin.

Ideas that incorporate or are based on metaphors have a very different relationship to empirical discovery than do standard scientific hypotheses with clear empirical consequences that ensure their falsifiability. There are several ways in which this is evident.

First, different or even opposing metaphors can both be 'true'. This is because metaphors highlight different aspects of the target to which they are applied, a fact that has long been familiar to metaphor theorists (Lakoff & Johnson, 1980; Kittay, 1987). Metaphors can correspond to different, even incompatible, aspects of reality. That is why, when comparing 'selfish' genes with 'prisoner' or 'cooperative' genes, as I do in chapter 1 of *The Music of Life* (Noble, 2006), there is no empirical test that will unequivocally show which is correct, a point which was conceded long ago by Richard Dawkins at the beginning of his book *The Extended Phenotype*: 'I doubt that there is any experiment that could prove my claim' (Dawkins, 1982, p. 1). This point is analogous to the sense in which no experiment could ever disprove a geometry, whether Euclidean or not (Poincaré, 1902, 1968). Significantly, Dawkins uses a geometric illusion (the Necker Cube) to illustrate his point.

(*The Extended Phenotype* was an even stronger statement of the selfish gene idea since it argued that "the phenotypic effects of a gene. . .may extend far outside the body in which the gene sits" (Dawkins, 1982, p. vi) Even effects "at a distance" are seen as being "for the benefit" of the selfish gene.)

Second, metaphors often appear circular if interpreted like a scientific theory. I will show that the selfish gene metaphor shows this circularity.

Finally, even though there may be no single empirical fact that will distinguish between very different metaphors, this does not mean that empirical discovery has no impact on our choice of metaphor. The relationship is more nuanced than it may be for most scientific theories. It will usually require a judgment based on a large set of empirical facts to arrive at a conclusion. Much of the meaning associated with metaphorical statements is determined by viewpoints that are a matter of personal choice, even though influenced by empirical facts. I will illustrate this later in this paper.

What does 'selfish' mean in the selfish gene story?

First we must decide whether 'selfish' defines a property that is universal to all genes (or even all DNA sequences) or whether it is a characteristic that distinguishes some DNA sequences from others. This is not as easy as it may seem. I suspect that the original intention was that all genes could be represented as 'seeking' their own success in the gene pool, regardless of how effective they might be in achieving this. One reason for thinking this is that so-called junk DNA is represented in the selfish gene story as an arch-example of selfishness: hitching a ride even with no function.

But on that interpretation, the demonstration that the concept is of no utility in physiological science is trivially easy. Interpreted in this way, a gene cannot 'help' being selfish. That is simply the nature of any replicator. But since 'selfishness' would not itself be a difference between successful and unsuccessful genes (success being defined here as increasing frequency in the gene pool), nor between functional and non-functional genes, there would be no cashable value whatsoever for the idea in physiology. Physiologists study what makes systems work. It matters to us whether something is successful or not. Attributing selfishness to all genes therefore leaves us with nothing we could measure to determine whether 'selfishness' is a correct attribute. As metaphor, it may work. But as a scientific hypothesis it is empty.

Could we rescue the idea for physiological science? I doubt whether anyone would want to do that *ab initio*, but we live in a scientific culture that is now thoroughly permeated by the idea, and in a way that has strongly disfavoured physiology. The idea has either to be rejected or assimilated. One option would be to re-interpret selfishness to include reference to effectiveness. We could, for example, say that genes whose numbers of copies increase are selfish, or more selfish than their competitors. This move would give us an empirical handle on the idea.

It is a standard move in science to unpack a metaphor or simile in this way. Physicists make similar moves when they give empirical criteria for black holes, quarks, strings and many other strange new entities in their theories. Without an empirical handle they might as well not exist. Indeed, one of the arguments about string theory, for example, is precisely whether it has satisfied this fundamental criterion.

Moreover, including reference to effectiveness, which in evolutionary theory could be interpreted to be fitness, is surely the most relevant way to gain empirical leverage. We can measure changes in gene copies in a population. Now the question becomes whether we can develop the theory a bit further to become predictive. What, in a gene, could tell us whether or not it is selfish in this sense?

On the original definition of a gene as a hypothetical cause of a particular phenotype, this would have been fairly straightforward. We could look, at the functional level of the phenotype, for the reasons why a particular function would be adaptive. This is in practice what defenders of the selfish gene idea do. They refer to the gene (more strictly an allele) as 'the gene for' X or Y, where these are functional, phenotype characters. The phenotype view creeps back in through the terminology. Any 'selfishness' lies at least as much in the phenotype as in the genes.

But since we now define genes as particular DNA sequences, what in a DNA sequence could possibly tell us whether or not it is selfish? The answer is obvious: the sequences of Cs, Gs, As and Ts could never, by themselves, give us a criterion that would enable us to predict that the frequency of that sequence will increase in the gene pool. A DNA sequence only makes sense in the context of particular organisms in which it is involved in phenotypic characteristics which can be selected for. A sequence that may be very successful in one organism and/or environment, might be lethal in another. This is evident in the fact that almost all cross-species clones do not form an adult (see later for an important exception). The same, or similar, DNA sequence may contribute to different, even unrelated, functions in different species. The sequence, intrinsically, is neutral with regard to such functional questions.

The price therefore of giving the selfish gene idea some empirical leverage is to reveal yet again, though in a different way, that it is an empty hypothesis. There is no criterion independent of the only prediction that the hypothesis makes, i.e. that selfish genes increase their number. It is a strange hypothesis that uses its own definition of its postulated entity as its only prediction.

At this point, I suspect that a defender of the concept would shift back to referring to genes as hypothetical entities, defined as the cause(s) of particular phenotypes. Note, though, that this is to abandon the purely 'genes-eye' view since it shifts the focus back to the phenotype. As a physiologist, naturally I would say 'so it should'. I will discuss the consequences of that shift in a later section.

How is the selfish gene story related to the central dogma?

In one of the central paragraphs of *The Selfish Gene* (page 21), Dawkins writes:

> Now they swarm in huge colonies, safe inside gigantic lumbering robots, sealed off from the outside world, communicating with it by tortuous indirect routes, manipulating it by remote control. They are in you and me; they created us, body and mind; and their preservation is the ultimate rationale for our existence.

The phrase 'sealed off from the outside world' is a colourful statement of the idea that genes are uninfluenced by their environment, a view that was strongly buttressed by the central dogma of molecular biology, originally formulated by Crick (1958, 1970) and taken to exclude information flow other than from genes to proteins. In fact, of course, what the molecular biology showed was simply that amino acid sequences are not used as templates for forming nucleic acid sequences. The unjustified extension was to think that *information* cannot pass from proteins to nucleic acids, whereas this is precisely what must happen for genes to be activated and for expression patterns to be formed. This extension (which can be seen in phrases like "the inheritance of instructively acquired adaptation would violate the 'central dogma' of embryology" (Dawkins, 1982, p. 173) was a godsend to the neo-Darwinists since it provided a basis, right down at the level of DNA itself, for regarding genes as 'sealed off' from the outside world. The original experimental basis for this idea was the Weismann (1893) barrier.

A godsend, except that it is not correct in the relevant sense, and never has been. Even at the time the dogma was formulated, it was sufficient to ask the question how do different cells in the body, with exactly the same genome, end up as different as bone cells and heart cells? The answer of course is that the way in which the genome is read leads to completely different patterns of gene expression. This requires flow of information onto the genome itself, which, as Barbara McClintock (1984) said, should be regarded as an 'organ of the cell', not its dictator. There are feedbacks and restraints, not only between the products of the genes (which might be consistent with a genes-eye view), but right down onto the genome itself, determining when, where and how much of each gene product is formed. As Beurton *et al.* (2008) comment 'it seems that a cell's enzymes are capable of actively manipulating DNA to do this or that. A genome consists largely of semistable genetic elements that may be rearranged or even moved around in the genome thus modifying the information content of DNA.'

The central dogma, as a general principle of biology, has therefore been progressively undermined. The only aspect of it still left intact is its original strictly chemical sense, i.e. that protein sequences are not used as templates for forming DNA or RNA sequences. All other aspects of the way in which the dogma has been extended to buttress neo-Darwinism have been deconstructed – by molecular biology itself. Shapiro's (2009) article is the best account of the demolition from a biochemical viewpoint, while Werner (2005) does so from an informatics perspective.

Are genes the only immortals?

A central distinction in the selfish gene story is that between replicators and vehicles. The distinction is based on considering inheritance only of *changes*. While the vehicle is also 'inherited' (genes on their own do nothing and certainly are not sufficient to 'make' an organism – since we *must* also inherit a complete fertilised egg cell), the story goes that *changes* in the vehicle are not inherited (so no inheritance of acquired characteristics) while *changes* in the replicator (e.g. mutations) are inherited. This approach is what enables the wholesale inheritance of the vehicle to be ignored.

Yet, the vehicle (the cell, or each cell in a multicellular organism) clearly does reproduce (indeed, it is only through this reproduction that DNA itself is transmitted), and in doing so it passes on all the phenotype characteristics for which there are no nuclear DNA templates and which are necessary to interpret the inherited DNA. An obvious example is the transmission of mitochondria, chloroplasts and other organelles, which almost certainly originated as symbionts ('invading' or 'engulfed' bacteria) at an early stage of evolution when eukaryotes were first formed. Many other transmitted cytoplasmic factors also exist (Sun *et al.* 2005; Maurel & Kanellopoulos-Langevin, 2008). All these replicate and, in the selfish gene story would have to be given the status of 'honorary genes'.

The existence of such cellular inheritance requires the selfish gene theory to distinguish between replication and reproduction. The next step in the story is to claim that replicators are potentially immortal, whereas reproducers are not.

Biologically speaking, this is evident nonsense. Through germline cells I am connected via many reproductions to the earliest cells, *even to those without genomes*. In some sense, the cell as a whole has achieved at least equivalent immortality to that of its DNA. Cells, even those without genomes in the postulated pre-DNA world of RNA enzymes (Maynard Smith & Szathmáry, 1999), clearly reproduce themselves, and in doing so they also pass on any differences among them (Sonneborn, 1970; Sun *et al.* 2005). Any difference between replication and reproduction (which, after all, are just synonyms; the distinction is a linguistic confusion) does not entitle one to say that one is immortal and the other is not. What were all those cells without genomes doing in early life on earth? We wouldn't be here to tell the story if they

did not also form an 'immortal line'. As I have argued elsewhere (Noble, 2008) the main difference between DNA and non-DNA inheritance is simply that one is digital, the other analog. In developing the organism the 3D analog information is just as necessary as the 1D digital (DNA) information. Neither is sufficient by itself. They are mutually dependent. The amount of analog information can also be calculated to be comparable to that of the genome (Noble, 2011). Moreover, organisms are not in fact digital machines (Shapiro, 2005; Noble, 2010*a*).

The genetic differential effect problem

Clearly, many of the problems with the selfish gene story arise from unusual or imprecise use of the language of genetics, leading to untestable ideas. Another central muddle, both in neo-Darwinism and in the selfish gene story, is what I have called 'The genetic differential effect problem' (Noble, 2008, 2011), the idea that genetics is only about differences. This view is now unsustainable, since defining genes as DNA sequences clearly does identify a specific chemical entity whose effects are not merely attributable to differences in the sequence. We can say precisely for which proteins or RNAs the sequence acts as a template and analyse the physiological effects of those proteins or RNAs. The arguments for abandoning the difference perspective are overwhelming (see also Longo & Tendero, 2007).

Differences in DNA do not necessarily, or even usually, result in differences in phenotype. The great majority, 80%, of knockouts in yeast, for example, are normally 'silent' (Hillenmeyer *et al.* 2008). While there must be underlying effects in the protein networks, these are clearly buffered at the higher levels. The phenotypic effects therefore appear only when the organism is metabolically stressed, and even then they do not reveal the precise quantitative contributions for reasons I have explained elsewhere (Noble, 2011). The failure of knockouts to systematically and reliably reveal gene functions is one of the great (and expensive) disappointments of recent biology. Note, however, that the disappointment exists only in the gene-centred view. By contrast it is an exciting challenge from the systems perspective. This very effective 'buffering' of genetic change is itself an important systems property of cells and organisms.

Moreover, even when a difference in the phenotype does become manifest, it may not reveal the function(s) of the gene. In fact, it cannot do so, since *all* the functions shared between the original and the mutated gene are necessarily hidden from view. This is clearly evident when we talk of oncogenes. What we mean is that a particular change in DNA sequence predisposes to cancer. But this does not tell us the function(s) of the un-mutated gene, which would be better characterised in terms of its physiological function in, e.g., the cell cycle. Only a full physiological analysis of the roles of the protein it codes for in higher-level functions can reveal that. That will include identifying the real biological regulators as systems properties. Knockout experiments by themselves do not identify regulators (Davies, 2009).

So, the view that we can only observe *differences* in phenotype correlated with *differences* in genotype both leads to incorrect labelling of gene functions and falls into the fallacy of confusing the tip with the whole iceberg. We want to know what the relevant gene products do in the organism as a physiological whole, not simply by observing differences. Remember that most genes and their products, RNA and proteins, have multiple functions.

To see the poverty of the view that we can only observe differences, just ask the question what engineer would be satisfied simply to know the *difference* between the cement he used this time to construct his building compared to what he used previously, or to know just the differences between two electronic components in an aircraft? Of course, he might use the difference approach as one of his experimental tools (as genetics has in the past, to good effect), but the equations and models of an engineer represent the relevant totality of the function of each component of a system. So does physiological analysis of function, which is why physiology cannot be restricted to the limitations of the 'difference' approach.

Second, accurate replication of DNA is itself a system property of the cell as a whole, not just of DNA. DNA on its own is an extremely poor replicator. It requires a dedicated set of proteins to ensure correction of transcription errors and eventual faithful transmission. Both in ensuring faithfulness of DNA replication and in creating robustness against genetic defects, systems properties are the important ones. The cell as a whole 'canalises' the way in which DNA is interpreted, making it robust and reproducible. The famed 'immortality' of DNA is actually a property of the complete cell.

The distinction between replicator and vehicle is therefore out of date from a physiologist's viewpoint. It stems from the original 'genetic program' idea, in which organisms are viewed as Turing machines with the DNA being the digital tape of the computer (tape–computer is much the same distinction as replicator–vehicle – this was the basis of Jacob and Monod's concept of the 'genetic program'; Jacob, 1970). Organisms are interaction systems, not Turing machines (Shapiro, 2005; Noble, 2008). There is no clear distinction between replicator and vehicle (Coen, 1999).

Finally, the story implies that the 'vehicles' do not themselves evolve independently of their DNA. There is no reason why this should be true. In fact it is certainly false. Egg cells from different species are different. So much so that cross-species hybrids using nuclear transfer usually do not survive, and those that do, as in the elegant experiments of Sun *et al.* (2005) – see Fig. 2 –

transferring nuclei between different fish species, reveal precisely the influence of the species-specific cytoplasmic factors on development (see also Jaenisch, 2004; Yang *et al.* 2007). Crossing a common carp nucleus with a goldfish enucleated egg cell produces an adult fish that has an intermediate shape and a number of vertebrae closer to that of the goldfish. These factors can therefore determine a phenotype characteristic as fundamental as skeletal formations. Over 50 years ago, McLaren & Michie (1958) showed a similar phenomenon as a maternal effect in mice. The number of tail vertebrae (4 or 6 in the different strains) was determined by the surrogate mother, not the embryo. Of course, such cytoplasmic influences are dependent on the DNA of the mother, but these influences will necessarily include *patterns* of gene expression that are also dependent on other influences. There is interplay here between DNA and non-DNA inheritance, as there must always be. Moreover, maternal and paternal effects in response to the environment have been shown to be transmitted down two generations (grandparents to grandchildren) in humans (Pembrey *et al.* 2006) and could therefore be a target for natural selection.

Conclusions

As physiological and systems biological scientists, we need to reconnect to evolutionary theory. It was difficult to do this during most of the 20th century because the neo-Darwinist synthesis more or less excluded us, by relegating the organism to the role of a disposable vehicle. It also, unjustifiably, excluded Lamarck (Noble, 2010*b*). Darwin himself was not so sure; in the first edition of *The Origin of Species* (Darwin, 1859) he wrote 'I am convinced that natural selection has been the main, but not the exclusive means of modification', a statement he reiterated with increased force in the 1872, 6th edition. As many evolutionary biologists now acknowledge, the Modern Synthesis (neo-Darwinism) requires extending (Jablonka & Lamb, 2005; Pigliucci & Muller, 2010*b*).

If physiology is to make the contribution it should to the fields of evolution and development, we need to move on from the restrictions of the differential approach. The integrative approach can achieve this by reverse engineering using computational modelling, as I have shown elsewhere (Noble, 2011). The genes-eye view is only one way of seeing biology and it doesn't accurately reflect much of what modern biology has revealed. In fact, its central entity, the gene, 'begins to look like hardly definable temporary products of a cell's physiology' (Beurton *et al.* 2008).

Finally, I want to return to the role of metaphor and the selfish gene idea.

When I first read Richard Dawkins's acknowledgement in *The Extended Phenotype* ('I doubt that there is any experiment that could be done to prove my claim') I was strongly inclined to agree with it (both in relation to the original selfish gene idea and its development in *The Extended Phenotype*) since, if you compare the selfish gene metaphor with very different metaphors, such as genes as prisoners, it is impossible to think of an experiment that would distinguish between the two views, as I argued earlier in this paper. For any given case, I still think that must be true. But I have slowly changed my view on whether this must be true if we consider *many* cases, looking at the functioning of the organism as a whole. There are different ways in which empirical discovery can impact on our theoretical understanding. Not all of these are in the form of the straight falsification of a hypothesis, a point that has been well-understood in theoretical physics for many years (Poincaré, 1902, 1968). Sometimes it is the slow accumulation of the weight of evidence that eventually triggers a change of viewpoint. This is the case with insights that are expressed in metaphorical form (like 'selfish' and 'prisoners'), and that should not be intended to be taken literally. The first mistake of the differential approach was to interpret the selfish gene idea as literal truth. It is clearly metaphorical metaphysics, and rather poor metaphysics at that since, as we have seen, it is essentially empty as a scientific hypothesis, at least in physiological science. But in social evolution also, the idea is simply one of several viewpoints that can account for the same data (Okasha, 2010).

Figure 2. Cross-species clone
The nucleus of a common carp, *Cyprinus carpio* (middle), was transferred into the enucleated egg cell of a goldfish, *Carassius auratus* (left). The result is a cross-species clone (right) with a vertebral number closer to that of a goldfish (26–28) than of a carp (33–36) and with a more rounded body than a carp. The bottom illustrations are X-ray images of the animals in the top illustration. Figure kindly provided by Professor Yonghua Sun from the work of Sun *et al.* (2005).

The weight of evidence in the physiological sciences is now much more favourable to the metaphor of 'co-operation' than of 'selfishness'. Gene products all co-operate in robust networks one of whose functions is precisely to insulate the organism from many of the vagaries of gene mutation, and stochasticity at lower levels. Investigating these networks and their mechanisms is the way forward.

It is therefore time to move on and remove the conceptual barriers to integrating modern physiological science with evolutionary and developmental theory. The integrative approach can achieve this since it avoids the simplistic fallacies of the gene-centred differential approach and it is essentially what successful systems physiology has employed for many years.

Further reading

This article has been written for a physiological readership that may not be very familiar with the current debates in evolutionary and genetic theory. If you learnt evolutionary biology and genetics a decade or more ago you need to be aware that those debates have moved on very considerably, as has the experimental and field work on which they are based. Amongst the references cited, the following may help the reader to catch up: Margulis (1998); Jablonka & Lamb (2005); Noble (2006); Okasha (2006); Beurton *et al.* (2008); Shapiro (2009); Pigliucci & Müller (2010*b*). For those interested in the philosophical and social impacts of the metaphors used, Midgley (2010) gives a very readable account.

References

Beurton PJ, Falk R & Rheinberger H-J (ed) (2008). *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge University Press, Cambridge, UK.

Coen E (1999). *The Art of Genes*. Oxford University Press, Oxford.

Crick FHC (1958). On protein synthesis. *Symp Soc Exp Biol* **12**, 138–163.

Crick FHC (1970). Central dogma of molecular biology. *Nature* **227**, 561–563.

Darwin C (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Davies J (2009). Regulation, necessity, and the misinterpretation of knockouts. *Bioessays* **31**, 826–830.

Dawkins R (1976, 2006). *The Selfish Gene*. Oxford University Press, Oxford.

Dawkins R (1981). In defence of selfish genes. *Philosophy* **56**, 556–573.

Dawkins R (1982). *The Extended Phenotype*. Freeman, Oxford and San Francisco.

Goldenfeld N & Woese C (2011). Life is physics: Evolution as a collective phenomenon far from equilibrium. *Annu Rev Cond Matt Phys* **2** (in press).

Gould SJ (2002). *The Structure of Evolutionary Theory*, (see ch. 8, especially pp. 673–714). Harvard University Press, Cambridge, MA, USA.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C & Giaever G (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365.

Hunter P, Smaill BH, Smith NP, Young A, Nash M, Nielsen PF, Vaughan-Jones RD, Omholt S & Paterson DJ (2011). The Heart Physiome Project. *WIRE Syst Biol Med* (in press).

Huxley JS (1942). *Evolution: The Modern Synthesis*. Allen & Unwin, London.

Jablonka E & Lamb M (2005). *Evolution in Four Dimensions*. MIT Press, Cambridge, MA, USA.

Jacob F (1970). *La Logique du vivant, une histoire de l'hérédité*. Gallimard, Paris.

Jaenisch R (2004). Human cloning – the science and ethics of nuclear transplantation. *New Engl J Med* **351**, 2787–2791.

Johannsen W (1909). *Elemente der Exakten Erblichkeitslehre*. Gustav Fischer, Jena.

Keeling PJ & Palmer JD (2008). Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**, 605–618.

Keller EF (2000). *The Century of the Gene*. Harvard University Press, Cambridge, MA, USA.

Kittay EF (1987). *Metaphor: Its Cognitive Force and Linguistic Structure*. Oxford University Press, Oxford.

Kohl P, Crampin E, Quinn TA & Noble D (2010). Systems biology: an approach. *Clin Pharmacol Ther* **88**, 25–33.

Kohl P & Noble D (2009). Systems biology and the virtual physiological human. *Mol Syst Biol* **5**, 291–296.

Lakoff G & Johnson M (1980). *Metaphors We Live By*. University of Chicago Press, Chicago.

Longo G & Tendero P-E (2007). The differential method and the causal incompleteness of programming theory in molecular biology. *Found Sci* **12**, 337–366.

Margulis L (1998). *Symbiotic Planet: A New Look at Evolution*. Basic Books, New York.

Maurel M-C & Kanellopoulos-Langevin C (2008). Heredity – venturing beyond genetics. *Biol Reprod* **79**, 2–8.

Maynard Smith J & Szathmáry E (1999). *The Origins of Life*. Oxford University Press, New York.

McClintock B (1984). The significance of responses of the genome to challenge. *Science* **226**, 792–801.

McLaren A & Michie D (1958). An effect of uterine environment upon skeletal morphology of the mouse. *Nature* **181**, 1147–1148.

Midgley M (1979). Gene-juggling. *Philosophy* **54**, 439–458.

Midgley M (2010). *The Solitary Self. Darwin and The Selfish Gene*. Acumen, Durham.

Noble D (2006). *The Music of Life*. Oxford University Press, Oxford.

Noble D (2008). Genes and causation. *Philos Transact A Math Phys Eng Sci* **366**, 3001–3015.

Noble D (2010*a*). Biophysics and systems biology. *Philos Transact A Math Phys Eng Sci* **368**, 1125–1139.

Noble D (2010*b*). Letter from Lamarck. *Physiol News* **78**, 31.

Noble D (2011). Differential and integral views of genetics in computational systems biology. *J Roy Soc Interface Focus* **1**, 7–15.

Okasha S (2006). *Evolution and the Levels of Selection*. Oxford University Press, Oxford.

Okasha S (2010). Altruism researchers must cooperate. *Nature* **467**, 653–655.

Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjostrom M, Golding J & ALSPAC_study_team (2006). Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* **14**, 159–166.

Pichot A (1999). *Histoire de la Notion de Gène*. Flammarion, Paris.

Pigliucci M & Müller GB (2010*a*). Elements of an extended evolutionary synthesis. In *Evolution: The Extended Synthesis*, ed. Pigliucci M & Muller GB, pp. 3–17. MIT Press, Cambridge, MA, USA.

Pigliucci M & Müller GB (ed) (2010*b*). *Evolution*: *The Extended Synthesis*. MIT Press, Cambridge, MA, USA.

Poincaré H (1902, 1968). *La science et l'hypothèse*. Flammarion, Paris.

Romanes GJ (1883). Letter to the Editor. *Nature* **27**, 528–529.

Shapiro JA (2005). A 21st century view of evolution: genome system architecture, repetitive DNA, and natural genetic engineering. *Gene* **345**, 91–100.

Shapiro JA (2009). Revisiting the central dogma in the 21st century. *Ann N Y Acad Sci* **1178**, 6–28.

Sonneborn TM (1970). Gene action on development. *Proc R Soc Lond B Biol Sci* **176**, 347–366.

Sun YH, Chen SP, Wang YP, Hu W & Zhu ZY (2005). Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. *Biol Reprod* **72**, 510–515.

Weismann A (1893). *The Germ-Plasm: A Theory of Heredity*. Charles Scribner's Sons, New York.

Werner E (2005). Genome semantics. In silico multicellular systems and the central dogma. *FEBS Lett* **579**, 1779–1782.

Woese CR & Goldenfeld N (2009). How the micobial world saved evolution from the scylla of molecular biology and the charybdis of the modern synthesis. *Microbiol Mol Biol Rev* **73**, 14–21.

Yang X, Smith SL, Tian XC, Lewin HA, Renard J-P & Wakayama T (2007). Nuclear reprogramming of cloned embryos and its implications for therapeutic cloning. *Nat Genet* **39**, 295–302.

Acknowledgements

I should like to acknowledge long and deep discussions with the organisers of the Balliol College, Oxford seminars on conceptual foundations of Systems Biology: Jonathan Bard, Tom Melham and Eric Werner; and the organisers and participants of the 'Homage to Darwin' debate (http://www.voicesfromoxford.com/homagedarwin_part1.html) held in Oxford in May 2009: Stephen Bell, Martin Brasier, Richard Dawkins and Lynn Margulis. I received criticism of early drafts of this paper from David Vines, David Cleevely, Nicholas Beale and Stig Omholt. I also acknowledge discussions with Peter Kohl, Ray Noble and James Shapiro. Providing valuable input and feedback does not of course signify assent to the claims of my paper. I consulted on a wide range of opinion. Work in the author's laboratory is funded by the PreDiCT project of the European Union under FP7.

Interface Focus

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

Review

A theory of biological relativity: no privileged level of causation

Denis Noble*

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

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

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

1. INTRODUCTION

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

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

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

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

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

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

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

Received 1 July 2011
Accepted 17 October 2011

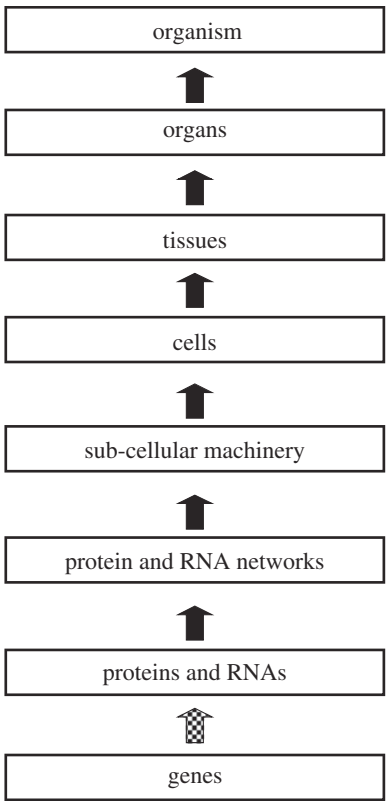


Figure 1. Upward causation: the reductionist causal chain in biology. This is a gross simplification, of course. No one today seriously believes that this diagram represents all causation in biology. Reductive biological discourse, however, privileges this form of causation and regards it as the most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

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

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

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

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

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

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

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

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

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

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

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

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

4. INSIGHTS FROM EXPERIMENTAL AND MODELLING WORK ON HEART CELLS

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

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

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

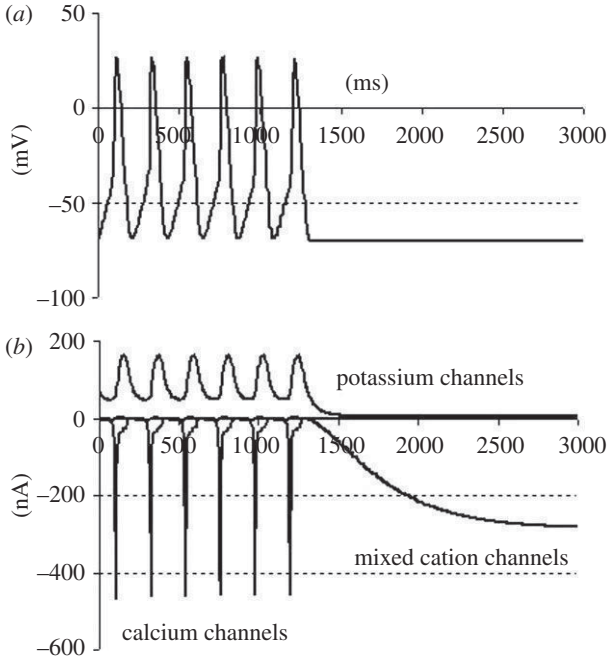


Figure 2. Computer model of pacemaker rhythm in the heart [27]. For the first six beats, the model is allowed to run normally and generates rhythm closely similar to a real cell. Then the feedback from cell voltage (*a*) to protein channels ((*b*) currents in nanoamps) is interrupted by keeping the voltage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

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

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

5. GENERALIZATION OF THE ARGUMENT IN MATHEMATICAL TERMS

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

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

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

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

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

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

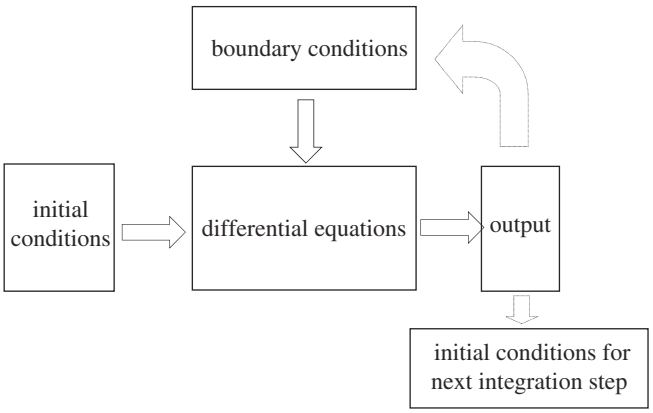


Figure 3. Many models of biological systems consist of differential equations for the kinetics of each component. These equations cannot give a solution (the output) without setting the initial conditions (the state of the components at the time at which the simulation begins) and the boundary conditions.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

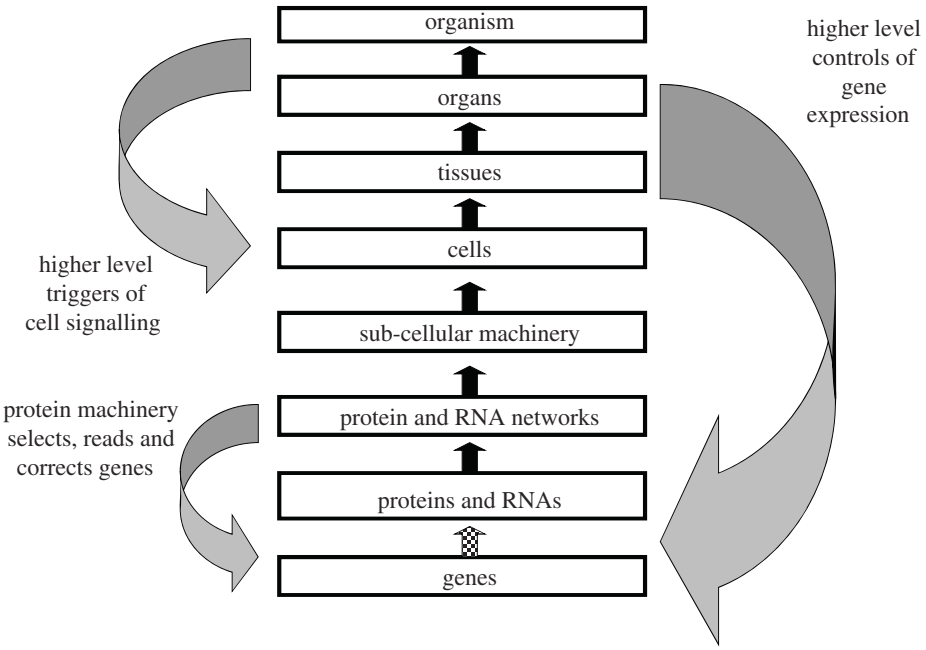


Figure 4. The completion of figure 1 with various forms of downward causation that regulates lower level components in biological systems. In addition to the controls internal to the organism, we also have to take account of the influence of the environment on all the levels (not shown in this diagram).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

familiar with the basic ideas on multi-level control ever since Claude Bernard formulated the concept of control of the internal environment in his book *Introduction à l'étude de la médecine expérimentale* in 1865 [32] and Walter B. Cannon developed the idea of homeostasis in *The wisdom of the Body* in 1932 [33]. So, how has mainstream biology tended to ignore it, as has physiology also with some exceptions, for example Guyton's modelling of the circulation [34]? I think the main culprit here has been neo-Darwinism and particularly the popularizations of this theory as a purely gene-centric view [31].

The essential idea of gene-centric theories is what I have called the differential view of the relationships between genes and phenotypes [35–38]. The idea is essential in the sense that it excludes alternative theories by arguing that what matters in evolutionary terms are *changes* in the genotype that are reflected in *changes* in phenotype. Selection of the phenotype is therefore, according to this logic, fundamentally equivalent to selection of particular genes (or, more strictly, gene alleles). This view might have been appropriate for a time when genes were regarded as hypothetical entities defined as

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

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

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

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

7. A (BIOLOGICAL) THEORY OF RELATIVITY

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

8. CONNECTING LEVELS

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

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

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

9. EMERGENCE AND BOUNDARY CONDITIONS

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

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

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

10. COMPUTABILITY

The first is the problem of computability.

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

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

11. SCALE RELATIVITY

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

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

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

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

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

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

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

12. CONCLUSIONS

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

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

REFERENCES

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

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

59 Winslow, R., Varghese, A., Noble, D., Adlakha, C. & Hoythya, A. 1993 Generation and propagation of triggered activity induced by spatially localised Na-K pump inhibition in atrial network models. *Proc. R. Soc. Lond. B* **254**, 55–61. (doi:10.1098/rspb.1993.0126)

60 Nash, M. P. & Hunter, P. J. 2001 Computational mechanics of the heart. *J. Elast.* **61**, 113–141. (doi:10.1023/A:1011084330767)

61 Smith, N. P., Pullan, A. J. & Hunter, P. J. 2001 An anatomically based model of transient coronary blood flow in the heart. *SIAM J. Appl. Math.* **62**, 990–1018. (doi:10.1137/S0036139999359860)

62 Brenner, S. 2010 Sequences and consequences. *Phil. Trans. R. Soc. B* **365**, 207–212. (doi:10.1098/rstb.2009.0221)

63 Feytmans, E., Noble, D. & Peitsch, M. 2005 Genome size and numbers of biological functions. *Trans. Comput. Syst. Biol.* **1**, 44–49. (doi:10.1007/978-3-540-32126-2_4)

64 Lewontin, R. C. 1974 *The genetic basis of evolutionary change.* New York, NY: Columbia University Press.

65 Siegelmann, H. T. 1995 Computation beyond the Turing limit. *Science* **268**, 545–548. (doi:10.1126/science.268.5210.545)

66 Garny, A., Noble, D. & Kohl, P. 2005 Dimensionality in cardiac modelling. *Progr. Biophys. Mol. Biol.* **87**, 47–66. (doi:10.1016/j.pbiomolbio.2004.06.006)

67 Nottale, L. 1993 *Fractal space-time and microphysics: towards a theory of scale relativity.* Singapore: World Scientific.

68 Nottale, L. 2000 *La relativité dans tous ses états. Du mouvements aux changements d'échelle.* Paris, France: Hachette.

69 Nottale, L. 2011 *Scale relativity and fractal space-time: a new approach to unifying relativity and quantum mechanics.* London, UK: Imperial College Press.

70 Auffray, C. & Nottale, L. 2008 Scale relativity theory and integrative systems biology. I. Founding principles and scale laws. *Progr. Biophys. Mol. Biol.* **97**, 79–114. (doi:10.1016/j.pbiomolbio.2007.09.002)

71 Nottale, L. & Auffray, C. 2008 Scale relativity and integrative systems biology. II. Macroscopic quantum-type mechanics. *Progr. Biophys. Mol. Biol.* **97**, 115–157. (doi:10.1016/j.pbiomolbio.2007.09.001)

72 Hinch, R., Greenstein, J. L., Tanskanen, A. J. & Xu, L. 2004 A simplified local control model of calcium-induced calcium release in cardiac ventricular myocytes. *Biophys. J.* **87**, 3723–3736. (doi:10.1529/biophysj.104.049973)

73 Hinch, R., Greenstein, J. L. & Winslow, R. L. 2006 Multi-scale modelling of local control of calcium induced calcium release. *Progr. Biophys. Mol. Biol.* **90**, 136–150. (doi:10.1016/j.pbiomolbio.2005.05.014)

74 Tanskanen, A. J., Greenstein, J. L., Chen, A., Sun, X. & Winslow, R. L. 2007 Protein geometry and placement in the cardiac dyad influence macroscopic properties of calcium-induced calcium release. *Biophys. J.* **92**, 3379–3396. (doi:10.1529/biophysj.106.089425)

75 Einstein, A. 1936 Lens-like action of a star by the deviation of light in the gravitational field. *Science* **84**, 506–507. (doi:10.1126/science.84.2188.506)

76 Petters, A. O., Levine, H. & Wambsganss, J. 2001 *Singularity theory and gravitational lensing.* Boston, MA: Birkhäuser.

77 Bitbol, M. In press. Downward causation without foundations. *Synthese.* (doi:10.1007/s11229-010-9723-5)

President's Lecture

Physiology is rocking the foundations of evolutionary biology

Denis Noble

Department of Physiology, Anatomy & Genetics, Oxford, UK

New Findings

- **What is the topic of this review?**
  Have recent experimental findings in evolutionary biology concerning the transmission of inheritance opened the way to a reintegration of physiology with evolutionary biology?
- **What advances does it highlight?**
  The answer is yes, and that this requires a new synthesis between evolutionary theory and experimental physiology.

The 'Modern Synthesis' (Neo-Darwinism) is a mid-20th century gene-centric view of evolution, based on random mutations accumulating to produce gradual change through natural selection. Any role of physiological function in influencing genetic inheritance was excluded. The organism became a mere carrier of the real objects of selection, its genes. We now know that genetic change is far from random and often not gradual. Molecular genetics and genome sequencing have deconstructed this unnecessarily restrictive view of evolution in a way that reintroduces physiological function and interactions with the environment as factors influencing the speed and nature of inherited change. Acquired characteristics can be inherited, and in a few but growing number of cases that inheritance has now been shown to be robust for many generations. The 21st century can look forward to a new synthesis that will reintegrate physiology with evolutionary biology.

(Received 27 February 2013; accepted after revision 9 April 2013; first published online 12 April 2013)
Corresponding author D. Noble: Department of Physiology, Anatomy & Genetics, Sherrington Building, Parks Road, Oxford OX1 3PT, UK. Email: denis.noble@dpag.ox.ac.uk

Introduction

As 2012 came to a close, an article appeared in the *Proceedings of the National Academy of Sciences of the United States of America* with a title that would have been inconceivable in such a prestigious journal only 5–10 years ago. 'Rocking the foundations of molecular genetics' (Mattick, 2012) is a commentary on a ground-breaking original experimental article (Nelson *et al.* 2012) in the same issue of the journal showing epigenetic maternal inheritance over several generations. My title echoes that of Mattick, but it also goes further. It is not only the standard 20th century views of molecular genetics that are in question. Evolutionary theory itself is already in a state of flux (Jablonka & Lamb, 2005; Noble, 2006, 2011; Beurton *et al.* 2008; Pigliucci & Müller, 2010; Gissis & Jablonka, 2011; Shapiro, 2011). In this article, I will show that all the central assumptions of the Modern Synthesis (often also called Neo-Darwinism) have been disproved. Moreover, they have been disproved in ways that raise the tantalizing prospect of a totally new synthesis; one that would allow a reintegration of physiological science with evolutionary biology. It is hard to think of a more fundamental change for physiology and for the conceptual

This article is based on the President's Lecture at the IUPS Congress, Birmingham, UK on 21 July 2013.

DOI: 10.1113/expphysiol.2012.071134

foundations of biology in general (Melham *et al.* 2013). The Modern Synthesis (Fisher, 1930; Huxley, 1942; Mayr, 1982) attributed genetic change solely to chance events, about which physiology could say very little. The germ line was thought to be isolated from any influence by the rest of the organism and its response to the environment, an idea that was encapsulated in the Weismann barrier (Weismann, 1893). Note that this was animal specific and did not apply to other life forms. But if acquired changes can be inherited through many generations, then physiology becomes relevant again, because it is precisely the study of function and functional changes. These are what determine epigenetic processes.

I start with some definitions. I will use the term 'Modern Synthesis' rather than 'Neo-Darwinism'. Darwin was far from being a Neo-Darwinist (Dover, 2000; Midgley, 2010), so I think it would be better to drop his name for that idea. As Mayr (1964) points out, there are as many as 12 references to the inheritance of acquired characteristics in *The Origin of Species* (Darwin, 1859) and in the first edition he explicitly states 'I am convinced that natural selection has been the main, but not the exclusive means of modification', a statement he reiterated with increased force in the 1872, 6th edition. In some respects, my article returns to a more nuanced, less dogmatic view of evolutionary theory (see also Müller, 2007; Mesoudi *et al.* 2013), which is much more in keeping with the spirit of Darwin's own ideas than is the Neo-Darwinist view.

Summary of the Modern Synthesis

The central assumptions of the Modern Synthesis that are relevant to this article are fourfold (see also the summary by Koonin, 2011).

First, genetic change is random. Interpreted in modern terms as referring to DNA, the changes can be thought of as restricted to single step changes in one (or a very few) bases brought about, for instance, by copying errors, radiation or any other random event. The concept of a purely random event is not easy to define. The physicochemical nature of biological molecules will, in any case, ensure that some changes are more likely to happen than others. Randomness cannot therefore be defined independently of asking 'random with respect to what'? I will use the definition that the changes are assumed to be random with respect to physiological function and could not therefore be influenced by such function or by functional changes in response to the environment. This is the assumption that excludes the phenotype from in any way influencing or guiding genetic change.

Second, genetic change is gradual. Since random events are best thought of as arising from microscopic stochasticity, it will generally be the case that many such events would have to accumulate to generate a major change in genome and phenotype. Of course, there are point mutations that can have a dramatic effect on the phenotype, but these are rare. The prediction would be that the evolution of gene sequences and the amino acid sequences of the proteins formed should not occur in ways that would require large domains to move around within and between genomes.

Third, following genetic change, natural selection leads to particular gene variants (alleles) increasing in frequency within the population. Those variants are said to confer an advantage in terms of fitness on the individuals concerned, which therefore increasingly dominate the population. By this process and other mechanisms, including genetic drift and geographic isolation, new species can arise.

Fourth, the inheritance of acquired characteristics is impossible. This is the main thrust of the synthesis and it is the means by which Darwin's ideas were represented as distinct from those of Lamarck (1994, originally published 1809). This assumption also excludes any notion of what Lamarck called 'le pouvoir de la vie', a life force that could in some way be seen as directing evolution through increasing complexity or through adaptation. Lamarckism was excluded not only by the experiments of Weismann (1893) but also by the central dogma of molecular biology (Crick, 1970). Both claim that the genetic material is isolated from the organism and its environment; 'sealed off from the outside world', to use *The Selfish Gene* popularization of the idea (Dawkins, 1976, 2006).

All these assumptions have been disproved in various ways and to varying degrees, and it is also important to note that a substantial proportion of the experimental work that has revealed these breaks has come from within molecular biology itself. Molecular biology can now be seen to have systematically deconstructed its own dogmas (Shapiro, 2009, 2011).

Are mutations random?

'It is difficult (if not impossible) to find a genome change operator that is truly random in its action within the DNA of the cell where it works. All careful studies of mutagenesis find statistically significant non-random patterns of change, and genome sequence studies confirm distinct biases in location of different mobile genetic elements' (Shapiro, 2011, p. 82). Shapiro gives large numbers of references on the non-random nature of mutations. As already noted, though, the key question is not so much whether changes are truly random (there can be no such thing independent of context) but whether they are chance events from the viewpoint of function. The evidence is that both the speed and the location of genome change can be influenced functionally. Changes in the speed of change are well known already from the way in which genome change occurs in immunological processes. The germ line has only a finite amount of DNA. In order to react to many different antigens, lymphocytes 'evolve' quickly

to generate extensive antigen-binding variability. There can be as many as 10^{12} different antibody specificities in the mammalian immune system, and the detailed mechanisms for achieving this have been known for many years. The mechanism is directed, because the binding of the antigen to the antibody itself activates the proliferation process. Antigen activation of B-cell proliferation acts as a selective force. The targeting of the genomic changes, which maintains the functional structure of the antibody while diversifying antigen recognition, occurs by protein–DNA binding specificity (VDJ joining; Shapiro, 2011, p. 173), coupling to transcription signals (somatic hypermutation) and lymphokine-directed transcription of heavy chain switch regions (class switch recombination; Shapiro, 2011, pp. 66–69).

Similar targeted genomic changes occur outside the context of the immune system. The reader is referred to table II.7 (Shapiro, 2011, pp. 70–74; http://shapiro.bsd.uchicago.edu/TableII.7.shtml) for many examples of the stimuli that have been shown to activate this kind of 'natural' genetic engineering, while table II.11 from the same book (pp. 84–86; http://shapiro.bsd.uchicago.edu/TableII.11.shtml) documents the regions of the genomes targeted. Thirty-two examples are given. One example will suffice to illustrate this. P element homing in fruit flies involves DNA transposons that insert into the genome in a functionally significant way, according to the added DNA. There is up to 50% greater insertion into regions of the genome that are related functionally to DNA segments included within the P element. Thus, 'Insertion of a binding sequence for the transcriptional regulator Engrailed targets a large fraction of insertions to chromosomal regions where Engrailed is known to function.' (Shapiro, 2011, p. 83). A possible explanation is that the donor element and the target site may be brought close together in the nucleus, i.e. organization of the genome is important. This kind of information is also therefore 'genetic'. We should not limit the concept of a 'gene' and the description 'genetic' to protein-template regions of the genome, particularly as we now know that 80% of the non-protein regions are transcribed, although it is uncertain how much is functional (http://www.genome.gov/10005107; http://genome.ucsc.edu/ENCODE/). It was clearly premature to label this DNA as 'junk'. Structural organization also represents information that is transmitted down the generations. DNA is not merely a one-dimensional sequence. It is a highly complex physiological system that is regulated by the cells, tissues and organs of the body. This will become even clearer in the next section.

Is genetic change gradual?

It was the Nobel Prize-winner Barbara McClintock who introduced the idea that the genome is 'an organ of the cell' (McClintock, 1984). She won her prize for physiology or medicine in 1983 over 40 years after she had made the ground-breaking discovery of chromosome transposition (now called mobile genetic elements). She worked on maize, and early reactions to her work were so sceptical that she stopped publishing her research in 1953 (Keller, 1983). The consequences for evolutionary theory were also ignored, because the phenomenon was not thought to occur in animals. We now know that animal genomes are full of transposons. About 3500 of the estimated 26,000 human protein-template regions contain exons originating from mobile elements (Shapiro, 2011, p. 109). This contrasts with a much lower number, 1200, in mice, even though the number of protein template regions is similar in both genomes. This suggests that transposons may have played a major role in primate and human evolution. Over two-thirds of the human genome is derived from mobile elements (de Koning *et al.* 2011), and there have been well over 3 million transposition events in its evolution.

McClintock could not have anticipated the evidence that would later emerge from whole-genome sequencing studies in various species, but it fully vindicates the general and widespread significance of her discovery. The *Nature* 2001 report (International Human Genome Mapping Consortium, 2001) compared protein-template regions for several classes of proteins from yeast, nematode worms, *Drosophila*, mice and humans. In the case of transcription factors (Figure 45 of the *Nature* report) and chromatin-binding proteins (Figure 42 of the *Nature* report) the evidence shows that whole domains up to hundreds of amino acids in length have been amplified and shifted around among different genetic loci in the genome. Of course, the sequencings were done on the contemporary species. We do not therefore know precisely when in the evolutionary process the transpositions may have occurred. However, a number of the domains and combinations are restricted to certain lineages. And of course, gradual changes also occurred within the sequences. The experimental evidence on genome sequencing shows multiple ways in which evolutionary change has occurred. Note also that domain shuffling and the polyphyletic origins of genomes were established facts well before the full sequencing of genomes (Gordon, 1999; Shapiro, 2011).

The mechanisms of transposable elements illustrate one of the important breaks with the central dogma of molecular biology. Retrotransposons are DNA sequences that are first copied as RNA sequences, which are then inserted back into a different part of the genome using reverse transcriptase. DNA transposons may use a cut-and-paste mechanism that does not require an RNA intermediate. As Beurton *et al.* (2008) comment, 'it seems that a cell's enzymes are capable of actively manipulating DNA to do this or that. A genome consists largely of

semi-stable genetic elements that may be rearranged or even moved around in the genome thus modifying the information content of DNA.' The central dogma of the 1950s, as a general principle of biology, has therefore been progressively undermined until it has become useless as support for the Modern Synthesis (Werner, 2005; Mattick, 2007; Shapiro, 2009) or indeed as an accurate description of what happens in cells. As Mattick (2012) says, 'the belief that the soma and germ line do not communicate is patently incorrect.'

An important point to note is the functionally significant way in which this communication can occur. In bacteria, starvation can increase the targeted transposon-mediated reorganizations by five orders of magnitude, i.e. by a factor of over 100,000 (Shapiro, 2011, p. 74).

Mobile transposable elements that have been involved in evolution come in more forms than only retrotransposons and DNA transposons. They include the movement and/or fusion of whole genomes between species. Symbiogenesis is the mechanism by which eukaryotes developed from prokaryotes, with mitochondria and chloroplasts being the most well-known examples, having originated as bacteria that invaded (or were engulfed by) the 'parent' cell (Margulis, 1981; Brown & Doolittle, 1997; Margulis & Sagan, 2003). During evolution, some of the acquired DNA transferred to the nucleus. Horizontal transfer of DNA is ubiquitous in the prokaryote world, but also far from absent amongst eukaryotes (Shapiro, 2011). Other forms of mobile DNA include plasmids, viruses and group II introns, which are all prokaryotic elements. To these we can add group I introns and inteins (Raghavan & Minnick, 2009), multiple classes of transposons (Curcio & Derbyshire, 2003), multiple classes of retrotransposons (Volff & Brosius, 2007) and various forms of genomic DNA derived from reverse transcription (Baertsch *et al.* 2008). One of the major developments of Darwin's concept of a 'tree of life' is that the analogy should be more that of a 'network of life' (Doolittle, 1999; Woese & Goldenfeld, 2009). As with other breaks from the Modern Synthesis, that synthesis emerges as only part of the evolutionary story.

The inheritance of acquired characteristics

In 1998, the great contributor to the development of the Modern Synthesis, John Maynard Smith, made a very significant and even prophetic admission when he wrote 'it [Lamarckism] is not so obviously false as is sometimes made out' (Maynard Smith, 1998), a statement that is all the more important from being made by someone working within the Modern Synthesis framework. The time was long overdue for such an acknowledgement. Nearly 50 years before, Waddington had written 'Lamarck is the only major figure in the history of biology whose name has become to all extents and purposes, a term of abuse. Most scientists' contributions are fated to be outgrown, but very few authors have written works which, two centuries later, are still rejected with an indignation so intense that the skeptic may suspect something akin to an uneasy conscience. In point of fact, Lamarck has, I think, been somewhat unfairly judged.' (Waddington, 1954).

So why, given his extraordinary (but completely correct) admission, did Maynard Smith not revise his view of the mechanisms of evolution? The reason he gave in 1999 was that 'it is hard to conceive of a mechanism whereby it could occur; this is a problem' (Maynard Smith, 1999). At that time, the examples of the inheritance of acquired characteristics could be counted on the fingers of one hand. They included Waddington's work on genetic assimilation (Waddington, 1959) and Sonneborn's work on the inheritance of non-genetic changes in *Paramecium* membrane–cilia orientation (Sonneborn, 1970). The flow of papers during the last 5 years showing non-Mendelian inheritance is, however, now becoming a flood of evidence. Sadly, Maynard Smith is no longer with us to comment on this important development. Let us try, though, to look at the evidence through his eyes, because although he saw a problem, he also added that it was 'not I think insuperable' (Maynard Smith, 1999).

The examples he had in 1998 were not only few and relatively old, they were also fairly easy to assimilate into the Modern Synthesis or ignore as special cases. Waddington's work could be dismissed, because it was not certain that no mutations were involved, although this would be very unlikely on the time scale of his experiments. Any variation that was necessary was almost certainly already present in the gene pool. His work on fruit flies essentially consisted in selecting for certain combinations of existing DNA sequences in the population gene pool by selective breeding from flies with unusual phenotypes induced by treating embryos with heat or ether (Bard, 2008). He was the first to call this mechanism 'epigenetics' (i.e. over and above genetics), but he did not mean the specific form that we now understand by that term, i.e. the marking of chromatin to change the patterns of expression.

The Modern Synthesists should not have dismissed Waddington's experiments, for example, as simply 'a special case of the evolution of phenotypic plasticity' (Arthur, 2010). Of course, the Modern Synthesis can account for the inheritance of the potential for plasticity, but what it cannot allow is the inheritance of a specific acquired form of that plasticity. Waddington's experiments demonstrate precisely inheritance of specific forms of acquired characteristics, as he claimed himself in the title of his paper (Waddington, 1942). After all, the pattern of the genome is as much inherited as its individual components, and those patterns can be determined by the environment.

But I can see why Modern Synthesists thought the way they did. Giving up such a central tenet of the Synthesis

would have been difficult anyway, not least because of the extraordinary distinction of the 20th century biologists who developed it. We are talking, after all, of Julian Huxley, Sewell Wright, J. B. S. Haldane, R. A. Fisher, George Price and Bill Hamilton, to name but a few. Waddington's genetic assimilation process was discounted as a break with the Modern Synthesis precisely because it did not involve gradual accumulations of mutations and was not viewed as a challenge to that process. But that is to put the question the wrong way round. It is precisely whether gradual mutations form the only mechanism that is in question. Waddington's work was a proven alternative additional mechanism. Even 70 years ago, the Modern Synthesis could have been admitted to be incomplete.

In a different way, Sonneborn's work was brushed aside as being on a unicellular organism, with no separate germ line. The Modern Synthesis has always had a strongly zoological basis, tending to ignore prokaryotes, unicellular organisms and plants, even though these cover more than 80% of the whole duration of the evolutionary process long before 'zoology' could even have a meaning in evolutionary history.

But the evidence for the inheritance of acquired characteristics has now moved right into the zoological domain. All the remaining examples I shall quote here are on multicellular organisms, including mammals, and they refer to pioneering work done in the last 7 years.

Anway *et al.* (2006*a*,*b*) demonstrated that an endocrine disruptor, vinclozolin (an anti-androgenic compound), can induce transgenerational disease states or abnormalities that are inherited for at least four generations in rats. The transmission is via epigenetic modifications carried by the male germ line and may involve either marking of the genome or transmission of RNAs. More recent work from the same laboratory has shown that the third generation granulosa cells carry a transgenerational effect on the transcriptome and epigenome through differential DNA methylation (Nilsson *et al.* 2012). The sperm nucleus contains much more than the genome (Johnson *et al.* 2011).

An alternative approach to determining how the organism as a whole may influence the genome and whether such influences can be transmitted transgenerationally is to study cross-species clones, e.g. by inserting the nucleus of one species into the fertilized but enucleated egg cell of another species. Following the gene-centric view of the Modern Synthesis, the result should be an organism determined by the species from which the genome was taken. In the great majority of cases, this does not happen. Incompatibility between the egg cytoplasm and the transferred nuclear genome usually results in development freezing or completely failing at an early stage. That fact already tells us how important the egg cell expression patterns are. The genome does not succeed in completely dictating development regardless of the cytoplasmic state. Moreover, in the only case where this process has resulted in a full adult, the results also do not support the prediction. Sun *et al.* (2005) performed this experiment using the nucleus of a carp inserted into the fertilized but enucleated egg cell of a goldfish. The adult has some of the characteristics of the goldfish. In particular, the number of vertebrae is closer to that of the goldfish than to that of a carp. This result echoes a much earlier experiment of McLaren and Michie, who showed an influence of the maternal uterine environment on the number of tail vertebrae in transplanted mice embryos (McLaren & Michie, 1958). Many maternal effects have subsequently been observed, and non-genomic transmission of disease risk has been firmly established (Gluckman & Hanson, 2004; Gluckman *et al.* 2007). A study done in Scandinavia clearly shows the transgenerational effect of food availability to human grandparents influencing the longevity of grandchildren (Pembrey *et al.* 2006; Kaati *et al.* 2007).

Epigenetic effects can even be transmitted independently of the germ line. Weaver and co-workers showed this phenomenon in rat colonies, where stroking and licking behaviour by adults towards their young results in epigenetic marking of the relevant genes in the hippocampus that predispose the young to showing the same behaviour when they become adults (Weaver *et al.* 2004; Weaver, 2009). (This field is growing so rapidly that there is not space in this review to cover it. A more extensive bibliography can be found at http://shapiro.bsd.uchicago.edu/Transgenerational_Epigenetic_Effects.html.)

Molecular mechanisms

The results I have described so far establish the existence of transgenerational non-Mendelian inheritance. This section describes recent studies that demonstrate the molecular biological mechanisms and that the transmission can be robust for many generations.

Rechavi *et al.* (2011) worked on *Caenorhabditis elegans* and the non-Mendelian inheritance of the worm's response to viral infection. This is achieved by the infection inducing the formation of an RNA silencer. They crossed worms with this response with worms that do not have it and followed the generations until they obtained worms that did not have the DNA required to produce the silencing RNA but which nevertheless had inherited the acquired resistance. The mechanism is that transmission of RNA occurs through the germ line and is then amplified by using RNA polymerase. The inheritance of the acquired characteristic is robust for over 100 generations.

The work of Nelson *et al.* (2012) that stimulated Mattick's article in *Proceedings of the National Academy of Sciences of the United States of America*, with which I began this review, is from the laboratory of Joe Nadeau

at the Institute of Systems Biology in Seattle. Their article begins by noting that many environmental agents and genetic variants can induce heritable epigenetic changes that affect phenotypic variation and disease risk in many species. Moreover, these effects persist for many generations and are as strong as conventional genetic inheritance (Richards, 2006; Jirtle & Skinner, 2007; Youngson & Whitelaw, 2008; Cuzin & Rassoulzadegan, 2010; Nelson & Nadeau, 2010; Guerrero-Bosagna & Skinner, 2012). The challenge now is to understand their molecular basis. The experiments of Nelson and co-workers were on the Deadend1 (*Dnd1*) gene, which enhances susceptibility to testicular germ cell tumours in mice, in part by interacting epigenetically with other testicular germ cell tumour modifier genes in previous generations. They showed that genetically engineered deficiency of Apobec1 modifies susceptibility, either alone or in combination with *Dnd1*, and either in a conventional or a transgenerational manner. The heritable epigenetic changes persisted for multiple generations and were fully reversed after consecutive crosses through the alternative germ lineage. The Apobec family is an unusual protein family of cytidine deaminases that can insert mutations in DNA and RNA (Conticello, 2008).

A further example of a molecular mechanism is that of paramutation, which consists in the interaction between two alleles at a single locus. This can induce permanent epigenetic changes in organisms from maize to mice (Chandler, 2007, 2010; Cuzin *et al.* 2008; Sidorenko *et al.* 2009; Arteaga-Vazquez *et al.* 2010; Erhard & Hollick, 2011).

These examples of robust inheritance of acquired characteristics reveal a wide array of mechanisms by which such inheritance can be achieved. Nature seems to work through the cracks, as it were, of the gene-centric view. Those cracks have now been discovered to be great fissures, through which functionally significant inherited changes occur. Such mechanisms could not have been foreseen at the time when the Modern Synthesis was formulated, or even a decade ago. To Maynard Smith's (1999) comment ('it is hard to conceive of a mechanism whereby it could occur'), the reply must be that some of those mechanisms have now been found and they are robust.

In addition to establishing the molecular mechanisms, these experiments help to explain an otherwise puzzling finding. Conventional genetic inheritance often accounts for $<10\%$ of observed inherited risk. Similar conclusions have been drawn from genome-wide association studies and from studies on identical twins (Roberts *et al.* 2012). This observation, in itself, creates problems for the gene-centric view, and it is now clear that non-Mendelian inheritance may provide a large part of the explanation (Slatkin, 2009).

What went wrong in the mid-20th century that led us astray for so long? The answer is that all the way from the Weismann barrier experiments in 1893 (which were very crude experiments indeed) through to the formulation of the central dogma of molecular biology in 1970, too much was claimed for the relevant experimental results, and it was claimed too dogmatically. Demonstrating, as Weismann did, that cutting the tails off many generations of mice does not result in tail-less mice shows, indeed, that this particular induced characteristic is not inherited, but it obviously could not exclude other mechanisms. The mechanisms found recently are far more subtle. Likewise, the demonstration that protein sequences do not form a template for DNA sequences should never have been interpreted to mean that information cannot pass from the organism to its genome. Barbara McClintock deservedly gets the last laugh; the genome is indeed an 'organ of the cell'.

Table 1. Comparison between the Modern Synthesis and the proposed Integrative Synthesis

Before: Modern Synthesis	Now: towards an Integrative Synthesis
Gene-centred view of natural selection	Selection is multilevel
Impossibility of inheritance of acquired characteristics	Acquired characters can be inherited
Distinction between replicator (genes) and vehicle (phenotype)	The genome is an 'organ of the cell', not its dictator. Control is distributed
The central dogma of molecular biology	Genomes are not isolated from organism and environment

Towards a new synthesis between physiology and evolutionary biology?

This review has been written for a primarily physiological audience, but its implications are profound for biological science in general. It shows that, through recent discoveries on the inheritance of acquired characteristics, the analysis of physiological function can be important to the mechanisms of evolutionary change. The full extent of this feedback from function to inheritance remains to be assessed, but it cannot be doubted that it runs counter to the spirit of the Modern Synthesis. The challenge now is how to construct a new Synthesis to take account of this development. In Table 1, I call this the Integrative Synthesis. I believe that in the future, the Modern Synthesis and the elegant mathematics that it gave rise to, for example in the various forms and developments of the Price equation, will be seen as only one of the processes involved, a special case in certain circumstances, just as Newtonian mechanics remains as a special case in the theory of relativity. The mathematics of evolutionary theory is developing to take additional processes into account (e.g. Bonduriansky & Day, 2009; Slatkin, 2009;

Nowak *et al.* 2010). In many cases, that is already implicit, for example where the 'gene' is really an inherited phenotype regardless of the mechanism of inheritance. Where the mechanism matters, for instance in allowing blending rather than discrete inheritance, the mathematics will be interestingly different. There are also important implications for the rate of evolutionary change, because an adaptive characteristic may be acquired by many individuals simultaneously, thus avoiding the slow process of a chance mutation in an individual spreading through the population.

A central feature of the Integrative Synthesis is a radical revision of the concept of causality in biology. *A priori* there is no privileged level of causation. This is the principle that I have called the theory of biological relativity (Noble, 2008, 2012). As Werner puts it, 'all levels have an equal contributing value' (Werner, 2003). Control is therefore distributed, some of which is inherited independently of DNA sequences. The revision of the concept will also recognize the different forms of causality. DNA sequences are best viewed as passive causes, because they are used only when the relevant sequences are activated. DNA on its own does nothing. The active causes lie within the control networks of the cells, tissues and organs of the body.

Conclusions

We are privileged to live at a time of a major change in the conceptual foundations of biology. That change is set to bring the physiological study of function right back into centre stage. It is worth quoting the relevant paragraph from Mattick's commentary on the work of Nelson *et al.* (2012):

> The available evidence not only suggests an intimate interplay between genetic and epigenetic inheritance, but also that this interplay may involve communication between the soma and the germline. This idea contravenes the so-called Weismann barrier, sometimes referred to as Biology's Second Law, which is based on flimsy evidence and a desire to distance Darwinian evolution from Lamarckian inheritance at the time of the Modern Evolutionary Synthesis. However, the belief that the soma and germline do not communicate is patently incorrect.

The only parts of this statement that I would change are, first, to remind readers, as I noted earlier in this article, that Darwin himself did not exclude the inheritance of acquired characteristics and, second, to remind us that Lamarck himself did not invent 'Lamarckism' (Noble, 2010). As we move on beyond the unnecessary restrictions of the Modern Synthesis we move back towards a more genuinely 'Darwinian' viewpoint and we also move towards a long-overdue rehabilitation of Lamarck. Of course, neither Darwinism nor Lamarckism remains unchanged. Neither could have anticipated the work of the 21st century. But we can now see the Modern Synthesis as too restrictive and that it dominated biological science for far too long. Perhaps the elegant mathematics and the extraordinary reputation of the scientists involved blinded us to what now seems obvious; the organism should never have been relegated to the role of mere carrier of its genes.

References

Anway MD, Leathers C & Skinner MK (2006*a*). Endocrine disruptor vinclozolin induced epigenetic transgenerational adult-onset disease. *Endocrinology* **147**, 5515–5523.

Anway MD, Memon MA, Uzumcu M & Skinner MK (2006*b*). Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *J Androl* **27**, 868–879.

Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC & Chandler VL (2010). RNA-mediated *trans*-communication can establish paramutation at the *b1* locus in maize. *Proc Natl Acad Sci U S A* **107**, 12986–12991.

Arthur W (2010). *Evolution, a Developmental Approach*. Wiley Blackwell, Oxford, UK.

Baertsch R, Diekhans M, Kent WJ, Haussler D & Brosius J (2008). Retrocopy contributions to the evolution of the human genome. *BMC Genomics* **9**, 466.

Bard JBL (2008). Waddington's legacy to developmental and theoretical biology. *Biological Theory* **3**, 188–197.

Beurton PJ, Falk R & Rheinberger H.-J. (2008). *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge University Press, Cambridge, UK.

Bonduriansky R & Day T (2009). Non-genetic inheritance and its evolutionary implications. *Annual Review of Ecology, Evolution, and Systematics* **40**, 103–125.

Brown JR & Doolittle WF (1997). *Archaea* and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* **61**, 456–502.

Chandler VL (2007). Paramutation: from maize to mice. *Cell* **128**, 641–645.

Chandler VL (2010). Paramutation's properties and puzzles. *Science* **330**, 628–629.

Conticello SG (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol* **9**, 229.

Crick FHC (1970). Central dogma of molecular biology. *Nature* **227**, 561–563.

Curcio MJ & Derbyshire KM (2003). The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol* **4**, 865–877.

Cuzin F, Grandjean V & Rassoulzadegan M (2008). Inherited variation at the epigenetic level: paramutation from the plant to the mouse. *Curr Opin Genet Dev* **18**, 193–196.

Cuzin F & Rassoulzadegan M (2010). Non-Mendelian epigenetic heredity: gametic RNAs as epigenetic regulators and transgenerational signals. *Essays Biochem* **48**, 101–106.

Darwin C (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Dawkins R (1976). *The Selfish Gene*. OUP, Oxford, UK.
de Koning AP, Gu W, Castoe TA, Batzer MA & Pollock DD (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **7**, e1002384.
Doolittle WF (1999). Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128.
Dover G (2000). *Dear Mr Darwin. Letters on the Evolution of Life and Human Nature*. Weidenfeld and Nicolson, London. UK.
Erhard KF Jr & Hollick JB (2011). Paramutation: a process for acquiring trans-generational regulatory states. *Curr Opin Plant Biol* **14**, 210–216.
Fisher RA (1930). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK.
Gissis SB & Jablonka E, eds (2011). *Transformations of Lamarckism. From Subtle Fluids to Molecular Biology*. MIT Press, Cambridge, MA, USA.
Gluckman P & Hanson M (2004). *The Fetal Matrix. Evolution, Development and Disease*. Cambridge University Press, Cambridge, UK.
Gluckman PD, Hanson MA & Beedle AS (2007). Non-genomic transgenerational inheritance of disease risk. *Bioessays* **29**, 145–154.
Gordon M (1999). The concept of monophyly: a speculative essay. *Biology and Philosophy* **14**, 331–348.
Guerrero-Bosagna C & Skinner MK (2012). Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Mol Cell Endocrinol* **354**, 3–8.
Huxley JS (1942). *Evolution: the Modern Synthesis*. Allen & Unwin, London, UK.
International Human Genome Mapping Consortium (2001). A physical map of the human genome. *Nature* **409**, 934–941.
Jablonka E & Lamb M (2005). *Evolution in Four Dimensions*. MIT Press, Boston, MA, USA.
Jirtle RL & Skinner MK (2007). Environmental epigenomics and disease susceptibility. *Nat Rev Genet* **8**, 253–262.
Johnson GD, Lalancette C, Linnemann AK, Leduc F, Boissonneault G & Krawetz SA (2011). The sperm nucleus: chromatin, RNA, and the nuclear matrix. *Reproduction* **141**, 21–36.
Kaati G, Bygren LO, Pembrey M & Sjöström M (2007). Transgenerational response to nutrition, early life circumstances and longevity. *Eur J Hum Genet* **15**, 784–790.
Keller EF (1983). *A Feeling for the Organism: The Life and Work of Barbara McClintock*. W.H. Freeman San Francisco, CA, USA.
Koonin EV (2011). The fundamentals of evolution: Darwin and Modern Synthesis. In *Logic of Chance: The Nature and Origin of Biological Evolution*, ed. Koonin EV, FT Press Upper Saddle River, NJ, Chapter **1**, pp 14–17.
Lamarck J-B (1994). *Philosophie Zoologique, original edition of 1809 with introduction by Andre Pichot*. Flammarion, Paris, France.
McClintock B (1984). The significance of responses of the genome to challenge. *Science* **226**, 792–801.
McLaren A & Michie D (1958). An effect of uterine environment upon skeletal morphology of the mouse. *Nature* **181**, 1147–1148.
Margulis L (1981). *Symbiosis in Cell Evolution*. W.H. Freeman Co., London, UK.
Margulis L & Sagan D (2003). *Acquiring Genomes: A Theory of the Origins of Species*. Basic Books, New York, NY, USA.
Mattick JS (2007). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann N Y Acad Sci* **1178**, 29–46.
Mattick JS (2012). Rocking the foundations of molecular genetics. *Proc Natl Acad Sci U S A* **109**, 16400–16401.
Maynard Smith J (1998). *Evolutionary Genetics*. Oxford University Press, New York, NY, USA.
Maynard Smith J (1999). The 1999 Crafoord Prize Lectures. The idea of information in biology. *Q Rev Biol* **74**, 395–400.
Mayr E (1964). Introduction. *The Origin of Species*. Harvard, Cambridge, MA, USA.
Mayr E (1982). *The Growth of Biological Thought*. Harvard, Cambridge, MA, USA.
Melham T, Bard J, Werner E, Noble D. (2013). Foreword: conceptual foundations of systems biology. *Prog Biophys Mol Biol* **111**, 55–56.
Mesoudi A, Blanchet S, Charmentier A, Danchin E, Fogarty L, Jablonka E, Laland KN, Morgan TJH, Mueller GB, Odling-Smee FJ & Pojol B. (2013). Is non-genetic inheritance just a proximate mechanism? A corroboration of the extended evolutionary synthesis. *Biological Theory* **7**, 189–195.
Midgley M (2010). *The Solitary Self. Darwin and The Selfish Gene*. Acumen, Durham, UK.
Müller GB (2007). Evo–devo: extending the evolutionary synthesis. *Nat Rev Genet* **8**, 943–949.
Nelson VR, Heaney JD, Tesar PJ, Davidson NO & Nadeau JH (2012). Transgenerational epigenetic effects of *Apobec1* deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc Natl Acad Sci U S A* **109**, E2766–E2773.
Nelson VR & Nadeau JH (2010). Transgenerational genetic effects. *Epigenomics* **2**, 797–806.
Nilsson E, Larsen G, Manikkam N, Guerrero-Bosagna C, Savenkova MI & Skinner MK (2012). Environmentally induced epigenetic transgenerational inheritance of ovarian disease. *PLoS One* **7**, e36129.
Noble D (2006). *The Music of Life*. Oxford University Press, Oxford, UK.
Noble D (2008). Genes and causation. *Philos Transact A Math Phys Eng Sci* **366**, 3001–3015.
Noble D (2010). Letter from Lamarck. *Physiol News* **78**, 31.
Noble D (2011). Neo-Darwinism, the Modern Synthesis, and selfish genes: are they of use in physiology? *J Physiol* **589**, 1007–1015.
Noble D (2012). A Theory of biological relativity: no privileged level of causation. *Interface Focus* **2**, 55–64.
Nowak MA, Tarnita CE & Wilson EO (2010). The evolution of eusociality. *Nature* **466**, 1057–1062.
Pembrey ME, Bygren LO, Kaati G, Edvinsson S, Northstone K, Sjöström M & Golding J; ALSPAC Study Team (2006). Sex-specific, male-line transgenerational responses in humans. *Eur J Hum Genet* **14**, 159–166.
Pigliucci M & Müller GB (2010). *Evolution – The Extended Synthesis*. MIT Press, Cambridge, MA, USA.

Raghavan R & Minnick MF (2009). Group I introns and inteins: disparate origins but convergent parasitic strategies. *J Bacteriol* **191**, 6193–6202.
Rechavi O, Minevish G & Hobert O (2011). Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248–1256.
Richards EJ (2006). Inherited epigenetic variation – revisiting soft inheritance. *Nat Rev Genet* **7**, 395–401.
Roberts NJ, Vogelstein JT, Parmigiani G, Kinzler KW, Vogelstein B & Velculescu VE (2012). The predictive capacity of personal genome sequencing. *Sci Transl Med* **4**, 133ra58.
Shapiro JA (2009). Revisiting the central dogma in the 21st century. *Annal N Y Acad Sci* **1178**, 6–28.
Shapiro JA (2011). *Evolution: a View from the 21st Century*. Pearson Education Inc., Upper Saddle River, NJ, USA.
Sharma A (2013). Transgenerational epigenetic inheritance: focus on soma to germline information transfer. *Prog Biophys Mol Biol*, doi: 10.1016/j.pbiomolbio.2012.12.003.
Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Viyas M, Kermicle J, Jurcin D, Brzeski J, Cai J & Chandler VL (2009). A dominant mutation in *mediator of paramutation2*, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. *PLoS Genet* **5**, e1000725.
Slatkin M (2009). Epigenetic inheritance and the missing heritability problem. *Genetics* **182**, 845–850.
Sonneborn TM (1970). Gene action in development. *Proc R Soc Lond B Biol Sci* **176**, 347–366.
Sun YH, Chen SP, Wang YP, Hu W & Zhu ZY (2005). Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. *Biol Reprod* **72**, 510–515.
Volff JN & Brosius J (2007). Modern genomes with retro-look: retrotransposed elements, retroposition and the origin of new genes. *Genome Dyn* **3**, 175–190.
Waddington CH (1942). Canalization of development and the inheritance of acquired characteristics. *Nature* **150**, 563–565.
Waddington CH (1954). Evolution and epistemology. *Nature* **173**, 880–881.
Waddington CH (1959). Canalization of development and genetic assimilation of acquired characteristics. *Nature* **183**, 1654–1655.
Weaver ICG (2009). Life at the interface between a dynamic environment and a fixed genome. In *Mammalian Brain Development*. ed. Janigro D, pp. 17–40. Humana Press, Springer, New York, NY, USA.
Weaver ICG, Cervoni N, Champagne FA, D'Alessio AC, Sharma S, Seckl JR, Dymov S, Szyf M & Meaney MJ (2004). Epigenetic programming by maternal behavior. *Nat Neurosci* **7**, 847–854.
Weismann A (1893). *The Germ-Plasm: A Theory of Heredity*. Charles Scribner's Sons, New York, NY, USA.
Werner E (2003). *In silico* multicellular systems biology and minimal genomes. *Drug Discov Today* **8**, 1121–1127.
Werner E (2005). Genome semantics, in silico multicellular systems and the Central Dogma. *FEBS Lett* **579**, 1779–1782.
Woese CR & Goldenfeld N (2009). How the micobial world saved evolution from the Scylla of molecular biology and the Charybdis of the modern synthesis. *Microbiol Mol Biol Rev* **73**, 14–21.
Youngson NA & Whitelaw E (2008). Transgenerational epigenetic effects. *Annu Rev Genom Human Genet* **9**, 233–257.

Additional information

Competing interests

None declared.

Acknowledgements

This article is based on lectures given in New Delhi, India, in December 2011 (http://www.appicon2011.org/), Suzhou, China, in November 2012 (http://www.voicesfromoxford.org/video/physiology-and-the-revolution-in-evolutionary-biology/184), the Rupert Riedl lecture at the University of Vienna (http://medienportal.univie.ac.at/uniview/veranstaltungen/detailansicht/artikel/rupert-riedl-lecture-the-music-of-life/) in March 2013, and the forthcoming President's Lecture at the IUPS Congress in the UK in July 2013 (http://www.iups2013.org/). I would like to thank Jonathan Bard, Nicholas Beale, Richard Boyd, Georges Christé, Dario DiFrancesco, Malcolm Gordon, Gerhard Müller, Raymond Noble, David Paterson, Etienne Roux, James Shapiro, Ania Sher, Eric Werner and Michael Yudkin for valuable discussions, some of whom gave specific feedback on this article. Further relevant reading can be found in two focused issues of *Progress in Biophysics and Molecular Biology* (see Melham *et al.* 2013; Sharma, 2013).

J Physiol 592.11 (2014) pp 2237–2244

The Journal of Physiology

EDITORIAL

Evolution evolves: physiology returns to centre stage

Denis Noble[1], Eva Jablonka[2], Michael J. Joyner[3], Gerd B. Müller[4] and Stig W. Omholt[5]

[1] *University of Oxford, Department of Physiology, Anatomy and Genetics, Oxford, UK*

[2] *Tel Aviv University, Cohn Institute for the History and Philosophy of Science and Ideas, Ramat Aviv, Israel*

[3] *Mayo Clinic, Rochester, MN, USA*

[4] *University of Vienna, Department of Theoretical Biology, Vienna, Austria*

[5] *Norwegian University of Science and Technology, Faculty of Medicine, Trondheim, Norway*

Email: denis.noble@physiol.ox.ac.uk

Introduction

This issue of *The Journal of Physiology* is devoted to the integration of evolutionary biology with physiological science. The immediate trigger was a very successful symposium on this theme held during the IUPS Congress in Birmingham in July 2013. The symposium followed an opening plenary lecture based on an article that had recently been published by one of us in the sister journal *Experimental Physiology* (Noble, 2013) and previously in *The Journal of Physiology* (Noble, 2011). The title of that article was ambitious, describing physiology as 'rocking the foundations' of biology. Strong language, perhaps? Yes, but that title was merely reflecting a rising tide of recently published articles in major scientific journals, including *Nature Reviews Genetics* (Müller, 2007), *Proceedings of the National Academy of Sciences of the USA* (Mattick, 2012), *Nature* (Ball, 2013), *Biological Journal of the Linnean Society* (Bateson, 2014) and *Science* (Rosenberg & Queitsch, 2014). It was also prompted by important books that have appeared recently (Margulis & Sagan, 2003; Jablonka & Lamb, 2014; Noble, 2006; Beurton *et al.* 2008; Pigliucci & Müller, 2010; Bateson & Gluckman, 2011; Gissis & Jablonka, 2011; Shapiro, 2011). Those books also propose either significant extensions of existing evolutionary theory or the replacement of the Modern Synthesis by a new synthesis. Despite the radical presentation of the *Experimental Physiology* article, therefore, it contains little that was not already known to those biologists who have been keeping abreast of recent literature. It is becoming increasingly difficult to keep up with this literature because it is widely spread amongst very many scientific journals. A focused issue of a journal, like this one, can therefore be very valuable. We intend that this should be a seminal resource for future research and teaching.

The questions addressed in the papers published here include the following.

- What are the major new developments in evolutionary biology and how do they challenge the Modern Synthesis?
- Which of these developments have implications for how the physiological sciences should further their understanding of health and disease?
- If the Modern Synthesis is to be extended or replaced by a new explanatory structure, what is the role of physiology in the development of this structure?

Function

Why have these questions become important? One answer is that they change the way in which physiological function is relevant to evolutionary biology. We define function here as the role that a part, a process or a mechanism plays within an encompassing system, a role that contributes to the goal-directed behaviour of that system. This definition covers different notions, such as those presented by Wright (1973), Cummins (1975) and Kitcher (1993). There is a possible confusion in discussing function in the context of evolution because current utility is not necessarily how the trait evolved. Further reading on these issues can be found in the articles by Tinbergen (1963), Bateson & Laland (2013) and the one in this issue by Roux (2014).

We are also using a broad definition of physiology as a discipline at the intersection of ecology, behavioural biology, developmental biology and molecular biology. As will be evident in the articles of this focused issue, the new developments encompass all these fields, often in combination.

In standard selection theory, usually called the Modern Synthesis (MS) and sometimes called Neo-Darwinism, function is relevant only to postgenomic change in populations through determining which individuals are successful in reproducing. One of the dogmas of the Modern Synthesis is the impossibility of the inheritance of acquired developmental dispositions. Genomic change, which is seen within the MS framework as a synonym to hereditary change, is assumed to be random with respect to function. Function therefore plays a role only in so far as it determines the fitness of the individual organism in its reproductive success after genomic mutations have created the possibility of an advantage. In contrast, the inheritance of some acquired epigenetic characteristics and other forms of non-DNA inheritance enables function to be involved in pre-genomic change by influencing hereditary change more directly before selection could play a role. Furthermore, mechanisms of genomic change have been identified that were not envisaged by the founders of the Modern Synthesis, including symbiogenesis and natural genetic engineering.

Making a categorical prohibition a central part of a theory can be useful for a time. The Modern Synthesis served an important function in the mid-20th century in stimulating much mathematical work in population genetics, for example. But we have to recognize that by encouraging a dogmatic use of the theory it may also have inhibited many lines of research that have now been found to be important. Theories with categorical prohibitions court their own demise, requiring either fundamental extensions or even complete replacement when contrary experimental evidence emerges. The articles in this issue demonstrate that evidence. The mechanism of random change followed by selection becomes only one of many possible mechanisms of evolutionary change. Moreover, all those mechanisms can interact. We have entered a period of a systems approach to evolution science that contrasts markedly with the parsimonious reductionism of the Modern Synthesis. In this respect, it echoes the move towards a systems approach in many other areas of biology (Melham *et al.* 2013).

DOI: 10.1113/jphysiol.2014.273151

The genotype–phenotype relation

The genotype–phenotype relation, which is at the heart of our view of heredity and development, has turned out to be much more subtle than what the Modern Synthesis made room for, and it is increasingly acknowledged that a better understanding of this relation is key to understanding a range of evolutionary phenomena beyond the explanatory reach of the Modern Synthesis. Considering that the disciplinary goals of physiology are 'the study of the functions and activities of living matter (as of organs, tissues, or cells) as such and of the physical and chemical phenomena involved' (*Webster's Third New International Dictionary*), it is clear that the mechanistic aspects of the genotype–phenotype relation lie within the explanatory domain of physiology. Hence, physiology must of necessity become the backbone of any mature evolutionary theory pretending to merge the proximate and ultimate explanatory domains. The consequence is that we will have to go back to a broader, more inclusive view of heredity, which was captured by William Bateson's original definition of genetics as 'The Physiology of Descent' (Bateson, 1906; see Olby, 2000). A physiological view of heredity enables the integration of the extended evolutionary synthesis view of evolution with the physiological sciences.

More specifically, the genotype–phenotype concept that is currently in wide use within evolutionary theory conceals the facts that it is an abstraction of a relation that is the outcome of very complex dynamics that in many cases are intimately connected to the environment (Gjuvsland *et al.* 2013), and that DNA does not have the privileged place in the chain of causality many attribute to it. As described in more detail by Omholt (2013), if one tries to interpret the function of DNA in systemic terms one finds that DNA allows a system to induce perturbations of its own dynamics as a function of the system's own state (its phenome). In this systems view, the causality flows from the system state through a change in use of DNA that results in a change in the production of RNA and protein, which in turn perturbs the system's dynamics. In those cases where variations in DNA cause changes in the perturbation regimen, it may lead to different system dynamics and thus physiological variation. Thus, the genotype–phenotype relation cannot be understood outside a systems-physiology framework, whatever causes variations in DNA. And any evolutionary theory aiming to explain the manifestation of biological form across time and space needs to be highly articulate about this relation.

Physiology in a broad sense, therefore, now moves to centre stage in evolutionary biology as we are finally in a position to step conceptually and technologically out of the narrow frames of the Modern Synthesis and take explanatory responsibility for a much wider set of evolutionary phenomena and patterns across time and space. Some of the articles in this issue address the consequences that this new intellectual spotlight has for the discipline of physiology itself, including possible consequences for health and disease; it is noteworthy that some of the new mechanisms manifest themselves in the inheritance of the chances of acquired disease states.

The ways in which a systems approach can be applied to the complex dynamics and evolution of organisms are addressed in this issue by Badyaev (2014), who explores 'whether epigenetic effects facilitate adaptive modulation of complex phenotypes by effectively reducing the dimensionality of their deterministic networks'; Baverstock & Rönkkö (2014), who regard the cell 'as a complex dissipative natural process' that 'minimizes the free energy of their ecosystems', a process where genetic variation is largely irrelevant; Jaeger & Monk (2014) showing 'how dynamical systems theory can provide a unifying conceptual framework for evolution of biological regulatory systems'; Lamm (2014), who 'applies the conceptual toolkit of Evolutionary Developmental Biology (evo–devo) to the evolution of the genome and the role of the genome in organism development'; Levin (2014), who analyses 'the control of anatomy by bioelectricity and the evolutionary implications of its top-down causal efficacy'; and Danchin & Pocheville (2014), who discuss the ways in which 'non-genetic inheritance shatters the frontier between physiology and evolution'.

Mechanisms of inheritance

The molecular mechanisms by which non-standard inheritance can occur are diverse.

Natural genetic engineering refers to reorganization of genomes. The mechanisms discovered since McClintock (1950, 1984) first demonstrated mobile genetic elements in plants are many. As Beurton *et al.* (2008) write, 'it seems that a cell's enzymes are capable of actively manipulating DNA to do this or that. A genome consists largely of semi-stable genetic elements that may be rearranged or even moved around in the genome thus modifying the information content of DNA.' In this issue, Shapiro (2014) shows that 'the genome is best modelled as a read–write (RW) data storage system rather than a read-only memory (ROM)'.

Symbiogenesis has been involved in the most dramatic examples of genome re-organization, i.e. the acquisition of DNA from other organisms through lateral gene transfer. As is now well known, this is thought to explain the origin of mitochondria, chloroplasts and other organelles.

Lateral gene transfer is now recognized to be much more extensive and widespread than it was previously assumed to be; occurring in most orders and often among them. Recent examples include mechanisms of transfer from prokaryotes to eukaryotes generally (Redrejo-Rodríguez *et al.* 2012) and transfer from bacteria to insects (Acuña *et al.* 2012).

Epigenetic mechanisms that lead to persistent developmentally induced changes in gene activity include diverse processes and factors. One type of system, the chromatin marking system, includes methylation of cytosines and histone modifications, which interact with each other and with other epigenetic control factors (such as small RNAs). Chromatin marks were originally thought to be wiped clean during transmission between generations. It is now clear that this is not always true. Moreover, recent work has shown 'heritable epigenetic changes [that] persisted for multiple generations and were fully reversed after consecutive crosses through the alternative germ-lineage' (Nelson *et al.* 2012). For example, induced epigenetic (methylation) changes affecting a wide range of characteristics were transmitted for three generations following ancestral exposure to fungicides (e.g. Anway *et al.* 2006), and conditioned fear to an odorant was transmitted for two generations in mice (Dias & Ressler, 2014). Transmission of epigenetic variations through the germ line is, however, not necessary for inheritance between generations. Chromatin marks can be transmitted across generations

by epigenetically marking the genome in the newborn, leading, through their physiological and behavioural effects, to the reconstruction of developmental conditions in the offspring (Weaver, 2009). Such genomic marking may also underlie inherited maternal (Gluckman *et al.* 2007) and nutritional effects (Kaati *et al.* 2007). Another non-standard inheritance system, the RNAi-mediated inheritance system, which interacts with the chromatin marking mechanisms, underlies the transmission of many important characteristics in both plants and animals. An example of RNA-transmitted resistance to viruses has been shown to be transmitted stably for 100 generations in nematodes (Rechavi *et al.* 2011). In this issue, Stern *et al.* (2014) demonstrate that 'exposure to [antibiotic] stress reduces the maternal levels of *Polycomb* in the offspring embryos and [that] this reduction contributes to the inheritance of induced expression'. Also in this issue, Bateson *et al.* (2014) discuss a form of developmental plasticity, the predictive adaptive response (PAR), 'in which cues received in early life influence the development of a phenotype that is normally adapted to the environmental conditions of later life'. Sela *et al.* (2014) suggest 'that non-coding RNAs synchronize the different transgenerational epigenetic effects by interacting with and therefore surveying both the transcriptome and the genome'.

The physiological adjustment of organisms to changes in conditions within and between generations involves corresponding epigenetic changes. Selection for the stabilization of the physiological adjustments can lead both to the selection of epigenetic changes that are inherited between generations and/or to the selection of genetic changes that further stabilize, expand or otherwise improve the physiological adjustments. This process, genetic assimilation, was first demonstrated by Waddington (1957), who also introduced the term 'epigenetics', though not with its current usage. A more inclusive term, 'genetic accommodation', was suggested by Mary-Jane West-Eberhard (2003). This process can lead to the stabilization and canalization of previous developmentally induced changes, to an increase in plasticity and to the buffering of potentially deleterious side-effects. In all cases, the processes are usually initiated by developmental changes that induce new patterns of gene activity in alleles that already exist in the population (but not in that combination in any individual) and expose the new allelic combination to natural selection. No new mutations are required in this process, although a new mutation can contribute to it. Given that it is gene combinations and developmental networks that are the targets of selection, genetic accommodation is yet another process showing the advantages of focusing on networks of interactions rather than on individual 'genes' (we return to the definition of 'gene' later). Thinking through the process of genetic accommodation requires consideration of the interactions between different developmental mechanisms at different levels of biological organization. Following genetic accommodation, the inheritance becomes standard DNA inheritance; therefore, it would be difficult to determine from genomic sequencing whether this process had occurred. However, comparisons of chromatin marking and small RNA profiles in populations that are at the initial stages of evolutionary divergence can uncover the epigenetic correlates of the physiological adjustments that drive genetic assimilation and can point to epigenetic factors that are inherited and contribute to the stabilization of the new adjustments. Further valuable insights on these questions can be found in the article in this issue by Bateson *et al.* (2014).

Physiological changes can accompany and stabilize cultural changes. Poverty and ethnic conflicts are cultural phenomena that may have long-term, heritable physiological effects. For example, young people living in developing countries in conditions of social and political insecurity, such as ongoing political conflicts, are likely to be exposed to hunger, psychological stress and toxic pollutants, which can alter their epigenetic profiles and adversely affect them and their offspring. This concern is highlighted by data from the 'Dutch Starvation Winter' of 1944–1945, which has shown that a deprived *in utero* environment can have lifelong effects, including the incidence of many chronic non-communicable diseases (Portrait *et al.* 2011; van Abeelen *et al.* 2012). Adverse effects also develop rapidly in the switch from low-calorie to high-calorie environments, as is now happening in China and India, with serious consequences in, for example, the prevalence of type 2 diabetes. The physiology of culture and of cultural inheritance emerges today as a new and urgent concern.

The neglect of physiological responsiveness may also lead to unwarranted, gene-centric, adaptationist interpretations. Organisms adapt to their environment at many levels that challenge a strict genotype-to-phenotype world view. For example, it has been suggested that positive selection pressure led to an increase in the prevalence of the EDARV370A variant of the human ectodysplasin receptor in the Han Chinese. This variant is associated with increased eccrine sweat gland function (Kamberov *et al.* 2013), and the idea is that it facilitated thermoregulation and thus survival in a warm, humid environment. This gene-centric interpretation fails to account for the fact that thermoregulation is highly adaptable in humans and that sweat rate can double with only a few weeks of heat exposure (Robinson *et al.* 1943; Wyndham, 1967).

Sun & Zhu (2014) in this issue show the limitations of the gene-centric view in the study of cross-species clones that provide 'an ideal system to study the relative role and crosstalk between egg cytoplasm and zygotic nucleus in development', emphasizing that 'the developmental process should be interpreted in a systemic way, rather than in a way that solely focuses on the role of nuclear genome.'

The question now, therefore, is not whether developmental plasticity and non-standard forms of inheritance occur but how often they occur and to what extent they contribute to evolutionary change. It is also important to incorporate these changes into mathematical models (Tal *et al.* 2010; Danchin *et al.* 2011) and to define the differences in the regulatory architecture that underlie, for example, broad and narrow sense inheritability (Wang *et al.* 2013). It will be important to assess the contribution these regulatory mechanisms may have made to the speed of evolution and how interactions between the mechanisms, such as genetic assimilation, contribute. These are all open and difficult questions. Nature is even more wondrous than the architects of the Modern Synthesis thought, and involves processes we thought were impossible.

Relevance to health and disease

The Modern Synthesis has also been a driver of biomedical research priorities and experimental diagnostic and therapeutic thinking since at least the US 'War on Cancer', which started in 1971. A

key idea was that discrete genetic and molecular dysfunction led to specific cancer phenotypes. If these could be identified and then targeted with drugs, cancer could be cured. This view is now being abandoned, and cancer is seen as a far more complex problem, involving many pathways, frequently trigged by environmental or behavioural factors, with only limited evidence for marked genetic risk in common cancers (Gatenby & Gillies, 2008; Watson, 2013). Paradoxically, successes in the War on Cancer have largely been through prevention, most notably via tobacco control.

In a similar vein, the human genome project saw a tight linkage between genotype and phenotype, with two major outcomes envisioned. For diseases with known genetic causes, cures based on gene therapy or other forms of genetic engineering would emerge. For more common non-communicable diseases, such as diabetes and heart disease, common gene variants would explain much of the lifetime risk of the disease and lead to pre-emptive medicine. In other words, people could be screened for high-risk genes and then given either lifestyle advice or drugs to prevent disease.

This latter strategy has been marked by a general failure to identify common gene variants that place large numbers of people at high risk for common non-communicable diseases. Instead, a large number of variants with small effect sizes have been identified. In general, the inclusion of genetic information in risk-prediction algorithms does little to improve risk prediction beyond simple questionnaires and blood tests for conditions such as diabetes and cardiovascular disease (Thanassoulis & Vasan, 2010; Echouffo-Tcheugui *et al.* 2013). The current worldwide rise in obesity seems so driven by the combination of high calories and low physical activity that some have concluded that the search for obesity-risk genes is futile (Veerman, 2011). Finally, even if such predictive information were available, would the average person change their behaviour or would low-risk individuals feel free generally to ignore well-known health guidelines? These issues are dealt with in more detail in the article by Joyner & Prendergast (2014) in this issue.

There is also a parallel story for rare phenotypes. In the case of extreme longevity (>100 years) the search for a clear-cut genotype–phenotype narrative (Sebastiani & Perls, 2012) has been slow to emerge and hard to unravel. For sudden death in young athletes, most commonly caused by hypertrophic cardiomyopathy, multiple causative rare genetic defects have emerged (Landstrom & Ackerman, 2010). However, even within the same family siblings with the potentially lethal gene variant do not always manifest the tragic phenotype.

At some level, biomedical research driven by the Modern Synthesis is being repackaged again. The idea is that certain gene variants might offer new therapeutic targets for common diseases. A notable recent example is the targeting of pathways associated with the *PCSK9* gene (Steinberg & Witztum, 2009) to reduce cholesterol. The extent to which this new strategy is more effective than the earlier focuses on genetic engineering or the common variant common phenotype remains to be seen.

Based on the above overview, it might be argued that the biomedical efforts informed by the Modern Synthesis have stalled or at least underperformed. In contrast, progress in epidemiology and public policy marches on, with ever more evidence showing the powerful effects of behaviour, environment and social circumstances on health (McGinnis *et al.* 2002; Wilkinson & Marmot, 2003; Bortz, 2005; Kuznetsova, 2012).

The extent to which the genome project has not influenced medical practice is striking (Editorial, 2010). For example, several recent clinical trials have shown little or no benefit of genetic testing to improve the dosing of the commonly used anticoagulant warfarin. Additionally, the need to design clinical trials to evaluate personalized therapy objectively, based on individual genetic markers, is critically needed.

The ubiquity and abundance of between-generation epigenetic inheritance has implications for assessing disease risk and the responses to ecological stresses. New methods for identifying and estimating the extent of heritable, epigenetic variation in populations are necessary. One method for doing this has been developed by Tal *et al.* (2010), who have combined a classical quantitative genetics approach with information about the number of opportunities for epigenetic reset between generations and assumptions about environmental induction to estimate the heritable epigenetic variance and epigenetic transmissibility. The application of this or similar methods to epidemiological data can help to uncover the epigenetic correlates and causes of complex metabolic and environmental diseases and help in finding adequate treatments. Further relevant material can be found in the article on the Predictive Adaptive Response (PAR) in this issue (Bateson *et al.* 2014).

Relevance for an extended evolutionary synthesis

It is clear, therefore, that evolutionary theory is undergoing ferment. Advances in the empirical and conceptual approaches to evolution prompt a renewed appreciation of the multiplicity of processes interacting in evolutionary change, leading to an expanded theoretical framework beyond the standard population genetic account (Margulis & Sagan, 2003; Beurton *et al.* 2008; Pigliucci & Müller, 2010; Gissis & Jablonka, 2011; Shapiro, 2011). Physiological science has an important role in this encompassing reform of evolutionary theory, because of three major contributions it can make, namely the reintroduction of function, the addition of higher order organizing principles and an account of organismal systems properties.

In the classical view of the Modern Synthesis, function – in general – was all but excluded from having any role in the generation of selectable variation, the directionality of evolutionary change (which was assumed to be the consequence of selection alone) or the kind of information transmitted from one generation to the next. The contributions to this issue demonstrate that this view is unwarranted on all three accounts. Hence, a representation of functional principles is required in the evolutionary framework. Indeed, while functional and evolutionary explanation were once regarded as distinct (Mayr, 1961), since the 1980s function has been re-appreciated, mostly in terms of constraints acting on the generation of phenotypic variation (Wagner, 1984; Maynard-Smith *et al.* 1985). More recently, functional principles have come to be addressed via evolutionary studies of gene regulation, embryonic development, comparative behaviour, ecological systems and, in particular, physiology. The trigger for this was the desire to achieve a better mechanistic understanding of the genotype–phenotype relation in the evolutionary process. It is hardly surprising that the emphasis has been, and still is, on the molecular analysis of gene

action, through functional genomics, transgenic techniques and genetic engineering. Essentially, this provides a means of experimental testing of the predictions made by statistical genetic inference (Dean & Thornton, 2007), thus adding a new level of analysis to evolutionary science.

While these aspects of function improve our mechanistic understanding of the genotype–phenotype relation, physiology brings function to evolution also in a different way, through the higher order control that physiological systems exert over basic molecular processes. Hormonal activity, metabolic networks or electrolyte regulation, to name but a few, represent physiological systems that are not restricted to specific gene activity, but affect the behaviour of numerous cells, tissues and developmental processes at once. Such functional systems may themselves be a target of selection, but, more importantly, they can also affect the pace and directionality of evolutionary change. In these cases, the phenotypic outcome is not an immediate consequence of natural selection, but a consequence of the functional properties of the given system. For instance, physiological activity during development, such as embryonic movement, when altered through evolution, leads to specific morphological consequences, e.g. the loss or gain of skeletal elements (Müller, 2003). Moreover, the functional properties of proteins already present in unicellular organisms, when mobilized in a multicellular context, may dictate the possible arrangements of primary metazoan body plans (Newman *et al.* 2006).

Functional systems affect evolutionary processes also through their influence on inheritance, e.g. via epigenetic marking or gene silencing. Epigenetic models show that the rate and direction of evolutionary change can differ markedly from that inferred from population genetic models (Day & Bonduriansky, 2011; Geoghegan & Spencer, 2012), and epigenetic inheritance may accelerate genetic accommodation processes (e.g. Klironomos *et al.* 2013). Heritable epigenetic changes may also accompany ecological and genomic shocks and contribute to macroevolutionary change, for example in speciation events (Jablonka & Lamb, 1995, 2014). Furthermore, epigenetic DNA methylation, which leads to tissue-specific gene silencing, can greatly accelerate the rate of fixation of beneficial recessive mutations (Chess, 2012) and adaptive evolution by gene duplication (Rodin *et al.* 2005). These effects strongly modify the standard picture of evolutionary theory and induce further questions about the role and the evolutionary sophistication of epigenetic mechanisms during the major transitions in evolution (Jablonka & Lamb, 2006).

Another way in which functional systems shape evolution is through their multilevel interactions. Biological functions interconnect at many different levels of organization, from molecules to whole organisms, some aspects of which can now be quantified through systems biological approaches, such as the physiome project (Hunter *et al.* 2002; Hunter & Borg, 2003). Hunter & de Bono (2014) in this issue combine 'a multiscale hierarchy of functional tissue units (FTUs) with the corresponding application of physical laws to describe molecular interaction networks and flow processes over continuum fields within these units' to explore the 'biophysical constraints on tissue evolution'. Newman (2014) also discusses how the application of physical laws in biology can show that 'large-scale changes in organismal form now [provide] a scientific basis other than gradualistic natural selection based on adaptive advantage'.

In developmental processes that generate biological form, for instance, cellular architecture, tissue activity, physiological regulation and gene activation play together in intricate functional networks, without any privileged level of control. Evolutionary modification of such multilevel dynamics, be it through mutation, natural selection or environmental induction, will always affect the entire system. By necessity, such multilevel systems exhibit emergent properties (Badyaev, 2011) and produce threshold effects that influence the phenotypic outcome (Lange *et al.* 2013; Čapek *et al.* 2014). On the evolutionary scale, such properties can lead to non-linear dynamics in population change (Jaeger *et al.* 2012). By connecting levels of organization and by defining the effective parameters and boundary conditions for functional interactions among them, the physiological sciences can make a major contribution towards the explanation of non-gradual evolutionary dynamics and macro-evolutionary events.

Thus, function in general, and physiological function in particular, does affect the generation of selectable variation, the directionality of evolutionary change and the transmission of genetic and non-genetic information. Hence, evolutionary biologists should genuinely be interested in the functional physiological approach. First steps are being made, and a functional synthesis between molecular biology and evolutionary biology has been proposed (Dean & Thornton, 2007). What we advocate here is different; not only does molecular function need to be reconciled with statistical gene variation, but the rules of higher order functional principles need to become part of a major reform of the general evolutionary framework that is currently taking place through the inclusion of new concepts from evo–devo, niche construction [see the article by Laland *et al.* (2014) in this issue], epigenetic inheritance and other areas (Pigliucci & Müller, 2010). Consideration of function permits the integration of this extended synthesis view of evolution with physiology. The hallmark of such a reform is a relinquishment of any privileged levels of causation in the evolutionary process and a replacement of gene reductionism by systems principles (Noble, 2012, 2013). Aware of the fact that many of the relevant processes now have become accessible to empirical research, Morange (2011) noted correctly: 'the obstacles for a merging of functional and evolutionary biology have potentially disappeared'.

Consequences for concepts and definitions

Finally, we note some consequences for the definitions of key elements and concepts, focusing on the concept of the gene. The articles by Keller (2014), Roll-Hansen (2014) and Roux (2014) in this issue should be consulted for important accounts on the history and philosophy of the relevant concepts and for their interpretations of the consequences.

The concept of 'gene' is primary amongst these, because the Modern Synthesis is a gene-centred theory of evolution. There has always been a tension between its original definition as a discrete, inheritable phenotype following Mendelian laws and the modern molecular biological definition of a gene as a template for a specific protein (Keller, 2000; Noble, 2008). The tension was manageable for so long as it was thought that the relations between genotype and phenotype were at least fairly direct, even if people long ago gave up 'the silent assumption [that] was made

almost universally that there is a 1:1 relation between genetic factor (gene) and character' (Mayr, 1982) to acknowledge that many genes are involved in each physiological function. From a physiological viewpoint, even this concession is not enough. Organisms are remarkably well buffered against DNA changes through built in back-up mechanisms. In the heart's pacemaker, multiple back-up mechanisms exist, so that targeting any one protein may result in only small changes in rhythm (Noble *et al.* 1992; Noble & Noble, 2011). In yeast, 80% of single knock-outs are silent in normal physiological conditions (Hillenmeyer *et al.* 2008). The relation between DNA and the phenotype is better represented as being mediated by functional networks, in which not all the components are specified in DNA sequences (Kohl *et al.* 2010). To this problem we need to add that posed by genetic assimilation, which, as we argued earlier, cannot be represented properly in terms of individual genes, but rather as networks of alleles; to which we can add the difficulty, also referred to already, that DNA sequences provide a relatively poor prediction of disease risks.

There has therefore been a new tendency within the Modern Synthesis view to represent this as a problem of 'missing inheritance', 'honorary genes' or 'phantom inheritability' (Zuk *et al.* 2012). This misleading terminology hides the problem in terms that have no role in scientific discourse. The better way forward is to recognize, quite simply, that we need a much better notion of inheritance through a systemic understanding of the genotype–phenotype relation. From such understanding we will, for example, be able to explain how the statistical concepts of broad and narrow senses of heritability are functions of regulatory anatomy and the environment (Wang *et al.* 2013).

It is also important to distinguish between different meanings of 'function' in physiology and in evolutionary biology. They are significantly different but often confused. As Roux (2014) says, '[since selectionist theories] restrict the functional attribution of a trait to its past selective value and not its current properties, these theories are inconsistent with the concept of function in physiology'. Many other terms in the discourse also need rethinking in the light of these considerations, such as 'genetic code', 'genetic programme' and 'book of life'.

Conclusions

The wide-ranging set of articles published in this issue reveal a major challenge both for the physiological sciences and for evolutionary biology. As the integration between the two proceeds, neither can remain unchanged. Evolutionary theory requires extension or even replacement, while physiological science needs to address the exciting possibilities opened up for the future. We hope that our article, and those published here, will enable both disciplines to respond effectively to that challenge.

References

Acuña R, Padilla B, Flórez-Ramosa CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH, Egan AN, Doyle JJ & Rose JKC (2012). Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A* **109**, 4197–4202.

Anway MD, Memon MA, Uzumcu M & Skinner MK (2006). Transgenerational effect of the endocrine disruptor vinclozolin on male spermatogenesis. *J Androl* **27**, 868–879.

Badyaev AV (2011). Origin of the fittest: link between emergent variation and evolutionary change as a critical question in evolutionary biology. *Proc Biol Sci* **278**, 1921–1929.

Badyaev AV (2014). Epigenetic resolution of the 'curse of complexity'. *J Physiol* **592**, 2251–2260.

Ball P (2013). DNA: celebrate the unknowns. *Nature* **496**, 419–420.

Bateson P (2014). New thinking about biological evolution. *Biol J Linn. Soc* DOI: 10.1111/bij.12125.

Bateson W (1906). In Report of the Third International Conference 1906 on Genetics 91 (Royal Horticultural Society, London, 1907).

Bateson P & Gluckman P (2011). *Plasticity, Robustness, Development & Evolution.* Cambridge University Press, Cambridge, UK.

Bateson P, Gluckman PD & Hanson M (2014). The biology of developmental plasticity and the hypothesis of the Predictive Adaptive Response. *J Physiol* **592**, 2357–2368.

Bateson P & Laland K (2013). Tinbergen's four questions: an appreciation and an update. *Trends Ecol Evol* **28**, 712–718.

Baverstock K & Rönkkö M (2014). The evolutionary origin of form and function. *J Physiol* **592**, 2261–2265.

Beurton PJ, Falk R & Rheinberger H-J (2008). *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives.* Cambridge University Press, Cambridge, UK.

Bortz WM (2005). Biological basis of determinants of health. *Am J Public Health* **95**, 389–392.

Čapek D, Metscher BD & Müller GB (2014). Thumbs down: a molecular-morphogenetic approach to avian digit homology. *J Exp Zool B Mol Dev Evol* **322**, 1–12.

Chess A (2012). Mechanisms and consequences of widespread random monoallelic expression. *Nat Rev Genet* **13**, 421–428.

Cummins R (1975). Functional analysis. *J Phil* **72**, 741–765.

Danchin E, Charmantier A, Champagne FA, Mesoudi A, Pujol B & Blanchet S (2011). Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet* **12**, 475–486.

Danchin E & Pocheville A (2014). Inheritance is where physiology meets evolution. *J Physiol* **592**, 2307–2317.

Day T & Bonduriansky R (2011). A unified approach to the evolutionary consequences of genetic and nongenetic inheritance *Am Nat* **178**, E18–E36.

Dean AM & Thornton JW (2007). Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet* **8**, 675–688.

Dias BG & Ressler KJ (2014). Parental olfactory experience influences behaviour and neural structure in subsequent generations. *Nat Neurosci* **17**, 89–96.

Echouffo-Tcheugui JB, Dieffenbach SD & Kengne AP (2013). Added value of novel circulating and genetic biomarkers in type 2 diabetes prediction: a systematic review. *Diabetes Res Clin Pract* **101**, 255–269.

Editorial (2010). The human genome at ten. *Nature* **464**, 649–650.

Gatenby RA & Gillies RJ (2008). A microenvironmental model of carcinogenesis. *Nat Rev Cancer* **8**, 56–61.

Geoghegan J & Spencer H (2012). Population-epigenetic models of selection. *Theor Popul Biol* **81**, 232–242.

Gissis SB & Jablonka E, ed. (2011). *Transformations of Lamarckism. From Subtle Fluids to Molecular Biology.* MIT Press, Cambridge, MA, USA.

Gjuvsland AB, Vik JO, Beard DA, Hunter PJ & Omholt SW (2013). Bridging the genotype-phenotype gap: what does it take? *J Physiol* **591**, 2055–2066.

Gluckman PD, Hanson MA & Beedle AS (2007). Non-genomic transgenerational inheritance of disease risk. *Bioessays* **29**, 145–154.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C & Giaever G (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362–365.

Hunter PJ & Borg TK (2003). Integration from proteins to organs: the Physiome Project. *Nat Rev Mol Cell Biol* **4**, 237–243.

Hunter PJ & De Bono B (2014). Biophysical constraints on the evolution of tissue structure and function. *J Physiol* **592**, 2389–2401.

Hunter PJ, Robbins P & Noble D (2002). The IUPS human Physiome Project. *Pflügers Arch* **445**, 1–9.

Jablonka E & Lamb M (1995). *Epigenetic Inheritance and Evolution. The Lamarckian Dimension*. OUP, Oxford, UK.

Jablonka E & Lamb M (2014). *Evolution in Four Dimensions*. MIT Press, Boston, MA, USA (revised edition).

Jablonka E & Lamb M (2006). The evolution of information in the major transitions. *J Theor Biol* **239**, 236–246.

Jaeger J, Irons D & Monk N (2012). The inheritance of process: a dynamical systems approach. *J Exp Zool B Mol Dev Evol* **318**, 591–612.

Jaeger J & Monk N (2014). Bioattractors: dynamical systems theory and the evolution of regulatory processes. *J Physiol* **592**, 2267–2281.

Joyner MJ & Prendergast FG (2014). Chasing Mendel: five questions for personalized medicine. *J Physiol* **592**, 2381–2388.

Kaati G, Bygren LO, Pembrey ME & Sjöström M (2007). Transgenerational response to nutrition, early life circumstances and longevity *Eur J Hum Genet* **15**, 784–790.

Kamberov Y, Wang S, Tan J, Gerbault P, Wark A, Tan L, Yang Y, Li S, Tang K, Chen H, Powell A, Itan Y, Fuller D, Lohmueller J, Mao J, Schachar A, Paymer M, Hostetter E, Byrne E, Burnett M, McMahon A, Thomas M, Lieberman D, Jin L, Tabin C, Morgan B & Sabeti P (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**, 691–702.

Keller EF (2000). *The Century of the Gene*. Harvard University Press, Cambridge, MA, USA.

Keller EF (2014). From Gene action to reactive genomes. *J Physiol* **592**, 2423–2429.

Kitcher P (1993). *The Advancement of Science*. Oxford University Press, New York, NY, USA.

Klironomos FD, Berg J & Collins S (2013). How epigenetic mutations can affect genetic evolution: model and mechanism. *Bioessays* **35**, 571–578.

Kohl P, Crampin E, Quinn TA & Noble D (2010). Systems biology: an approach. *Clin Pharmacol Ther* **88**, 25–33.

Kuznetsova D (2012). *Healthy Places. Towards a Health Improving Council*. New Local Government Network, UK.

Laland KN, Odling-Smee FJ & Turner S (2014). The role of internal and external constructive processes in evolution *J Physiol* **592**, 2413–2422.

Lamm E (2014). The genome as a developmental organ. *J Physiol* **592**, 2283–2293.

Landstrom AP & Ackerman MJ (2010). Mutation Type Is Not Clinically Useful in Predicting Prognosis in Hypertrophic Cardiomyopathy. *Circulation* **122**, 2441–2450.

Lange A, Nemeschkal HL & Müller GB (2013). Biased polyphenism in polydactylous cats carrying a single point mutation: the Hemingway model for digit novelty. *Evol Biol* DOI 10.1007/s11692-013-9267-y.

Levin M (2014). Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. *J Physiol* **592**, 2295–2305.

McClintock B (1950). The origin and behaviour of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**, 344–355.

McClintock B (1984). The significance of responses of the genome to challenge *Science* **226**, 792–801.

McGinnis JM, Williams-Russo P & Knickman JR (2002). The case for more active policy attention to health promotion. *Health Affairs* **21**, 78–93.

Margulis L & Sagan D (2003). *Acquiring Genomes: A Theory of the Origins of Species* Basic Books, New York.

Mattick JS (2012). Rocking the foundations of molecular genetics. *Proc Natl Acad Sci U S A* **109**, 16400–16401.

Maynard-Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D & Wolpert L (1985). Developmental constraints and evolution. *Q Rev Biol* **60**, 265–287.

Mayr E (1961). Cause and effect in biology. *Science* **134**, 1501–1506.

Mayr E (1982). *The Growth of Biological Thought*. Harvard, Cambridge, MA, USA.

Melham T, Bard J, Werner E & Noble D (2013). Conceptual foundations of systems biology. *Prog Biophys Mol Biol* **111**, 55–56.

Morange M (2011). What will result from the interaction between functional and evolutionary biology? *Stud Hist Philos Biol Biomed Sci* **42**, 69–74.

Müller GB (2003). Embryonic motility: environmental influences and evolutionary innovation. *Evol Dev* **5**, 56–60.

Müller GB (2007). Evo–devo: extending the evolutionary synthesis. *Nat Rev Genet* **8**, 943–949.

Nelson VR, Heaney JD, Tesar PJ, Davidson NO & Nadeau JH (2012). Transgenerational epigenetic effects of *Apobec1* deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc Natl Acad Sci U S A* **109**, E2766–E2773.

Newman SA (2014). Form and function remixed: developmental physiology in the evolution of vertebrate body plans *J Physiol* **592**, 2403–2412.

Newman SA, Forgacs G & Muller GB (2006). Before programs: the physical origination of multicellular forms. *Int J Dev Biol* **50**, 289–299.

Noble D (2006). *The Music of Life*. OUP, Oxford, UK.

Noble D (2008). Genes and causation. *Philos Trans A Math Phys Eng Sci* **366**, 3001–3015.

Noble D (2011). Neo-Darwinism, the modern synthesis and selfish genes: are they of use in physiology? *J Physiol* **589**, 1007–1015.

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

Noble D (2013). Physiology is rocking the foundations of evolutionary biology *Exp Physiol* **98**, 1235–1243.

Noble D, Denyer JC, Brown HF & DiFrancesco D (1992). Reciprocal role of the inward currents $i_{b,Na}$ and i_f in controlling and stabilizing pacemaker frequency of rabbit sino-atrial node cells. *Proc Biol Sci* **250**, 199–207.

Noble PJ & Noble D (2011). A historical perspective on the development of models of rhythm in the heart. In *Heart Rate and Rhythm: Molecular Basis, Pharmacological Modulation and Clinical Implications*, ed. Tripathi ON, Ravens U & Sanguinetti MC, pp. 155–173. Springer, Heidelberg, Germany.

Olby RC (2000). Horticulture: the font for the baptism of genetics. *Nat Rev Genet* **1**, 65–70.

Omholt SW (2013). From sequence to consequence and back. *Prog Biophys Mol Biol* **111**, 75–82.

Pigliucci M & Müller GB (2010). *Evolution – the Extended Synthesis*. MIT Press, Cambridge, MA, USA.

Portrait F, Teeuwiszen E & Deeg D (2011). Early life undernutrition and chronic diseases at older ages: the effects of the Dutch famine on cardiovascular diseases and diabetes. *Soc Sci Med* **73**, 711–718.

Rechavi O, Minevish G & Hobert O (2011). Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248–1256.

Redrejo-Rodríguez M, Muñoz-Espín D, Holguera I, Mencía M & Salas M (2012). Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages. *Proc Natl Acad Sci U S A* **109**, 18482–18487.

Robinson S, Turrell ES, Belding HS & Horvath SM (1943). Rapid acclimatization to work in hot climates. *Am J Physiol* **140**, 168–176.

Rodin SN, Parkhomchuk DV, Rodin AS, Holmquist GP & Riggs AD (2005). Repositioning-dependent fate of duplicate genes. *DNA Cell Biol* **24**, 529–542.

Roll-Hansen N (2014). The holist tradition in 20th century genetics. Wilhelm Johannsen's genotype concept. *J Physiol* **592**, 2431–2438.

Rosenberg SM & Queitsch C (2014). Combating Evolution to fight disease *Science* **343**, 1088–1089.

Roux E (2014). The concept of function in modern physiology. *J Physiol* **592**, 2245–2249.

Sebastiani P & Perls TT (2012). The genetics of extreme longevity: lessons from the New England Centenarian Study. *Front Genet* **3**, 277.

Sela M, Kloog Y & Rechavi O (2014). Non-coding RNAs as the bridge between epigenetic mechanisms, lineages and domains of life. *J Physiol* **592**, 2369–2373.

Shapiro JA (2011). *Evolution: A View from the 21st Century*. Pearson Education Inc., Upper Saddle River, NJ, USA.

Shapiro JA (2014). Physiology of the read–write (RW) genome. *J Physiol* **592**, 2319–2341.

Steinberg D & Witztum JL (2009). Inhibition of PCSK9: a powerful weapon for achieving ideal LDL cholesterol levels. *Proc Natl Acad Sci U S A* **106**, 9546–9547.

Stern S, Snir O, Mizrachi E, Galili M, Zaltsman I & Soen Y (2014). Reduction in maternal *Polycomb* levels contributes to transgenerational inheritance of responses to toxic stress in flies. *J Physiol* **592**, 2343–2355.

Sun Y-H & Zhu Z-Y (2014). Cross-species cloning: influence of cytoplasmic factors on development. *J Physiol* **592**, 2375–2379.

Tal O, Kisdi E & Jablonka E (2010). Epigenetic contribution to covariance between relatives. *Genetics* **184**, 1037–1050.

Thanassoulis G & Vasan RS (2010). Genetic cardiovascular risk prediction – will we get there? *Circulation* **122**, 2323–2334.

Tinbergen N (1963). On aims and methods of ethology. *Z Tierpsychol* **20**, 410–433.

van Abeelen AF, Elias SG, Bossuyt PM, Grobbee DE, van der Schouw YT, Roseboom TJ & Uiterwaal CS (2012). Cardiovascular consequences of famine in the young. *Eur Heart J* **33**, 538–545.

Veerman JL (2011). On the futility of screening for genes that make you fat. *PLoS Med* **8**, e1001114.

Waddington CH (1957). *The Strategy of the Genes*. Allen and Unwin, London, UK.

Wagner GP (1984). Coevolution of functionally constrained characters: prerequisites for adaptive versatility. *Biosystems* **17**, 51–55.

Wang Y, Vik J, Omholt S & Gjuvsland A (2013). Effect of regulatory architecture on broad *versus* narrow sense heritability. *PLoS Comput Biol* **9**, e1003053.

Watson J (2013). Oxidants, antioxidants and the current incurability of metastatic cancers. *Open Biol* **3**, 120144.

Weaver ICG (2009). Life at the interface between a dynamic environment and a fixed genome. In *Mammalian Brain Development*, ed. Janigro D, pp. 17–40. Humana Press, Springer, New York, NY, USA.

West-Eberhard MJ (2003). *Developmental Plasticity and Evolution*. Oxford University Press, New York, NY, USA.

Wilkinson R & Marmot M (2003). *Social Determinants of Health: the Solid Facts*, 2nd edition. WHO Regional Office for Europe, Copenhagen, Denmark.

Wright L (1973). Functions. *Phil Rev* **82**, 139–168.

Wyndham CH (1967). Effect of acclimatization on the sweat rate–rectal temperature relationship. *J Appl Physiol* **22**, 27–30.

Zuk O, Hechter E, Sunyaev SR & Lander ES (2012). The mystery of missing heritability: genetic interactions create phantom heritability *Proc Natl Acad Sci U S A* **109**, 1193–1198.

© 2015. Published by The Company of Biologists Ltd | The Journal of Experimental Biology (2015) 218, 7-13 doi:10.1242/jeb.106310

REVIEW

Evolution beyond neo-Darwinism: a new conceptual framework

Denis Noble*

ABSTRACT

Experimental results in epigenetics and related fields of biological research show that the Modern Synthesis (neo-Darwinist) theory of evolution requires either extension or replacement. This article examines the conceptual framework of neo-Darwinism, including the concepts of 'gene', 'selfish', 'code', 'program', 'blueprint', 'book of life', 'replicator' and 'vehicle'. This form of representation is a barrier to extending or replacing existing theory as it confuses conceptual and empirical matters. These need to be clearly distinguished. In the case of the central concept of 'gene', the definition has moved all the way from describing a necessary cause (defined in terms of the inheritable phenotype itself) to an empirically testable hypothesis (in terms of causation by DNA sequences). Neo-Darwinism also privileges 'genes' in causation, whereas in multi-way networks of interactions there can be no privileged cause. An alternative conceptual framework is proposed that avoids these problems, and which is more favourable to an integrated systems view of evolution.

KEY WORDS: Epigenetics, Genetic program, Modern synthesis, Lamarck, Systems biology

Origin of this article

This paper represents the culmination of ideas previously developed in a book, *The Music of Life* (Noble, 2006), and four related articles (Noble, 2011b; Noble, 2012; Noble, 2013; Noble et al., 2014). Those publications raised many questions from readers in response to which the 'Answers' pages (http://musicoflife.co.uk/Answers-menu.html) of *The Music of Life* website were drafted. Those pages, in particular the page entitled *The language of Neo-Darwinism*, were written in preparation for the present article. The ideas have been extensively honed in response to further questions and comments.

Introduction

The recent explosion of research on epigenetic mechanisms described in this issue and elsewhere (e.g. Noble et al., 2014), and most particularly work focused on trans-generational inheritance mediated by those mechanisms (e.g. Danchin et al., 2011; Dias and Ressler, 2014; Gluckman et al., 2007; Klironomos et al., 2013; Nelson et al., 2012; Nelson and Nadeau, 2010; Nelson et al., 2010; Rechavi et al., 2011; Sela et al., 2014), has created the need to either extend or replace the Modern (neo-Darwinist) Synthesis (Beurton et al., 2008; Gissis and Jablonka, 2011; Noble et al., 2014; Pigliucci and Müller, 2010). This paper explains why replacement rather than extension is called for. The reason is that the existence of robust mechanisms of trans-generational inheritance independent of DNA sequences runs strongly counter to the spirit of the Modern Synthesis. In fact, several new features of experimental results on inheritance and mechanisms of evolutionary variation are incompatible with the Modern Synthesis. Fig. 1 illustrates the definitions and relationships between the various features of Darwinism, the Modern Synthesis and a proposed new Integrative Synthesis. The diagram is based on an extension of the diagram used by Pigliucci and Müller (Pigliucci and Müller, 2010) in explaining the idea of an extended Modern Synthesis.

The shift to a new synthesis in evolutionary biology can also be seen to be part of a more general shift of viewpoint within biology towards systems approaches. The reductionist approach (which inspired the Modern Synthesis as a gene-centred theory of evolution) has been very productive, but it needs, and has always needed, to be complemented by an integrative approach, including a new theory of causation in biology (Noble, 2008), which I have called the theory of Biological Relativity (Noble, 2012). The approach to replace the Modern Synthesis could be called the Integrative Synthesis as it would be based on the integration of a variety of mechanisms of evolutionary change that must interact, rather than the single mechanism postulated by the Modern Synthesis (Noble, 2013). We are moving to a much more nuanced multi-mechanism theory of evolution, which, interestingly, is closer to some of Darwin's ideas than to neo-Darwinism. Darwin was not a neo-Darwinist. He recognised other mechanisms in addition to natural selection and these included the inheritance of acquired characteristics.

The language of neo-Darwinism

Many of the problems with the Modern Synthesis in accommodating the new experimental findings have their origin in neo-Darwinist forms of representation rather than in experimental biology itself. These forms of representation have been responsible for, and express, the way in which 20th century biology has most frequently been interpreted. In addition, therefore, to the need to accommodate unanticipated experimental findings, we have to review the way in which we interpret and communicate experimental biology. The language of neo-Darwinism and 20th century biology reflects highly reductionist philosophical and scientific viewpoints, the concepts of which are not required by the scientific discoveries themselves. In fact, it can be shown that, in the case of some of the central concepts of 'selfish genes' or 'genetic program', no biological experiment could possibly distinguish even between completely opposite conceptual interpretations of the same experimental findings (Noble, 2006; Noble, 2011b). The concepts therefore form a biased interpretive veneer that can hide those discoveries in a web of interpretation.

I refer to a web of interpretation as it is the whole conceptual scheme of neo-Darwinism that creates the difficulty. Each concept and metaphor reinforces the overall mind-set until it is almost impossible to stand outside it and to appreciate how beguiling it is. As the Modern Synthesis has dominated biological science for over half a century, its viewpoint is now so embedded in the scientific literature, including standard school and university textbooks, that many biological scientists may not recognise its conceptual nature,

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

*Author for correspondence (Denis.noble@dpag.ox.ac.uk)

The Journal of Experimental Biology (2015) doi:10.1242/jeb.106310

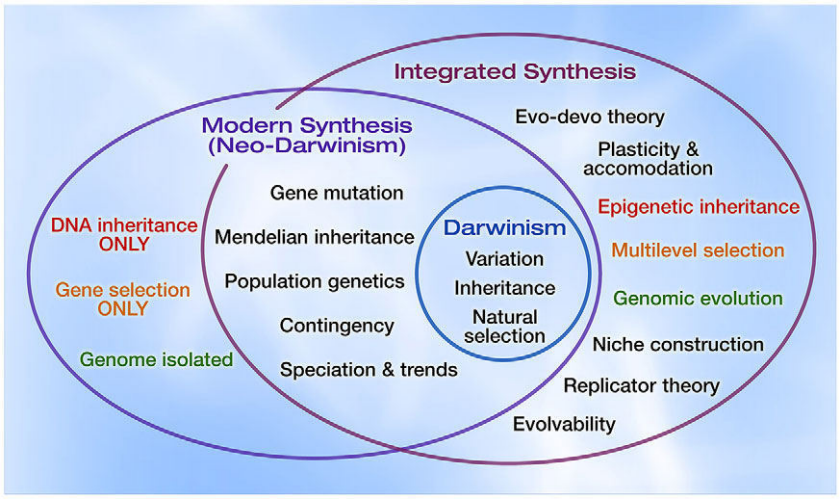


Fig. 1. Diagram illustrating definitions of Darwinism, Modern Synthesis (neo-Darwinism) and Integrated Synthesis. The diagram is derived from Pigliucci and Müller's (Pigliucci and Müller, 2010) presentation of an Extended Synthesis. All the elements are also present in their diagram. The differences are: (1) the elements that are incompatible with the Modern Synthesis are shown coloured on the right; (2) the reasons for the incompatibility are shown in the three corresponding coloured elements on the left. These three assumptions of the Modern Synthesis lie beyond the range of what needs to extend or replace the Modern Synthesis; (3) in consequence, the Modern Synthesis is shown as an oval extending outside the range of the extended synthesis, which therefore becomes a replacement rather than an extension.

let alone question incoherences or identify flaws. Many scientists see it as merely a description of what experimental work has shown: the idea in a nutshell is that genes code for proteins that form organisms via a genetic program inherited from preceding generations and which defines and determines the organism and its future offspring. What is wrong with that? This article analyses what I think is wrong or misleading and, above all, it shows that the conceptual scheme is neither required by, nor any longer productive for, the experimental science itself.

I will analyse the main concepts and the associated metaphors individually, and then show how they link together to form the complete narrative. We can then ask what would be an alternative approach better fitted to what we now know experimentally and to a new more integrated systems view. The terms that require analysis are 'gene', 'selfish', 'code', 'program', 'blueprint' and 'book of life'. We also need to examine secondary concepts like 'replicator' and 'vehicle'.

'Gene'

Neo-Darwinism is a gene-centred theory of evolution. Yet, its central notion, the 'gene', is an unstable concept. Surprising as it may seem, there is no single agreed definition of 'gene'. Even more seriously, the different definitions have incompatible consequences for the theory.

The word 'gene' was introduced by Johannsen (Johannsen, 1909). But the concept had already existed since Mendel's experiments on plant hybrids, published in 1866 (see Druery and Bateson, 1901), and was based on 'the silent assumption [that] was made almost universally that there is a 1:1 relation between genetic factor (gene) and character' (Mayr, 1982). Of course, no-one now thinks that there is a simple 1:1 relation, but the language of direct causation has been retained. I will call this definition of a 'gene' $gene_J$ to signify Johannsen's (but essentially also Mendel's) meaning. Since then, the concept of a gene has changed fundamentally. $Gene_J$ referred to the cause of a specific inheritable phenotype characteristic (trait), such as eye/hair/skin colour, body shape and mass, number of legs/arms/wings, to which we could perhaps add more complex traits such as intelligence, personality and sexuality.

The molecular biological definition of a gene is very different. Following the discovery that DNA forms templates for proteins, the definition shifted to locatable DNA sequences with identifiable beginnings and endings. Complexity was added through the discovery of regulatory elements (essentially switches), but the basic cause of phenotype characteristics was still thought to be the DNA sequence as that forms the template to determine which protein is made, which in turn interacts with the rest of the organism to produce the phenotype. I will call this definition of a 'gene' $gene_M$ (see Fig. 2).

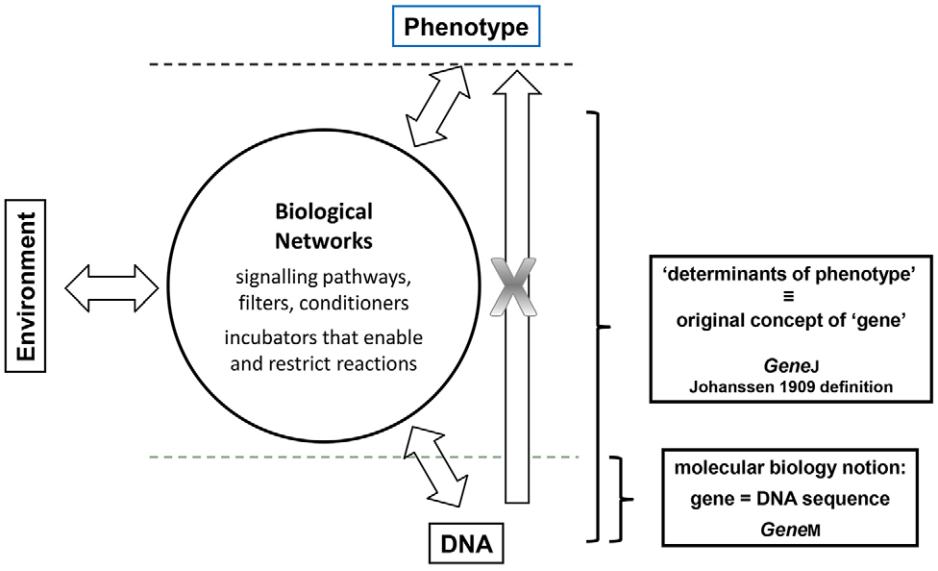


Fig. 2. Relationships between genes, environment and phenotype characters according to current physiological and biochemical understanding. This diagram represents the interaction between DNA sequences, environment and phenotype as occurring through biological networks. The causation occurs in both directions between all three influences on the networks. This view is very different from the idea that genes 'cause' the phenotype (right-hand arrow). This diagram also helps to explain the difference between the original concept of a gene as the cause of a particular phenotype ($gene_J$) and the modern definition as a DNA sequence ($gene_M$). For further description and analysis see Kohl et al. (Kohl et al., 2010).

The Journal of Experimental Biology (2015) doi:10.1242/jeb.106310

But unless all phenotype characteristics are attributable entirely to DNA sequences (which is false: DNA does not act outside the context of a complete cell), $gene_M$ cannot be the same as $gene_J$. According to the original view, $genes_J$ were necessarily the cause of inheritable phenotypes because that is how they were defined: as whatever in the organism is the cause of that phenotype. Johanssen even left the answer on what a gene might be vague: 'The gene was something very uncertain, "ein Etwas" ['anything'], with no connection to the chromosomes' (Wanscher, 1975). Dawkins (Dawkins, 1982) also uses this 'catch-all' definition as 'an inheritable unit'. It would not matter whether that was DNA or something else or any combination of factors. No experiment could disprove a 'catch-all' concept as anything new discovered to be included would also be welcomed as a $gene_J$. The idea becomes unfalsifiable.

The question of causation is now an empirical investigation precisely because the modern definition, $genes_M$, identifies them instead with DNA sequences alone, which omits reference to all other factors. To appreciate the difference, consider Mendel's experiments showing specific phenotypes, such as smooth or wrinkled surfaces of peas. $Gene_J$ was whatever in the plant caused the peas to be smooth or wrinkled. It would not make sense to ask whether $gene_J$ was the cause. That is how it was defined. It simply is everything that determines the inherited phenotype, i.e. the trait. (Of course, different questions of an empirical nature could be asked about $genes_J$, such as whether they follow Mendel's laws. Some do; some don't.) By contrast, it makes perfect sense to ask whether a specific DNA sequence, $gene_M$, is responsible for determining the phenotype. That question is open to experimental investigation. $Gene_J$ could only be the same as $gene_M$ if DNA alone determined the phenotype.

This difference between $gene_J$ (which refers to indeterminate entities that are necessarily the cause) and $gene_M$ (whose causation is open to experimentation) is central and I will use it several times in this article. The difference is in fact large as most changes in DNA do not necessarily cause a change in phenotype. Organisms are very good at buffering themselves against genomic change. Eighty per cent of knockouts in yeast, for example, are normally silent (Hillenmeyer et al., 2008), while critical biological oscillators like the cardiac pacemaker (Noble, 2011a) or circadian rhythm (Foster and Kreitzman, 2004) are buffered against genomic change through extensive back-up mechanisms.

The original concept of a gene has therefore been adopted, but then significantly changed by molecular biology. This led to a great clarification of molecular mechanisms, surely one of the greatest triumphs of 20th century biology, and widely acknowledged as such. But the more philosophical consequences of this change for higher level biology are profound and they are much less widely understood. Fig. 2 summarizes the difference.

Some biological scientists have even given up using the word 'gene', except in inverted commas. As Beurton et al. (Beurton et al., 2008) comment: 'It seems that a cell's enzymes are capable of actively manipulating DNA to do this or that. A genome consists largely of semi stable genetic elements that may be rearranged or even moved around in the genome thus modifying the information content of DNA.' This view is greatly reinforced by the fact that gene expression is stochastic (Chang et al., 2008) and that this itself opens the way to an extensive two-way interaction between the organism's functional networks and the structure and function of chromatin [e.g. figure 10.5 in Kupiec (Kupiec, 2014)].

The reason that the original and the molecular biological definitions have incompatible consequences for neo-Darwinism is that only the molecular biological definition, $gene_M$, could be compatible with a strict separation between the 'replicator' and the 'vehicle'. As illustrated in Fig. 2, a definition in terms of inheritable phenotypic characteristics (i.e. $gene_J$) necessarily includes much more than the DNA, so that the distinction between replicator and vehicle is no longer valid (Noble, 2011b). Note also that the change in definition of a gene that I am referring to here is more fundamental than some other changes that are required by recent findings in genomics, such as the 80% of 'non-coding' DNA that is now known to be transcribed (The_Encode_Project_Consortium, 2012) and which also might be included in the molecular biological definition. Those findings raise an empirical question: are those transcriptions as RNAs functional? That would extend $gene_M$ to include these additional functional sequences. The difference I refer to, by contrast, is a conceptual one. The difference between $gene_J$ and $gene_M$ would still be fundamental because it is the difference between necessary and empirically testable causality, not just an extension of the definition of $gene_M$.

'Selfish'

There is no biological experiment that could distinguish between the selfish gene theory and its opposites, such as 'imprisoned' or 'co-operative genes'. This point was conceded long ago by Richard Dawkins in his book *The Extended Phenotype*: 'I doubt that there is any experiment that could prove my claim' (Dawkins, 1982). A more complete dissection of the language and possible empirical interpretations of selfish gene theory can be found in Noble (Noble, 2011b).

'Code'

After the discovery of the double helical structure of DNA, it was found that each sequence of three bases in DNA or RNA corresponds to a single amino acid in a protein sequence. These triplet patterns are formed from any combination of the four bases U, C, A and G in RNA and T, C, A and G in DNA. They are often described as the genetic 'code', but it is important to understand that this usage of the word 'code' carries overtones that can be confusing. This section of the article is not intended to propose that the word 'code' should not be used. Its purpose is rather to ensure that we avoid those overtones.

A code was originally an intentional encryption used by humans to communicate. The genetic 'code' is not intentional in that sense. The word 'code' has unfortunately reinforced the idea that genes are active and even complete causes, in much the same way as a computer is caused to follow the instructions of a computer program. The more neutral word 'template' would be better. Templates are used only when required (activated); they are not themselves active causes. The active causes lie within the cells themselves because they determine the expression patterns for the different cell types and states. These patterns are communicated to the DNA by transcription factors, by methylation patterns and by binding to the tails of histones, all of which influence the pattern and speed of transcription of different parts of the genome. If the word 'instruction' is useful at all, it is rather that the cell instructs the genome. As the Nobel-prize winner Barbara McClintock said, the genome is an 'organ of the cell', not the other way round (McClintock, 1984).

Representing the direction of causality in biology the wrong way round is confusing and has far-reaching consequences. The causality is circular, acting both ways: passive causality by DNA sequences acting as otherwise inert templates, and active causality by the functional networks of interactions that determine how the genome is activated.

The Journal of Experimental Biology (2015) doi:10.1242/jeb.106310

'Program'

The idea of a 'genetic program' was introduced by the French Nobel laureates Jacques Monod and Francois Jacob. They referred specifically to the way in which early electronic computers were programmed by paper or magnetic tapes: 'The programme is a model borrowed from electronic computers. It equates the genetic material with the magnetic tape of a computer' (Jacob, 1982). The analogy was that DNA 'programs' the cell, tissues and organs of the body just as the code in a computer program causally determines what the computer does. In principle, the code is independent of the machine that implements it, in the sense that the code itself is sufficient to specify what will happen when the instructions are satisfied. If the program specifies a mathematical computation, for example, it would contain a specification of the computation to be performed in the form of complete algorithms. The problem is that no complete algorithms can be found in the DNA sequences. What we find is better characterised as a mixture of templates and switches. The 'templates' are the triplet sequences that specify the amino acid sequences or the RNA sequences. The 'switches' are the locations on the DNA or histones where transcription factors, methylation and other controlling processes trigger their effects. As a program, this is incomplete.

Where then does the full algorithmic logic of a program lie? Where, for example, do we find the equivalent of 'IF-THEN-ELSE' type instructions? The answer is in the cell or organism as a whole, not just in the genome.

Take as an example circadian rhythm. The simplest version of this process depends on a DNA sequence *Period* used as a template for the production of a protein PER whose concentration then builds up in the cytoplasm. It diffuses through the nuclear membrane and, as the nuclear level increases, it inhibits the transcription of *Period* (Foster and Kreitzman, 2004). This is a negative feedback loop of the kind that can be represented as implementing a 'program' like IF LEVEL X EXCEEDS Y STOP PRODUCING X, BUT IF LEVEL X IS SMALLER THAN Y CONTINUE PRODUCING X. But it is important to note that the implementation of this 'program' to produce a 24 h rhythm depends on rates of protein production by ribosomes, the rate of change of concentrations within the cytoplasm, the rate of transport across the nuclear membrane, and interaction with the gene transcription control site (the switch). All of this is necessary to produce a feedback circuit that depends on much more than the genome. It depends also on the intricate cellular, tissue and organ structures that are not specified by DNA sequences, which replicate themselves via self-templating, and which are also essential to inheritance across cell and organism generations.

This is true of all such 'programs'. To call them 'genetic programs' or 'gene networks' is to fuel the misconception that all the active causal determination lies in the one-dimensional DNA sequences. It doesn't. It also lies in the three-dimensional static and dynamic structures of the cells, tissues and organs.

The postulate of a 'genetic program' led to the idea that an organism is fully defined by its genome, whereas in fact the inheritance of cell structure is equally important. Moreover, this structure is specific to different species. Cross-species clones do not generally work. Moreover, when, very rarely, cross-species clones do work, the outcome is determined by the cytoplasmic structures and expression patterns as well as the DNA (Sun et al., 2005). In this connection it is worth noting that the basic features of structural organisation both of cells and of multicellular organisms must have been determined by physical constraints before the relevant genomic information was developed (Müller and Newman, 2003; Newman et al., 2006).

As with 'code', the purpose of this section is to warn against simplistic interpretations of the implications of the word 'program'. In the extended uses to which the word has been put in biology, and in modern computing science where the concept of a distributed program is normal, 'program' can be used in many different ways. The point is that such a 'program' does not lie in the DNA alone. That is also the reason why the concept of a 'genetic program' is not testable. By necessarily including non-DNA elements, there is no way of determining whether a 'genetic program' exists. At the limit, when all the relevant components have been added in, the 'program' is the same as the function it is supposed to be programming. The concept then becomes redundant [p. 53 of Noble (Noble, 2006)]. Enrico Coen (Coen, 1999) put the point beautifully when he wrote: 'Organisms are not simply manufactured according to a set of instructions. There is no easy way to separate instructions from the process of carrying them out, to distinguish plan from execution.'

'Blueprint'

'Blueprint' is a variation on the idea of a program. The word suffers from a similar problem to the concept of a 'program', which is that it can be mistaken to imply that all the information necessary for the construction of an organism lies in the DNA. This is clearly not true. The complete cell is also required, and its complex structures are inherited by self-templating. The 'blueprint', therefore, is the cell as a whole. But that destroys the whole idea of the genome being the full specification. It also blurs and largely nullifies the distinction between replicator and vehicle in selfish gene theory.

'Book of life'

The genome is often described as the 'book of life'. This was one of the colourful metaphors used when projecting the idea of sequencing the complete human genome. It was a brilliant public relations move. Who could not be intrigued by reading the 'book of life' and unravelling its secrets? And who could resist the promise that, within about a decade, that book would reveal how to treat cancer, heart disease, nervous diseases, diabetes, with a new era of pharmaceutical targets. As we all know, it didn't happen. An editorial in *Nature* spelt this out:

> '*The activity of genes is affected by many things not explicitly encoded in the genome, such as how the chromosomal material is packaged up and how it is labelled with chemical markers. Even for diseases like diabetes, which have a clear inherited component, the known genes involved seem to account for only a small proportion of the inheritance…the failure to anticipate such complexity in the genome must be blamed partly on the cosy fallacies of genetic research. After Francis Crick and James Watson cracked the riddle of DNA's molecular structure in 1953, geneticists could not resist assuming it was all over bar the shouting. They began to see DNA as the "book of life," which could be read like an instruction manual. It now seems that the genome might be less like a list of parts and more like the weather system, full of complicated feedbacks and interdependencies.*'
> (Editorial, 2010)

The 'book of life' represents the high watermark of the enthusiasm with which the language of neo-Darwinism was developed. Its failure to deliver the promised advances in healthcare speaks volumes. Of course, there were very good scientific reasons for sequencing whole genomes. The benefits to evolutionary and comparative biology in particular have been immense, and the sequencing of genomes will eventually contribute to healthcare

The Journal of Experimental Biology (2015) doi:10.1242/jeb.106310

when the sequences can be better understood in the context of other essential aspects of physiological function. But the promise of a peep into the 'book of life' leading to a cure for all diseases was a mistake.

The language of neo-Darwinism as a whole

All parts of the neo-Darwinist forms of representation encourage the use and acceptance of the other parts. Once one accepts the idea that the DNA and RNA templates form a 'code', the idea of the 'genetic program' follows naturally. That leads on to statements like 'they [genes] created us body and mind' (Dawkins, 1976; Dawkins, 2006), which gets causality wrong in two ways. First, it represents genes as active causes, whereas they are passive templates. Second, it ignores the many feedbacks on to the genome that contribute to circular causality, in which causation runs in both directions. Those mistakes lead to the distinction between replicators and vehicles. The problem lies in accepting the first step, the idea that there is a 'code' forming a complete program.

The distinction between the replicator and the vehicle can be seen as the culmination of the neo-Darwinist way of thinking. If all the algorithms for the processes of life lie in the genome then the rest of the organism does seem to be a disposable vehicle. Only the genome needs to replicate, leaving any old vehicle to carry it.

The distinction, however, is a linguistic confusion and it is incorrect experimentally (Noble, 2011b). The DNA passed on from one generation to the next is based on copies (though not always perfect). The cell that carries the DNA is also a copy (also not always perfect). In order for a cell to give rise to daughter cells, both the DNA and the cell have to be copied. The only difference between copying a cell and copying DNA is that the cell copies itself by growing (copying its own detailed structure gradually, which is an example of self-templating) and then dividing so that each daughter cell has a full complement of the complex cell machinery and its organelles, whereas copying DNA for the purpose of inheritance occurs only when the cell is dividing. Moreover, the complexity of the structure in each case is comparable: 'It is therefore easy to represent the three-dimensional image structure of a cell as containing as much information as the genome' (Noble, 2011a). Faithful genome replication also depends on the prior ability of the cell to replicate itself because it is the cell that contains the necessary structures and processes to enable errors in DNA replication to be corrected. Self-templating must have been prior to the development of the relevant DNA (Müller and Newman, 2003; Newman et al., 2006).

My germ line cells are therefore just as much 'immortal' (or not) as their DNA. Moreover, nearly all of my cells and DNA die with me. Those that do survive, which are the germ cells and DNA that help to form the next generation, do not do so separately. DNA does not work without a cell. It is simply an incorrect playing with words to single the DNA out as uniquely immortal.

I was also playing with words when I wrote that 'DNA alone is inert, dead' (Noble, 2011b). But at least that has a point in actual experiments. DNA alone does nothing. By contrast, cells can continue to function for some time without DNA. Some cells do that naturally, e.g. red blood cells, which live for about 100 days without DNA. Others, such as isolated nerve axons, fibroblasts (Cox et al., 1976; Goldman et al., 1973) or any other enucleated cell type, can do so in physiological experiments.

$Genes_M$ are best viewed therefore as causes in a passive sense. They do nothing until activated. Active causation lies with proteins, membranes, metabolites, organelles, etc., and the dynamic functional networks they form in interaction with the environment (Noble, 2008).

Notice also that the language as a whole is strongly anthropomorphic. This is strange, given that most neo-Darwinists would surely wish to avoid anthropomorphising scientific discovery.

An alternative form of representation

The alternative form of representation depends on two fundamental concepts. The first one is the distinction between active and passive causes. $Genes_M$ are passive causes; they are templates used when the dynamic cell networks activate them. The second concept is that there is no privileged level of causation. In networks, that is necessarily true, and it is the central feature of what I have called the theory of biological relativity, which is formulated in a mathematical context (Noble, 2012).

I will illustrate the second point in a more familiar non-mathematical way. Take some knitting needles and some wool. Knit a rectangle. If you don't knit, just imagine the rectangle. Or use an old knitted scarf. Now pull on one corner of the rectangle while keeping the opposite corner fixed. What happens? The whole network of knitted knots moves. Now reverse the corners and pull on the other corner. Again, the whole network moves, though in a different way. This is a property of networks. Everything ultimately connects to everything else. Any part of the network can be the prime mover, and be the cause of the rest of the network moving and adjusting to the tension. Actually, it would be better still to drop the idea of any specific element as prime mover. It is networks that are dynamically functional.

Now knit a three-dimensional network. Again, imagine it. You probably don't actually know how to knit such a thing. Pulling on any part of the three-dimensional structure will cause all other parts to move (cf. Ingber, 1998). It doesn't matter whether you pull on the bottom, the top or the sides. All can be regarded as equivalent. There is no privileged location within the network.

The three-dimensional network recalls Waddington's epigenetic landscape network (Fig. 3) and is quite a good analogy to biological networks as the third dimension can be viewed as representing the multi-scale nature of biological networks. Properties at the scale of cells, tissues and organs influence activities of elements, such as genes and proteins, at the lower scales. This is sometimes called downward causation, to distinguish it from the reductionist interpretation of causation as upward causation (Ellis et al., 2012). 'Down' and 'up' here are also metaphors and should be treated carefully. The essential point is the more neutral statement: there is no privileged scale of causality, beyond the representation of scales, perhaps. This must be the case in organisms, which work through many forms of circular causality. A more complete analysis of this alternative approach can be found in the article on Biological Relativity (Noble, 2012), from which Fig. 4 is taken. One of the consequences of the relativistic view is that $genes_M$ cease to be represented as active causes. Templates are passive causes, used when needed. Active causation resides in the networks, which include many components for which there are no DNA templates. It is the physics and chemistry of those dynamic networks that determine what happens.

In certain respects, my article reflects some of the points made over 30 years ago by Ho and Saunders (Ho and Saunders, 1979), who wrote: 'The intrinsic dynamical structure of the epigenetic system itself, in its interaction with the environment, is the source of non-random variations which direct evolutionary change, and that a proper study of evolution consists in the working out of the dynamics of the epigenetic system and its response to environmental stimuli as well as the mechanisms whereby novel developmental responses are canalized.' Their ideas also owe much to those of Conrad Waddington – the term 'canalised' is one that he often used.

The Journal of Experimental Biology (2015) doi:10.1242/jeb.106310

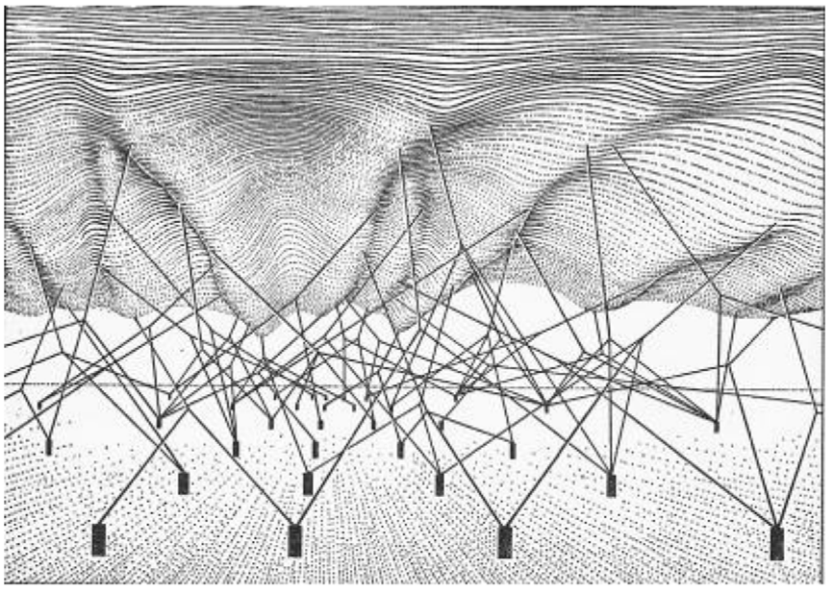


Fig. 3. Conrad Waddington's diagram of the epigenetic landscape. Genes (solid pegs at the bottom) are viewed as parts of complex networks so that many gene products interact between themselves and with the phenotype to produce the phenotypic landscape (top) through which development occurs. Waddington's insight was that new forms could arise through new combinations to produce new landscapes in response to environmental pressure, and that these could then be assimilated into the genome. Waddington was a systems biologist in the full sense of the word. If we had followed his lead many of the more naive 20th century popularisations of genetics and evolutionary biology could have been avoided. Image taken from *The Strategy of the Genes* (Waddington, 1957). Reprinted (2014) by Routledge Library Editions.

An important linguistic feature of the alternative, relativistic, concepts proposed here is that most or all the anthropomorphic features of the neo-Darwinist language can be eliminated, without contravening a single biological experimental fact. There may be other forms of representation that can achieve the same result. It doesn't really matter which you use. The aim is simply to distance ourselves from the biased conceptual scheme that neo-Darwinism has brought to biology, made more problematic by the fact that it has been presented as literal truth.

Conclusions

The extent to which the language of neo-Darwinism has dominated biological thought for over a century since George Romanes invented the term in a letter to *Nature* (Romanes, 1883) is remarkable. It is a tribute to the inventiveness and persuasiveness of many biologists and to their ability to communicate the original idea and its subsequent formulation as the Modern Synthesis to a very wide public. The integration of the early discoveries of molecular biology also contributed great momentum, particularly as the Central Dogma of Molecular Biology (Crick, 1970) was perceived (incorrectly as it subsequently turned out) to confirm a central assumption, which was that the genome was isolated from the lifestyle of the organism and its environment.

In retrospect, neo-Darwinism can be seen to have oversimplified biology and over-reached itself in its rhetoric. By so conclusively excluding anything that might be interpreted as Lamarckism, it assumed what couldn't be proved. As John Maynard Smith (Maynard Smith, 1998) admitted: 'It [Lamarckism] is not so obviously false as is sometimes made out', a statement that is all the more significant from being made by someone working entirely within the Modern Synthesis framework. His qualification on this statement in 1998 was that he couldn't see what the mechanism(s) might be. We can now do so thanks to some ingenious experimental research in recent years.

Nevertheless, the dogmatism was unnecessary and uncalled for. It damaged the reputation of Lamarck, possibly irretrievably.

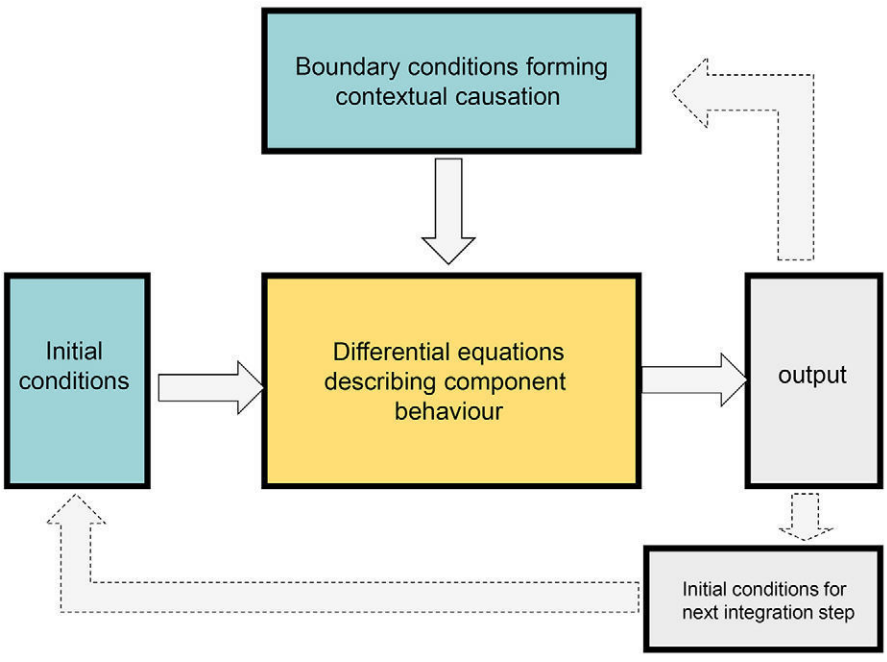


Fig. 4. Many models of biological systems consist of differential equations for the kinetics of each component. These equations cannot give a solution (the output) without setting the initial conditions (the state of the components at the time at which the simulation begins) and the boundary conditions. The boundary conditions define what constraints are imposed on the system by its environment and can therefore be considered as a form of contextual causation from a higher scale. This diagram is highly simplified to represent what we actually solve mathematically. In reality, boundary conditions are also involved in determining initial conditions and the output parameters can also influence the boundary conditions, while they in turn are also the initial conditions for a further period of integration of the equations. The arrows are not really unidirectional. The dotted arrows complete the diagram to show that the output contributes to the boundary conditions (although not uniquely), and determines the initial conditions for the next integration step. Legend and diagram are reproduced from Noble (Noble, 2012).

Lamarck should be recognised by biologists generally as one of the very first to coin and use the term 'biology' to distinguish our science, and by evolutionary biologists in particular for championing the transformation of species against some very powerful critics. Darwin praised Lamarck for this achievement: 'This justly celebrated naturalist…who upholds the doctrine that all species, including man, are descended from other species' (preface to the 4th edition of *The Origin of Species*, 1866).

Many others were damaged too, Waddington included. A little more humility in recognising the pitfalls that beset the unwary when they think they can ignore some basic philosophical principles would have been a wiser strategy. The great physicist Poincaré pointed out, in connection with the relativity principle in physics, that the worst philosophical errors are made by those who claim they are not philosophers (Poincaré, 1902; Poincaré, 1968). They do so because they don't even recognise the existence of the conceptual holes they fall into. Biology has its own version of those conceptual holes.

Acknowledgements

I thank Peter Hacker, Michael Joyner, Peter Kohl, Jean-Jacques Kupiec, Gerd Müller, Raymond Noble and Amit Saad for valuable discussions and comments on the paper itself, and the many correspondents who commented or asked further questions on the *Answers* pages on the Music of Life website (http://musicoflife.co.uk/Answers-menu.html). I thank Bryce Bergene, Senior Designer, Mayo Clinic Creative Media, for the design of Fig. 1. A video version of this figure in relation to the original extended synthesis figure can be viewed online (supplementary material Movie 1).

Competing interests

The author declares no competing financial interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Supplementary material

Supplementary material available online at
http://jeb.biologists.org/lookup/suppl/doi:10.1242/jeb.106310/-/DC1

References

Beurton, P. J., Falk, R. and Rheinberger, H.-J. (2008). *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge: Cambridge University Press.
Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. and Huang, S. (2008). Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-547.
Coen, E. (1999). *The Art of Genes*. Oxford: Oxford University Press.
Cox, R. P., Krauss, M. R., Balis, M. E. and Dancis, J. (1976). Studies on cell communication with enucleated human fibroblasts. *J. Cell Biol.* **71**, 693-703.
Crick, F. (1970). Central dogma of molecular biology. *Nature* **227**, 561-563.
Danchin, É., Charmantier, A., Champagne, F. A., Mesoudi, A., Pujol, B. and Blanchet, S. (2011). Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat. Rev. Genet.* **12**, 475-486.
Dawkins, R. (1982). *The Extended Phenotype*. London: Freeman.
Dawkins, R. (1976, 2006). *The Selfish Gene*. Oxford: Oxford University Press.
Dias, B. G. and Ressler, K. J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat. Neurosci.* **17**, 89-96.
Druery, C. T. and Bateson, W. (1901). Experiments in plant hybridization. *Journal of the Royal Horticultural Society* **26**, 1-32.
Editorial (2010). The human genome at ten. *Nature* **464**, 649-650.
Ellis, G. F. R., Noble, D. and O'Connor, T. (2012). Top-down causation: an integrating theme within and across the sciences. *Interface Focus* **2**, 1-3.
Foster, R. and Kreitzman, L. (2004). *Rhythms of Life*. London: Profile Books.
Gissis, S. B. and Jablonka, E. (2011). *Transformations of Lamarckism. From Subtle Fluids to Molecular Biology*. Cambridge, MA: MIT Press.
Gluckman, P. D., Hanson, M. A. and Beedle, A. S. (2007). Non-genomic transgenerational inheritance of disease risk. *BioEssays* **29**, 145-154.
Goldman, R. D., Pollack, R. and Hopkins, N. H. (1973). Preservation of normal behavior by enucleated cells in culture. *Proc. Natl. Acad. Sci. USA* **70**, 750-754.
Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St Onge, R. P., Tyers, M., Koller, D. et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**, 362-365.
Ho, M. W. and Saunders, P. T. (1979). Beyond neo-Darwinism – an epigenetic approach to evolution. *J. Theor. Biol.* **78**, 573-591.
Ingber, D. E. (1998). The architecture of life. *Sci. Am.* **278**, 48-57.
Jacob, F. (1982). *The Possible and the Actual*. New York, NY: Pantheon Books.
Johannsen, W. (1909). *Elemente der Exakten Erblichkeitslehre*. Jena: Gustav Fischer.
Klironomos, F. D., Berg, J. and Collins, S. (2013). How epigenetic mutations can affect genetic evolution: model and mechanism. *BioEssays* **35**, 571-578.
Kohl, P., Crampin, E. J., Quinn, T. A. and Noble, D. (2010). Systems biology: an approach. *Clin. Pharmacol. Ther.* **88**, 25-33.
Kupiec, J.-J. (2014). Cell differentiation is a stochastic process subjected to natural selection. In *Towards a Theory of Development* (ed. A. Minelli and T. Pradeu), pp. 155-173. Oxford: OUP.
Maynard Smith, J. (1998). *Evolutionary Genetics*. New York, NY: Oxford University Press.
Mayr, E. (1982). *The Growth of Biological Thought*. Cambridge, MA: Harvard.
McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* **226**, 792-801.
Müller, G. and Newman, S. A. (2003). Origination of organismal form: the forgotten cause in evolutionary theory. In *Origination of Organismal Form* (ed. G. Müller and S. A. Newman), pp. 3-10. Boston, MA: MIT Press.
Nelson, V. R. and Nadeau, J. H. (2010). Transgenerational genetic effects. *Epigenomics* **2**, 797-806.
Nelson, V. R., Spiezio, S. H. and Nadeau, J. H. (2010). Transgenerational genetic effects of the paternal Y chromosome on daughters' phenotypes. *Epigenomics* **2**, 513-521.
Nelson, V. R., Heaney, J. D., Tesar, P. J., Davidson, N. O. and Nadeau, J. H. (2012). Transgenerational epigenetic effects of Apobec1 deficiency on testicular germ cell tumor susceptibility and embryonic viability. *Proc. Natl. Acad. Sci. USA* **109**, E2766-E2773.
Newman, S. A., Forgacs, G. and Müller, G. B. (2006). Before programs: the physical origination of multicellular forms. *Int. J. Dev. Biol.* **50**, 289-299.
Noble, D. (2006). *The Music of Life*. Oxford: Oxford University Press.
Noble, D. (2008). Genes and causation. *Philos. Trans. R. Soc. A* **366**, 3001-3015.
Noble, D. (2011a). Differential and integral views of genetics in computational systems biology. *Interface Focus* **1**, 7-15.
Noble, D. (2011b). Neo-Darwinism, the modern synthesis and selfish genes: are they of use in physiology? *J. Physiol.* **589**, 1007-1015.
Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus* **2**, 55-64.
Noble, D. (2013). Physiology is rocking the foundations of evolutionary biology. *Exp. Physiol.* **98**, 1235-1243.
Noble, D., Jablonka, E., Joyner, M. J., Müller, G. B. and Omholt, S. W. (2014). Evolution evolves: physiology returns to centre stage. *J. Physiol.* **592**, 2237-2244.
Pigliucci, M. and Müller, G. B. (2010). Elements of an extended evolutionary synthesis. In *Evolution: The Extended Synthesis* (ed. M. Pigliucci and G. B. Muller), pp. 3-17. Cambridge, MA: MIT Press.
Poincaré, H. (1902, 1968). *La Science et l'Hypothèse*. Paris: Flammarion.
Rechavi, O., Minevich, G. and Hobert, O. (2011). Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248-1256.
Romanes, G. J. (1883). Letter to the Editor. *Nature* **27**, 528-529.
Sela, M., Kloog, Y. and Rechavi, O. (2014). Non-coding RNAs as the bridge between epigenetic mechanisms, lineages and domains of life. *J. Physiol.* **592**, 2369-2373.
Sun, Y. H., Chen, S. P., Wang, Y. P., Hu, W. and Zhu, Z. Y. (2005). Cytoplasmic impact on cross-genus cloned fish derived from transgenic common carp (*Cyprinus carpio*) nuclei and goldfish (*Carassius auratus*) enucleated eggs. *Biol. Reprod.* **72**, 510-515.
The_Encode_Project_Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.
Waddington, C. H. (1957). *The Strategy of the Genes*. London: Allen and Unwin.
Wanscher, J. H. (1975). An analysis of Wilhelm Johannsen's genetical term 'genotype' 1909-26. *Hereditas* **79**, 1-4.

© 2015. Published by The Company of Biologists Ltd | The Journal of Experimental Biology (2015) 218, 816-818 doi:10.1242/jeb.120071

CLASSICS

Conrad Waddington and the origin of epigenetics

Denis Noble discusses Conrad Waddington's classic paper, 'The genetic assimilation of the bithorax phenotype', published in *Evolution* in 1956.

In 1956, the British developmental biologist, Conrad Waddington, published a paper in the journal *Evolution* (Waddington, 1956) in which he succeeded in demonstrating the inheritance of a characteristic acquired in a population in response to an environmental stimulus. Much earlier, in 1890, August Weismann had tried and failed to achieve this. He amputated the tails of five successive generations of mice and showed absolutely no evidence for an effect on subsequent generations. Weismann's discovery that the effects of an environmental stimulus (tail amputation) cannot be transmitted to subsequent generations, together with his assumption that genetic change is random, formed the foundations of the Modern Synthesis (Neo-Darwinism) of our understanding of genetic inheritance.

Waddington's approach, however, was much more subtle and more likely to be successful because he realised that the way to test for the inheritance of acquired characteristics is first to discover what forms of developmental plasticity already exist in a population, or that the population could be persuaded to demonstrate with a little nudging from the environment. By exploiting plasticity that already existed he was much more likely to mimic a path that evolution itself could have taken.

He used the word 'canalised' for this kind of persuasion since he represented the developmental process as a series of 'decisions' that could be represented as 'valleys' and 'forks' in a developmental landscape (Fig. 1). He knew from his developmental studies that embryo fruit flies could be persuaded to show different thorax and wing structures, simply by changing the environmental temperature or by a chemical stimulus. In his landscape diagram, this could be represented as a small manipulation in slope that would lead to one channel in the landscape being favoured over another, so that the adult could show a different phenotype starting from the same genotype.

The next step in his experiment was to select for and breed from the animals that displayed the new characteristic. Exposed to the same environmental stimulus, these gave rise to progeny with an even higher proportion of adults displaying the new character. After a relatively small number of generations, he found that he could then breed from the animals and obtain robust inheritance of the new character even without applying the environmental stimulus. The characteristic had therefore become locked into the genetics of the animal. He called this process genetic assimilation. What he had succeeded in showing was that an acquired characteristic could first be inherited as what we would now call 'soft' inheritance, and that it could then be assimilated into becoming standard 'hard' genetic inheritance. Today, we call 'soft' inheritance epigenetic inheritance, and of course, we know many more mechanisms by which the same genome can be controlled to produce different epigenetic effects.

What was happening at the gene level in Waddington's experiments? A standard Neo-Darwinist explanation might be that some mutations occurred. That is possible, but extremely unlikely on the time scale of the experiment, which was only a few generations. Moreover, random mutations would occur in individuals, not in a whole group. Single small mutations would have taken very many generations to spread through whole populations, and many such mutations would have been required.

But I think there is a much simpler explanation. Recall that the experiment exploited plasticity that is already present in the population. That strongly suggests that all the alleles (gene variants) necessary for the inheritance of the characteristic were already present in the population, but not initially in any particular individuals in the correct combination. The experiment simply brings them together. This is a modification of the pattern of the genome in response to the environmental change, but not in a way that requires any new mutations. I came to this conclusion before reading Waddington's (1957) book, *The Strategy of the Genes*. But it is in fact one of Waddington's own ideas! He writes 'There is no … reason which would prevent us from imagining that all the genes which eventually make up the assimilated genotype were already present in the population before the selection began, and only required bringing together' (p. 176). Not only does he clearly see this possibility, he also tests it. He continues (p. 178) 'Attempts to carry out genetic assimilation starting from inbred lines have remained quite unsuccessful. This provides further evidence that the process depends on the utilisation of genetic variability in the foundation stock with which the experiment begins'. His text could not be clearer.

Orthodox Neo-Darwinists dismissed Waddington's findings as merely an example of the evolution of phenotype plasticity. That is what you will find in many of the biology textbooks even today (e.g. Arthur, 2010). I think that Waddington showed more than that. Of course, plasticity can evolve, and that itself could be by a Neo-Darwinist or any other mechanism. But Waddington was not simply showing the evolution of plasticity in general; he was showing how it could be exploited to enable a particular acquired characteristic in response to an environmental change to be inherited and be assimilated into the genome. Moreover, he departed from the strict Neo-Darwinist view by showing that this could happen even if no new mutations occur (Fig. 2).

Epigenetics means 'above genetics' and it was originally conceived by

Classics is an occasional column, featuring historic publications from the literature. These articles, written by modern experts in the field, discuss each classic paper's impact on the field of biology and their own work.

The Journal of Experimental Biology

The Journal of Experimental Biology (2015) 218, 816-818 doi:10.1242/jeb.120071

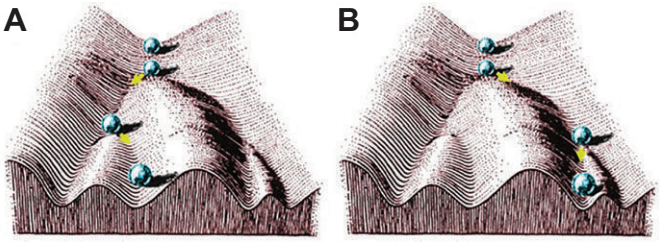


Fig. 1. Waddington's developmental landscape diagram. The landscape itself and the ball at the top are from his original diagram. The subsequent positions of the ball have been added to illustrate his point that development can be canalised to follow different routes (A and B). The plasticity to enable this to happen already exists in the wild population of organisms (modified diagram by K. Mitchell).

Waddington himself to describe the existence of mechanisms of inheritance in addition to (over and above) standard genetics (Bard, 2008). Waddington regarded himself as a Darwinist since Darwin also, in *The Origin of Species*, included the inheritance of acquired characteristics. But significantly, Waddington was not a Neo-Darwinist since Neo-Darwinism, following Weismann, specifically excludes such inheritance. Waddington was a profound thinker about biology, and much else too. *The Strategy of the Genes* is a masterly account of the many reasons why he dissented from Neo-Darwinism, and it has stood the test of time. It was reprinted over half a century later, in 2014. He did not describe himself as a Lamarckian, but by revealing mechanisms of inheritance of acquired characteristics, I think he should be regarded as such. The reason he did not do so is that Lamarck could not have conceived of the processes that Waddington revealed. Incidentally, it is also true to say that Lamarck did not invent the idea of the inheritance of acquired characteristics. But, whether historically correct or not, we are stuck today with the term 'Lamarckian' for inheritance of a characteristic acquired through an environmental influence.

Waddington's concepts of plasticity and epigenetics have been very influential in my own thinking about experiments on cardiac rhythm. We found that the heart's pacemaker is very robust, so much so that protein mechanisms normally responsible for a large part of the rhythm could be completely blocked or deleted (Noble et al., 1992). Only very small changes in rhythm occur, because other mechanisms come into play to ensure that pacemaker activity continues. The relation between individual genes and the phenotype is therefore mediated through networks of interactions that can buffer individual gene variation, just as Waddington envisaged in his diagrams of epigenetic effects and canalisation. This is one of the reasons why I became interested in evolutionary biology many years ago, and why I have also explored ways in which evolutionary theory can be integrated with recent discoveries in molecular and physiological biology (Noble et al., 2014).

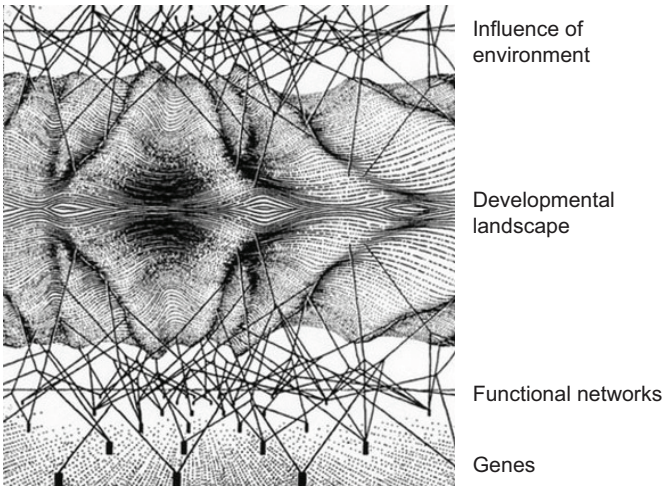


Fig. 2. Waddington's diagram to show how the developmental landscape relates to individual genes (bottom pegs) through networks of interactions in the organism. Since he also showed the influence of the external environment on canalisation of development, I have extended the diagram by adding the top part to represent the environmental influences. It is the combination of these influences that can lead to an evolutionary change without mutations (modified from Waddington, 1957).

Waddington's concepts are also highly relevant to biologists interested in the ways in which organisms adapt to their environment, and to comparative biologists interested in how this varies between species. Many of the ways in which modern epigenetics plays an essential role in these fields have been described in a special issue of this journal (see overview by Knight, 2015). The discovery of epigenetic marking of DNA and its associated chromatin proteins has opened up new vistas for experimental biology.

I conclude this article with a warning: if you are inspired to try to repeat Waddington's 1956 experiment, do remember that you will fail if you try to do it on a cloned laboratory population. The mechanism depends on using a wild population with natural genetic diversity. In this respect it resembles a phenomenon first noted by James Baldwin (1896). This is that individuals in a population with the 'correct' allele combinations could choose a new environment and so permanently change the evolutionary development in that environment. It resembles Waddington's idea, as he himself recognised, because it does not require new mutations. More recently, Karl Popper, the great logician of science, also noted the possible importance of genetic assimilation without mutations in evolutionary theory (Niemann, 2014; Noble, 2014). Popper and Waddington had both taken part in discussions on evolutionary biology during the 1930s and 1940s when the field of molecular biology was still developing (Niemann, 2014).

While celebrating the recent rapid rise in epigenetics research (see Hoppeler, 2015; Knight, 2015), let's also celebrate the father of epigenetics, Conrad Waddington, who opened our eyes to the rich opportunities of adaptation through epigenetic regulation.

Denis Noble
University of Oxford
denis.noble@physiol.ox.ac.uk

The Journal of Experimental Biology (2015) 218, 816-818 doi:10.1242/jeb.120071

References

Arthur, W. (2010). *Evolution, A Developmental Approach*. Oxford, UK: Wiley Blackwell.

Baldwin, J. M. (1896). A new factor in evolution. *Am. Nat.* **30**, 441-451.

Bard, J. B. L. (2008). Waddington's legacy to developmental and theoretical biology. *Biol. Theory* **3**, 188-197.

Hoppeler, H. H. (2015). Epigenetics in comparative physiology. *J. Exp. Biol.* **218**, 6.

Knight, K. (2015). A comparative perspective on epigenetics. *J. Exp. Biol.* **218**, 1-5.

Niemann, H.-J. (2014). *Karl Popper and The Two New Secrets of Life*. Tübingen: Mohr Siebeck.

Noble, D. (2014). Secrets of life from beyond the grave. *Physiol. News* **97**, 34-35.

Noble, D., Denyer, J. C., Brown, H. F. and DiFrancesco, D. (1992). Reciprocal role of the inward currents $i_{b,Na}$ and i_f in controlling and stabilizing pacemaker frequency of rabbit Sino-atrial node cells. *Proc. R. Soc. B. Biol. Sci.* **250**, 199-207.

Noble, D., Jablonka, E., Joyner, M. J., Müller, G. B. and Omholt, S. W. (2014). Evolution evolves: physiology returns to centre stage. *J. Physiol.* **592**, 2237-2244.

Waddington, C. H. (1956). The genetic assimilation of the bithorax phenotype. *Evolution* **10**, 1-13.

Waddington, C. H. (1957). *The Strategy of the Genes*. London: Allen and Unwin. Reprinted 2014.

© 2015. Published by The Company of Biologists Ltd | The Journal of Experimental Biology (2015) 218, 816-818 doi:10.1242/jeb.120071

Conrad Waddington and the origin of epigenetics

Denis Noble discusses Conrad Waddington's classic paper, 'The genetic assimilation of the bithorax phenotype', published in *Evolution* in 1956.

In 1956, the British developmental biologist, Conrad Waddington, published a paper in the journal *Evolution* (Waddington, 1956) in which he succeeded in demonstrating the inheritance of a characteristic acquired in a population in response to an environmental stimulus. Much earlier, in 1890, August Weismann had tried and failed to achieve this. He amputated the tails of five successive generations of mice and showed absolutely no evidence for an effect on subsequent generations. Weismann's discovery that the effects of an environmental stimulus (tail amputation) cannot be transmitted to subsequent generations, together with his assumption that genetic change is random, formed the foundations of the Modern Synthesis (Neo-Darwinism) of our understanding of genetic inheritance.

Waddington's approach, however, was much more subtle and more likely to be successful because he realised that the way to test for the inheritance of acquired characteristics is first to discover what forms of developmental plasticity already exist in a population, or that the population could be persuaded to demonstrate with a little nudging from the environment. By exploiting plasticity that already existed he was much more likely to mimic a path that evolution itself could have taken.

He used the word 'canalised' for this kind of persuasion since he represented the developmental process as a series of 'decisions' that could be represented as 'valleys' and 'forks' in a developmental landscape (Fig. 1). He knew from his developmental studies that embryo fruit flies could be persuaded to show different thorax and wing structures, simply by changing the environmental temperature or by a chemical stimulus. In his landscape diagram, this could be represented as a small manipulation in slope that would lead to one channel in the landscape being favoured over another, so that the adult could show a different phenotype starting from the same genotype.

The next step in his experiment was to select for and breed from the animals that displayed the new characteristic. Exposed to the same environmental stimulus, these gave rise to progeny with an even higher proportion of adults displaying the new character. After a relatively small number of generations, he found that he could then breed from the animals and obtain robust inheritance of the new character even without applying the environmental stimulus. The characteristic had therefore become locked into the genetics of the animal. He called this process genetic assimilation. What he had succeeded in showing was that an acquired characteristic could first be inherited as what we would now call 'soft' inheritance, and that it could then be assimilated into becoming standard 'hard' genetic inheritance. Today, we call 'soft' inheritance epigenetic inheritance, and of course, we know many more mechanisms by which the same genome can be controlled to produce different epigenetic effects.

What was happening at the gene level in Waddington's experiments? A standard Neo-Darwinist explanation might be that some mutations occurred. That is possible, but extremely unlikely on the time scale of the experiment, which was only a few generations. Moreover, random mutations would occur in individuals, not in a whole group. Single small mutations would have taken very many generations to spread through whole populations, and many such mutations would have been required.

But I think there is a much simpler explanation. Recall that the experiment exploited plasticity that is already present in the population. That strongly suggests that all the alleles (gene variants) necessary for the inheritance of the characteristic were already present in the population, but not initially in any particular individuals in the correct combination. The experiment simply brings them together. This is a modification of the pattern of the genome in response to the environmental change, but not in a way that requires any new mutations. I came to this conclusion before reading Waddington's (1957) book, *The Strategy of the Genes*. But it is in fact one of Waddington's own ideas! He writes 'There is no … reason which would prevent us from imagining that all the genes which eventually make up the assimilated genotype were already present in the population before the selection began, and only required bringing together' (p. 176). Not only does he clearly see this possibility, he also tests it. He continues (p. 178) 'Attempts to carry out genetic assimilation starting from inbred lines have remained quite unsuccessful. This provides further evidence that the process depends on the utilisation of genetic variability in the foundation stock with which the experiment begins'. His text could not be clearer.

Orthodox Neo-Darwinists dismissed Waddington's findings as merely an example of the evolution of phenotype plasticity. That is what you will find in many of the biology textbooks even today (e.g. Arthur, 2010). I think that Waddington showed more than that. Of course, plasticity can evolve, and that itself could be by a Neo-Darwinist or any other mechanism. But Waddington was not simply showing the evolution of plasticity in general; he was showing how it could be exploited to enable a particular acquired characteristic in response to an environmental change to be inherited and be assimilated into the genome. Moreover, he departed from the strict Neo-Darwinist view by showing that this could happen even if no new mutations occur (Fig. 2).

Epigenetics means 'above genetics' and it was originally conceived by

Classics is an occasional column, featuring historic publications from the literature. These articles, written by modern experts in the field, discuss each classic paper's impact on the field of biology and their own work.

The Journal of Experimental Biology (2015) 218, 816-818 doi:10.1242/jeb.120071

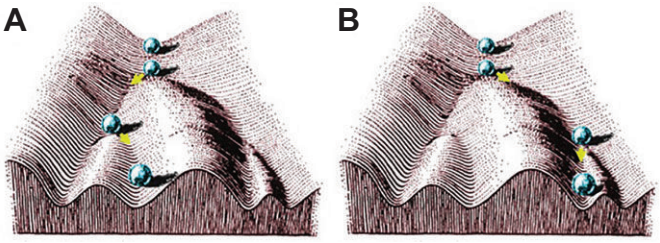


Fig. 1. Waddington's developmental landscape diagram. The landscape itself and the ball at the top are from his original diagram. The subsequent positions of the ball have been added to illustrate his point that development can be canalised to follow different routes (A and B). The plasticity to enable this to happen already exists in the wild population of organisms (modified diagram by K. Mitchell).

Waddington himself to describe the existence of mechanisms of inheritance in addition to (over and above) standard genetics (Bard, 2008). Waddington regarded himself as a Darwinist since Darwin also, in *The Origin of Species*, included the inheritance of acquired characteristics. But significantly, Waddington was not a Neo-Darwinist since Neo-Darwinism, following Weismann, specifically excludes such inheritance. Waddington was a profound thinker about biology, and much else too. *The Strategy of the Genes* is a masterly account of the many reasons why he dissented from Neo-Darwinism, and it has stood the test of time. It was reprinted over half a century later, in 2014. He did not describe himself as a Lamarckian, but by revealing mechanisms of inheritance of acquired characteristics, I think he should be regarded as such. The reason he did not do so is that Lamarck could not have conceived of the processes that Waddington revealed. Incidentally, it is also true to say that Lamarck did not invent the idea of the inheritance of acquired characteristics. But, whether historically correct or not, we are stuck today with the term 'Lamarckian' for inheritance of a characteristic acquired through an environmental influence.

Waddington's concepts of plasticity and epigenetics have been very influential in my own thinking about experiments on cardiac rhythm. We found that the heart's pacemaker is very robust, so much so that protein mechanisms normally responsible for a large part of the rhythm could be completely blocked or deleted (Noble et al., 1992). Only very small changes in rhythm occur, because other mechanisms come into play to ensure that pacemaker activity continues. The relation between individual genes and the phenotype is therefore mediated through networks of interactions that can buffer individual gene variation, just as Waddington envisaged in his diagrams of epigenetic effects and canalisation. This is one of the reasons why I became interested in evolutionary biology many years ago, and why I have also explored ways in which evolutionary theory can be integrated with recent discoveries in molecular and physiological biology (Noble et al., 2014).

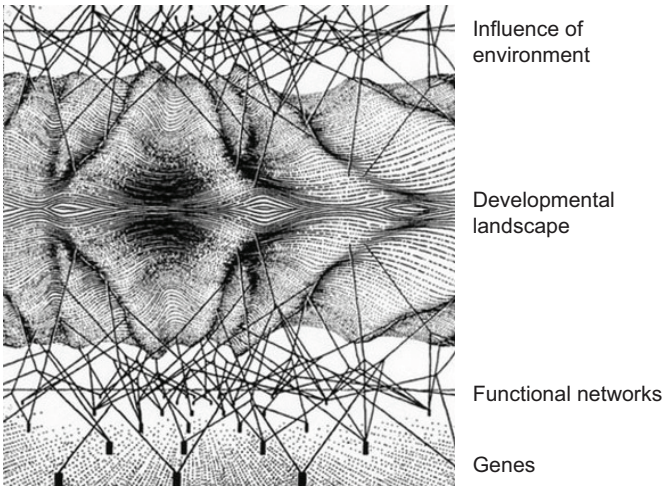


Fig. 2. Waddington's diagram to show how the developmental landscape relates to individual genes (bottom pegs) through networks of interactions in the organism. Since he also showed the influence of the external environment on canalisation of development, I have extended the diagram by adding the top part to represent the environmental influences. It is the combination of these influences that can lead to an evolutionary change without mutations (modified from Waddington, 1957).

Waddington's concepts are also highly relevant to biologists interested in the ways in which organisms adapt to their environment, and to comparative biologists interested in how this varies between species. Many of the ways in which modern epigenetics plays an essential role in these fields have been described in a special issue of this journal (see overview by Knight, 2015). The discovery of epigenetic marking of DNA and its associated chromatin proteins has opened up new vistas for experimental biology.

I conclude this article with a warning: if you are inspired to try to repeat Waddington's 1956 experiment, do remember that you will fail if you try to do it on a cloned laboratory population. The mechanism depends on using a wild population with natural genetic diversity. In this respect it resembles a phenomenon first noted by James Baldwin (1896). This is that individuals in a population with the 'correct' allele combinations could choose a new environment and so permanently change the evolutionary development in that environment. It resembles Waddington's idea, as he himself recognised, because it does not require new mutations. More recently, Karl Popper, the great logician of science, also noted the possible importance of genetic assimilation without mutations in evolutionary theory (Niemann, 2014; Noble, 2014). Popper and Waddington had both taken part in discussions on evolutionary biology during the 1930s and 1940s when the field of molecular biology was still developing (Niemann, 2014).

While celebrating the recent rapid rise in epigenetics research (see Hoppeler, 2015; Knight, 2015), let's also celebrate the father of epigenetics, Conrad Waddington, who opened our eyes to the rich opportunities of adaptation through epigenetic regulation.

Denis Noble
University of Oxford
denis.noble@physiol.ox.ac.uk

The Journal of Experimental Biology (2015) 218, 816-818 doi:10.1242/jeb.120071

References

Arthur, W. (2010). *Evolution, A Developmental Approach*. Oxford, UK: Wiley Blackwell.

Baldwin, J. M. (1896). A new factor in evolution. *Am. Nat.* **30**, 441-451.

Bard, J. B. L. (2008). Waddington's legacy to developmental and theoretical biology. *Biol. Theory* **3**, 188-197.

Hoppeler, H. H. (2015). Epigenetics in comparative physiology. *J. Exp. Biol.* **218**, 6.

Knight, K. (2015). A comparative perspective on epigenetics. *J. Exp. Biol.* **218**, 1-5.

Niemann, H.-J. (2014). *Karl Popper and The Two New Secrets of Life*. Tübingen: Mohr Siebeck.

Noble, D. (2014). Secrets of life from beyond the grave. *Physiol. News* **97**, 34-35.

Noble, D., Denyer, J. C., Brown, H. F. and DiFrancesco, D. (1992). Reciprocal role of the inward currents $i_{b,Na}$ and i_f in controlling and stabilizing pacemaker frequency of rabbit Sino-atrial node cells. *Proc. R. Soc. B. Biol. Sci.* **250**, 199-207.

Noble, D., Jablonka, E., Joyner, M. J., Müller, G. B. and Omholt, S. W. (2014). Evolution evolves: physiology returns to centre stage. *J. Physiol.* **592**, 2237-2244.

Waddington, C. H. (1956). The genetic assimilation of the bithorax phenotype. *Evolution* **10**, 1-13.

Waddington, C. H. (1957). *The Strategy of the Genes*. London: Allen and Unwin. Reprinted 2014.