The Harnessing of Stochasticity

Physiological and philosophical consequences

Articles by Raymond and Denis Noble

This pdf collection of 9 articles brings together a developing story. It can be seen as the culmination of many years of discussions between the two of us on the nature of biology, what causes what and how, and what evolves and how. Those discussions long predate the first articles to appear.

The opening shot in print was the 2012 formulation in *Interface Focus* of the mathematical basis of the principle of biological relativity, meaning relativity of causation by different levels of organisation in organisms, which had already been implied in *The Music of Life* (2006). That principle is insufficient on its own since there has also to be a specific process by which lower level processes are constrained by higher-level organisation.

The beginnings of identifying that process as including the harnessing of stochasticity occurred at the 2016 meeting at The Royal Society on New Trends in Evolutionary Biology, published in *Interface Focus* in 2017. But that idea was not followed up in the Discussion Meeting itself, except for a very helpful chance meeting we had with the immunologist Richard Moxon during breakfast the next day.

What really set discussion alight was the 2017 article in *Biology*, which was the subject of a very lengthy interaction with a neo-darwinist referee. That was followed by a cascade of invitations for the subsequent articles published in a variety of journals leading up to the 2020 article in the *Journal for the General Philosophy of Science*.

Raymond and Denis Noble, October 2020.

1. The physiology of agency

Can Reasons and Values Influence Action: How Might Intentional Agency Work Physiologically? *Journal for General Philosophy of Science*. 2020. https://doi.org/10.1007/s10838-020-09525-3

2. "Active" Darwinism and Karl Popper

Rehabilitation of Karl Popper's ideas on evolutionary biology and the nature of biological science. In *Karl Popper: His Philosophy and Science*. Springer-Nature 2021 In press

3. Boundaries and causation

Biological Relativity Requires Circular Causality but Not Symmetry of Causation: So, Where, What and When Are the Boundaries? *Frontiers in Physiology*. 2019. **10**, 827.

4. Emergent properties

A-mergence of biological systems. In *The Routledge Handbook of Emergence*. 387-399. 2020. In press.

5. How organisms make choices.

Harnessing stochasticity: How do organisms make choices? 2018. *Chaos* **28** 106309.

6. Is evolution blind?

Was the Watchmaker Blind? Or Was She One-Eyed? 2017. *Biology*, **6**, 47.

7. Harnessing of Stochasticity

Evolution viewed from physics, physiology and medicine, *Interface Focus*, 7, 20160159.

8. The principle of Biological Relativity

A theory of biological relativity: no privileged level of causation. 2012. *Interface Focus*. **2**, 55-64.

9. Artificial Intelligence and Agency

Forum: Artificial Intelligence, Artificial Agency and Artificial Life
RUSI Journal, Vol. 164, No. 5, July 2019

Journal for General Philosophy of Science
https://doi.org/10.1007/s10838-020-09525-3

ARTICLE

Can Reasons and Values Influence Action: How Might Intentional Agency Work Physiologically?

Raymond Noble[1] · **Denis Noble**[2]

Accepted: 18 September 2020

Abstract
In this paper, we demonstrate (1) how harnessing stochasticity can be the basis of creative agency; (2) that such harnessing can resolve the apparent conflict between reductionist (micro-level) accounts of behaviour and behaviour as the outcome of rational and value-driven (macro-level) decisions; (3) how neurophysiological processes can instantiate such behaviour; (4) The processes involved depend on three features of living organisms: (a) they are necessarily open systems; (b) micro-level systems therefore nest within higher-level systems; (c) causal interactions must occur across all the boundaries between the levels of organization. The higher levels constrain the dynamics of lower levels. The experimental evidence and theoretical arguments are shown to be consistent with previous research on the neuronal mechanisms of conscious choice, and with the interconnected multi-level processes by which organisms harness stochasticity, whether conscious or unconscious.

Keywords Harnessing stochasticity · Free choice · Free will · Intentionality · Agency · Micro-level causation · Macro-level causation · Reductionism · Holism

1 Introduction

In recent articles (Noble 2017; Noble and Noble 2018; Noble and Noble 2017), we have shown how organisms can harness stochasticity in ways that serve functional needs. Organisms are not merely passive recipients of random variations at molecular and other levels. They can use stochasticity both to guide their behaviour creatively and to influence the directions in which they evolve. We have also shown that harnessing stochasticity can be the basis on which organisms make creative choices (Noble and Noble 2018). In that article, we drew attention to close parallels with the work of Karl Popper on evolution and free choice (Niemann 2014; Popper 1945; 1972; 1973; Popper and Eccles 1977).

✉ Denis Noble
denis.noble@dpag.ox.ac.uk

Raymond Noble
r.noble@ucl.ac.uk

[1] Institute for Women's Health, University College London, London WC1E 6AU, UK

[2] Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford OX1 3PT, UK

Published online: 28 October 2020

In this paper, we will explore whether harnessing stochasticity could be the basis of resolving a long-standing philosophical problem: the tension between what are thought to be standard empirical (usually mechanistic) scientific accounts of organism behaviour and views of life that regard behaviour as the outcome of rational and value-driven decisions. We will refer to the first as micro-level and the second as macro-level explanations.

On the macro-level view, organisms can act rationally in the sense that we can give (in the case of humans) or we can see (in the case of other organisms) good social reasons why they chose to act in the way they do. The micro-level view is often taken by empirical scientists to be that those reasons were not the cause of the behaviour. On that account, the behaviour was entirely accountable by observing the physical processes involved. This is the basis of the reductionist view of biology. Could both accounts be somehow correct, or must there be a, perhaps unresolvable, tension between them?

On any micro-level view, the tension is evident. If determinate molecular and other physical forces wholly and always determine the actions of organisms, and if there is causal closure, then how could a macro-level explanation involving rational or value-driven action possibly be correct? The philosopher Jaegwon Kim, for example, argues persuasively that if there is complete causal closure at the microphysical (e.g. molecular) level then there is no room for additional causation from macro-levels (Kim 2000). The feelings and thoughts we have as humans must then be an illusion (Heisenberg 2009; Midgley 2014), and to work it has to be assumed to be a very strong illusion. That is precisely what incompatibilist evolutionary biologists suppose:

> "The illusion of agency is so powerful that even strong incompatibilists like myself will always *act* as if we had choices, even though we know that we don't. We have no choice in this matter. But we can at least ponder why evolution might have bequeathed us such a powerful illusion."(Coyne 2014)

Intention and will can be admitted to exist (how could we possibly deny that they exist as *experienced* by us?), but on this view they are not causally involved.

In this article we challenge two assumptions in the purely micro-level view. The first is the possibility of causal closure in living organisms, meaning that there is no room for any other influence, particularly that of intention; the second is the assumption that any stochasticity involved cannot itself be functional.

The assumption of causal closure cannot be true of living organisms since they are necessarily open systems (for recent accounts of organisms as open systems, see (Capra and Luisi 2014; Noble 2016). In this paper, we will explain the consequences for causal explanation at both micro- and macro-levels (see Sect. 4g).

2 The Origins of Stochasticity

Thermal and quantal noise exists at the molecular, atomic and subatomic levels (Del Santo and Gisin 2019), while the causes that organisms encounter in their interactions with other organisms and with their environments include many forms of chance occurrence. However, these facts, by themselves, are insufficient to resolve the tension between macro- and micro-level explanations. There are degrees of uncertainty. Intention and will are not claimed to be just chance. They are claimed to be causally influential on the macro-level view. Thus, mimicking organisms by simply introducing stochastic equations into otherwise determinate algorithms in artificial intelligence systems can generate novel and

unpredictable behaviour but does not create what we call agency (Noble and Noble 2019), i.e. when organisms make decisions or have an intention to act. Introducing unharnessed stochasticity is merely an additional reductive process.

Furthermore, regarding biological systems as determinate, but requiring the addition of a creative element, is to misunderstand that biological systems *are* organisationally creative. We will also explain the kind of causality involved. For that, we need to demonstrate a rationalising biological process that enables organisms to make choices predicated on it.

We will outline physiological processes that can enable rational and value-driven decisions in behaviour to be causally influential. The evidence is that harnessing stochasticity in the way in which we described this process in (Noble and Noble 2018) could be the basis of agency and therefore of free, rational action. It is the organism that is doing what it is doing with the intention to achieve an objective. In a previous article (Noble et al. 2019) we took this approach further to characterise more precisely how downward (i.e. macro-level) causation is effective physiologically. We gave detailed examples where choices of life-style influence molecular events that control our genes.

We will also show that the processes we outline could be compatible with recent philosophical work on the nature of free action, specifically the account presented in the work of Christian List (2019)—see Sect. 6.

3 Causation

Some of the arguments in this paper turn on the meaning of causation. Not all causes are of the same type. We use the categories of causation derived from Aristotle. A version of these categories of use in multi-level biology can be found in Noble (2016, 176–181), which includes any factors whose alteration would result in changes in behaviour. Specifically we distinguish between the *dynamics* of a system and the *conditions* under which those dynamics play out. Those conditions can then be seen as constraining the dynamics, just as the dimensions and elasticity of the container of a gas determine the pressure and temperature generated by the dynamics of the molecular movements. In the biology of conscious organisms, constraints will necessarily include social/cultural factors. Some of those will be logical ones, including what in ordinary language we would refer to as the reasons for an organism's behaviour. This issue is analysed further in Sect. 7 of the paper.

4 The Argument

4.1 Why Do We Need a Macro-Level Account?

The tension between micro-level and macro-level accounts might be 'solved' by claiming that they are merely different viewpoints. The micro-level account is then regarded as the (only?) scientific one, while the macro-level account is the (necessary for ordinary social interactions?) philosophical one. However, this solution is unsatisfactory not only because it fails to explain how rational thought can influence behaviour, and how this can be a basis of 'free' choice, but also because it matters to society whether we believe we are living under a strong illusion (Vohs and Schooler 2008). Socially it begs the question what illusion we are living with. If rational thought is not influential, why then do we also need the

macro-level (common sense) explanation? Why is it necessary? And why should we even strive to act according to what we feel and want?

The answer is that, as Coyne acknowledges (we "always *act* as if we had choices"), we need the macro-level account because, in practice, we cannot manage without it. In the real world, we need to treat other people and many other organisms as intentional beings with real rational choices to make. When we ask someone else "why did you do that?" we do not usually expect the answer "my brain/genes, or whatever, made me do it." (Murphy and Brown 2007; Lim 2008). We expect a rational explanation. That is so even if it may be a rational*ised* explanation. Whether we or the other organism believe the explanation or not, we need it in order to continue a rational interaction and often to make choices. Interactions between predator and prey often involve anticipating the behaviour of each other. This is also true of human interactions since our intentions are predicated on the intentions of others. In this sense the 'illusion' of intention is powerful because we act upon it. Merely calling intentionality an illusion does not solve the problem; it merely parks it.

Such interaction may also change the other person's perception of their behaviour, e.g. by giving the reasons why their rationalisation, or reasoning, cannot be correct, and this interaction influences what they do. We do this all the time in argumentation, and in legal contexts, yet none of it appears to make sense in a purely micro-level explanation (Ellis 2016; Midgley 2014).

In fact, on the micro-level explanation, someone's reasons for the action they took would *never* be the 'real' causes. At best they might be shorthand for the micro-level account. This is the approach taken by some evolutionary biologists. As an example, Maynard Smith (1998, 213) wrote:

> I am prepared to think as loosely as necessary to give me an idea when I'm confronted with a new biological problem…..But when I've got an idea, I want to be able to write down the equations and show that the idea works.

Purposive accounts of behaviour or function in organisms are then temporary shorthand for mechanistic accounts in terms of gene variations and natural selection. It matters to decide whether this 'shorthand' view is correct. We will show in Sect. 4(g) that it matters what kind of mathematics is used.

4.2 Physiological Evidence for Micro-Level Causal Efficacy of Rational Choice

Recent physiological research shows that it is possible to measure micro-level effects contingent upon macro-level decision-making. As an example, consider the case where an athlete chooses, as his life-style, to train hard and regularly. This decision subjects his body to many physiological challenges as he deliberately pushes these physiological processes to their limits. Many of the decisions he or she makes will be influenced by reading articles on fitness, or interactions and advice from his trainer, and by observing others. These decisions are made at a social level and are logically based on an understanding of physiology and training. Any particular form of exercise is performed to produce a given result. We have discussed this case in detail in a previous article (Noble et al. 2019). What happens? One thing that happens over time is that the decision and training as an athlete cause RNA changes that enable the production of more of the right muscle proteins that enhance his athletic ability (Bathgate et al. 2018). Indeed, much more than that happens. His cardiovascular fitness improves, he gets better at making decisions, when to run faster and how to pace himself. He may also alter his diet and make choices about what to eat and

when. The rational process in making these decisions therefore had a profound effect at both macro- and micro-levels. It seems implausible to argue that all of these changes could be determined purely by the evolutionary process since they occur during the lifetime of the individual.

Gene-determinists might nevertheless argue that the athlete just had the right genes to enable him to be a successful athlete (Plomin 2018), whether or not he uses them. That must be true, of course, at least in an *enabling* sense (Epstein 2014). However, so did his identical twin brother who did not become an athlete (Bathgate et al. 2018). Identical twins are born with the same genome. The difference between them is in no small measure the decision by the athletic brother to exercise regularly, to change his lifestyle. Genes were a necessary condition, but they do not and obviously cannot fully explain the difference between identical twin brothers.

The difference between the identical twins might be explained in part by their different experiences in the womb. Being identical in a genome sense does not mean they developed identically in the womb. Blood supply is often not evenly distributed between twins, so their nutrient experience may be different, leading to a difference in metabolic strategy and also possibly in muscle type. They may also have experienced different levels of oestrogen and testosterone in the womb. We should be aware of this alternative explanation for differences between twins since those prenatal influences might predispose one twin rather than the other in a particular direction. But the fact remains that the RNA changes will occur only if and when the life-style choice is made and maintained. Anything that prevents that from happening would also prevent the micro-level RNA and similar changes from happening.

Physiologists are identifying many other molecular epigenetic changes consequent upon life-style social choices that change how gene expression levels are varied. These experiments show that even on the micro-level approach of measuring low-level molecular changes, rational choices have causal effects. Thus, they show that the issue between micro-level and macro-level accounts is not a purely conceptual problem. It is an empirical matter that it can be and is resolved in the sense that a rational choice influenced the physical events.

4.3 Evidence from Observations of Non-Human Animals

Other organisms are also capable of this kind of agency. When chimpanzees or bonobos use body language and signs, such as hand gestures and facial expression, to communicate with each other they do so anticipating the possible responses of each other. The precise meaning of a sign is dependent on the contextual logic. They act with reason and anticipate with reason. Language is like a tool to convey anticipatory states (Byrne et al. 2017; Graham et al. 2017, 2018; Hobsiter and Byrne 2014).

As a further example, when a chimpanzee uses a stone to crack-open a nut, it is not a hard-wired biogenically determined event. It is a creative choice. Furthermore, chimpanzees choose stones that are better suited for cracking nuts, and in that sense therefore the behaviour is rational (Whiten 2017). Chimpanzees learn this by observing others and by their experience, solving a problem by using tools. Tools are purposeful because of the agency of their creators. The question is not whether but how biological systems achieve this creativity.

If there is perceived to be a tension between micro-level and macro-level explanations, then we contend that physiological processes themselves, as open systems, show that the

tension is in principle resolved, and these physiological findings can demonstrate that. The question remaining is how is this possible? What kind of physiological processes could account for it?

4.4 The Proposed Resolution

We will now outline how such physiological changes can, *at the same time*, have identifiable molecular biological processes by which they occur and also satisfy the criterion of rational choice. To achieve that we need to outline a physiological process that would explain why we are not necessarily able to make predictions in advance but the choice will appear rational in retrospect. The philosopher John Lucas expressed this requirement in his book *The Freedom of the Will*:

> For the reasons which "determine" a composer to add a particular bar to his composition are parts of some sort of rational explanation, not a regularity one; they do not enable us to make predictions in advance, but only to see how right it was *ex post facto*. It is the mark of creative genius that it is original and unpredictable; although after it has manifested itself, its *rationale* is manifest also. (Lucas 1970)

Using the example of creative genius is simply a way of highlighting this difference between what we can say in advance and the rationale we can give afterwards. The difference is to varying degrees characteristic of all rational behaviour.

The reason why this conjunction seems impossible is that molecular processes, in themselves, cannot have values and meaning of the kind that is necessary to satisfy the criterion of rationality. A micro-level approach does not find the rational self. The idea that genes make us selfish, or that the overriding purpose of our behaviour is to preserve genes in a gene pool, is to impart purpose (albeit metaphorically) to an inanimate bit of the micro-level system yet deny it to a macro-level organism. DNA, RNA or protein sequences, and the lipids and metabolites with which they interact in functional biological networks have no meaning outside the context of a living organism that can give them meaning.

This point about gene sequences is valid in the same way in which alphabetic sequences in a language have no meaning outside the context of the language speakers who choose to give them meaning. Thus, the sequence B U T has completely different meanings in English and French. Similarly, gene sequences involved in the development of arms and legs have entirely different functional instantiations in organisms without arms or legs. A single gene (*per* in the fruit fly) can be involved in functions as different as circadian rhythm, long-term memory, cancer development, and courtship behaviour (Foster and Kreitzman 2004; Sakai et al. 2004). Genes are not by themselves causal. They can have no property of motivation or agency.

Furthermore, in organisms, the molecular sequences do not have meaning in isolation. Gene ontology (the giving of names to genes based on their functional involvement in the functions of the organism) may give the popular impression of 'genes for this' and 'genes for that'. What the genome-wide association studies (GWAS) have shown is that this is far from reality. Many DNA and RNA sequences are involved in every function of the body. Some GWAS scientists even favour the omnigenic hypothesis (Boyle et al. 2017), which surmises that all genes are involved in one way or another in all functions. That is another way of expressing the fact that all molecular sequences have no meaning outside the context of the complete organism.

These are important scientific discoveries. However, the central conceptual tension remains. If there was only one, determinate, molecular level response to a given challenging situation for the organism, the chances of it *also* happening to be the one that at the molecular level instantiates the rational decision at a macro-level would be so small as effectively to be zero. We base our claim on the fact that this is not how organisms react to challenging situations. We need to abandon the twin ideas of linear-sequenced causality and privileged level causality. As we showed in (Noble et al. 2019), the forms of causation that act across boundaries necessarily act simultaneously, and they are necessarily different forms of causation.

We will first show what organisms can do in a non-neuronal (or unconscious) process and then explain how similar processes of harnessing stochasticity must occur in nervous, including conscious, processes.

In both cases, organisms can harness stochasticity in ways that generate rational (i.e. guided) responses to environmental challenges. The empirical foundation was laid out in (Noble 2017) and applied to agency driven processes in (Noble and Noble 2017; Noble and Noble 2018). Organisms have demonstrably evolved *guided* random mutation and other molecular mechanisms that can respond rapidly and correctly to environmental challenges. These processes allow organisms and populations to harness stochasticity to evolve solutions to such challenges relatively fast compared to the accumulation of non-harnessed chance variations. It is the *harnessing* of stochasticity in guided responses to environmental challenges that achieves what blind chance alone could not possibly do. Ellis (2016: 163–168) refers to such processes as Adaptive Selection of Outcomes. Hoffmann refers to it as the process of extracting order from chaos in his book *Life's Ratchet* (Hoffmann 2012). He correctly identifies many of the molecular mechanisms that enable this extraction to occur, and concludes that

> reductionism is essential if we want to understand life. Without it, scientists would have long ago stopped looking at smaller and smaller scales and would have missed the marvels of molecular machinery. At the same time, molecular machines don't explain everything. Scientists must still answer the question of how these machines interact. The ultimate goal is always to explain the totality of life's processes….. reductionism and holism are two sides of the same coin. (Hoffmann 2012, 238)

Bronfman et al. (2016) refer to

> hierarchical predictive coding theory, according to which one of the organism's biggest challenges is to infer the (hidden) world-causes that give rise to the (observable) sensory signals the animal receives.

This point is relevant to ours about the anticipatory behaviour of organisms (Sect. 4(f)).

4.5 Example of Guided Reaction to New Environmental Stress: The Immune System

The immune system achieves such guided responses throughout the life of an organism. The mutation rate in the variable part of the genome that forms the template for an immunoglobulin protein can be accelerated by many orders of magnitude in response to a new antigen challenge. So far as is known, these mutations occur stochastically, and what is modified is the speed at which they occur. However, the location in the genome is certainly not a matter of chance. The functionality, in this case, lies precisely in targeting the relevant part of the genome. The arrival of the antigen itself activates the hypermutation

process, and the binding to a successful antibody triggers the proliferation of those cells that make it. Thus, the system targets the specific antigen.

What this process achieves is that all the other sequences in the DNA array forming a template for the immunoglobulin protein are held sufficiently constant to maintain functionality. Even more remarkably, all the functionality in the rest of the genome is also maintained. Considering the vast size of the entire genome, this is pin-point targeting requiring highly specific feedback processes to be successful.

By holding correct parts of the immunoglobulin sequence constant, the system finely tunes the rapid mutation to only *a tiny* part of the entire genome. Such tuning is one way in which organisms can dynamically respond to environmental change (Noble 2018). Another way is that their biological networks buffer the organism from the majority of molecular changes at the genetic level. The robustness of the networks acts like a cloud overlying the DNA (Noble and Noble 2017) so that, under favourable environmental conditions, as much as 80% of the genome changes have negligible effects (Hillenmeyer et al. 2008).

The targeted process in the immune system has been known and intensely studied for many years (Odegard and Schatz 2006; Shapiro 2011). The molecular action has been identified as abolishing or reducing the DNA copying error-correction process (Saribasak and Gearhart 2012). So, how did many people not realise that it is a physiologically guided process? The answer is that the guidance does not lie at the genome level. At the genome level, the process appears blind. It depends on purely stochastic mutation. The functionality guiding or targetting the process lies in the system as a whole. The system *harnesses* the stochasticity. The immune system is the paradigm example of this harnessing.

The system includes: (a) sensing the environmental challenge, i.e. the antigen invasion, (b) transmitting this signal to the nuclei of immune system cells to trigger hyper-mutation in just a tiny fraction of the genome, (c) then sensing of the correctness or otherwise of the outcome, followed by the "reproduce or die" signal: cells that do not produce an antibody to the antigen do not reproduce. At this stage, natural selection occurs amongst the population of immune system cells. Thus, this is a complete, finely-tuned physiological feedback and guided search system rapidly generating an acquired characteristic in response to an environmental challenge, and inherited through the surviving population of cells. By all the usual criteria this is a teleological, i.e. goal-directed, process.

The process does not have to be perfect. Not all keys have to be a perfect fit to open a lock. The system feels its way forward, harnessing stochasticity to create novelty while using targeted preservation of what already works. The targeted preservation is what gives the system its purpose: to maintain its integrity. It uses stochasticity to change what it must change, precisely because that is the part that does not work or does not work well enough.

It is crucial moreover to see that the goal, the directionality, exists *within* the organisms and their populations. The goals of organisms and populations of organisms have developed during the evolutionary process. They are harnessing the capacity to change to meet environmental or psycho-social challenges.

4.6 Harnessing Stochasticity Within the Nervous System

In a previous article on similar processes in the nervous system (Noble and Noble 2018), we represented the process as follows, where we divide what happens into steps only for clarity. There are significant analogies with, but also substantive differences from, processes in the immune system, which harnesses stochasticity to create antibodies to new challenges.

1. Influences from the environment and the organism's history may present a particular problem to the organism—for example, gaining access to a food supply or defending itself from a potential predator, or building a nest. We conjecture that such a problem can be viewed as a puzzle for which the organism needs a solution. The challenge facing the organism is in finding an answer to the puzzle. The puzzle is complex and multidimensional in space and time.

Behaviour is an ongoing dynamic process and is iterative and reiterative in solution and execution. Nor is the puzzle itself static, and the response is equally fluid or elastic, requiring recalculation and reassessment in finding a solution, for example in a predator stalking its prey. Once we engage a solution, more decisions are needed in addition to the original one. Problems do not present themselves in isolation, although we might dissect them as if they did. Thus, decisions of some kind are necessary even with routine outputs. The decision process is not a linear sequence, and the processing of action is continually assessed and adjusted; in a predator–prey situation where each party is continuously assessing, or anticipating the intent of the other.

2. Part of the process of choice is that organisms consider experience and memory of what may work as a solution. We hypothesise that the harnessing of stochasticity occurs continuously. Thus the organism is continually activating stochastic processes within and without itself so generating ongoing creative solutions. The stochastic options available may be extensive and at various levels (Atmanspacher and Rotter 2011; Brembs and Heisenberg 2018; Burns 1968; Hille 1992; Tchaptchet et al. 2015).

For example, the solution for a chimpanzee in getting termites would be to push a stick into the nest to get them. Both the use of the stick and choosing to hone it or modify it to better effect are biologically and socially creative and rational solutions in action. The behaviour can be understood both in terms of the biology and the situational logic. If we see a chimpanzee honing a stick, we (and other chimpanzees) might reasonably anticipate that it is going to use it to get termites.

Where are these stochastic processes located in the case of the nervous system? The answer is at all levels. Stochasticity is ubiquitous in biological systems.

At the cellular level, all life depends on a mostly stochastic process in generating a membrane potential. The generation of that potential depends on the concentration gradients of positive and negative ions driving flux across the membrane.

Control of these random fluxes is by changes in membrane permeability—opening and closing of specific ion channels. By opening ion channels in the membrane, nerve cells harness this stochasticity to generate discrete electrical signals, action potentials. Thus, neural processes are extensively stochastic—at all functional levels, from the opening and closing of ion channels via action potential generation, spontaneously or through synaptic transmission in neuronal networks, up to cognitive functions including decision making. All involve an ebb and flow of inhibitory and excitatory processes in complex networks of neurones. This balance is so even for the processing of first-order sensory inputs involving both localised and descending modulation of sensory processing in anticipation of a stimulus (Noble and Short 1989; Noble and Riddell 1989). To quote from the second of those papers:

> Nevertheless, it is clear from the present study that the PSDC [postsynaptic dorsal column neurones] system is amongst those somaesthetic pathways under the influ-

> ence of descending mechanisms which, whatever their precise physiological function, have the potential to modify and regulate the transmission of cutaneous sensory messages. (Noble and Riddell 1989, 181).

Thus, stochasticity is harnessed in the functional organisation of living systems composed of a manifold of nonlinear feedback loops that often are adjusted to operate in the neighbourhood of bifurcations where what happens can significantly depend on random influences, e.g. whether an ion channel is opened or remains closed or whether action potentials arise or not.

3. Potential solutions using these stochastic processes are continuously assessed as solutions to the puzzle. While a particular problem might be present at a given moment, the processes involved are in continuous engagement in anticipation of such problems forming part of perception and awareness.
4. The capacity to abstract the problem and potential solution may depend on the complexity of the organisation in a given species. Thus, in humans, we can envisage the potential solution and rehearse it before the behavioural output. Other species may also do so.

This process helps to explain the apparent paradox that we referred to earlier regarding the predictability or otherwise of what we call a free choice, and it ensures that, in retrospect, the choice may be what in the case of humans we call rational. There may be a greater or lesser degree of conformity to possible reasons why it was made. So the choice is both rational and creative.

We can now answer the earlier question in Sect. 4(d): how physiological changes can *at the same time* have identifiable molecular biological processes by which they occur and also correspond to rational choice at a macro-level. By harnessing stochasticity, lower-level attempts at a solution might occur, some of which correspond to an appropriate rational choice. The analogy with the immune system is now apparent in the harnessing of stochasticity, but the nervous system processes are different in many other respects.

4.7 Organisms as Open Systems

Organisms are open systems at both micro- and macro-levels. Thus, there are no hard boundaries to causation between levels; or as Capra and Luisi (2014) express it:

> In nature, there is no ‘above’ or ‘below’, and there are no hierarchies. There are only networks nesting within other networks.

‘Nesting’ is the key. Each level of function meshes with and so interacts with other levels (Noble et al. 2019). Molecular networks nest within cells, which in turn nest within cellular (tissue) networks, within organ networks, and whole organism networks, then within their social networks, and also in interaction with other species (ecology). This meshwork is what we mean by open systems—while there may be causal distance, there is no causal isolation. Thus, boundaries are not merely structural; they are also functional. Biological systems are not closed mechanical units. What happens at a social level of organisation influences what happens at the level of organs, tissues, cells and ultimately, molecules. Perceptions of the world, and each other, are influenced by culture and social interaction; it influences the development of our brains and our thoughts. Spinning the stochastic wheel occurs at all levels, allowing biological systems to be creative.

Where we differ from Capra and Luisi is on the question of hierarchy. In systems with agency, there is a hierarchy of sorts. The decisions of the agent, the organism, influence function at other levels, even as function at other levels forms the integral being of the agent. The level at which agency can exist constrains all the other levels. However, an agent depends on the capacity to act, which in turn is constrained by functionality. In that sense, there is no privileged level of causality; all levels engage in the process of agency, of making decisions and acting on them. What is the agent, but the organism, or the social entity? Being aware or conscious of our agency is also a functional level. It is not merely an 'illusion' but a potent level of organisation. Furthermore, the expression of our thoughts, our ideas, framing them in language and social interaction is a functional boundary.

This insight into the nature of open systems derives from a mathematical approach to the interactions across boundaries between different organisational levels, formulated as the principle of biological relativity (Noble 2012). It is important to note that the distinction between dynamics and constraints is evident in differential equation models but not in "static" algebraic models often used in evolutionary biology, for example in models of kin selection. When challenged by Stephen Rose on whether his work on algebraic models of kin selection and altruism (Smith and John 1964) could have political, philosophical and economic consequences, the evolutionary biologist John Maynard Smith famously quipped "what would he have me do? Fiddle the algebra?" (Smith 1979). Significantly, had Maynard Smith been using dynamic differential equation models he might have seen that constraints from macro-levels require no "fiddling" of the dynamic equations. Static models leave little room for the social dynamics of altruism and concepts of good and bad behaviour or thought, or individual agency. Static mathematics describes a system requiring stasis or equilibrium, or 'maintaining genes in a gene pool' and becomes the overriding goal of the system. In that case, the mathematics governs us rather than the other way round. However, the form of mathematics is a choice we have made. Evolutionary biology and its contributions to the debates on agency might look very different today had it not been so profoundly influenced by static (steady-state) mathematical models. So also would the philosophical debates on agency and free action. Organisms, as open systems, are never in a static equilibrium state.

5 Relevant Previous Physiological Work

Our proposed reconciliation between micro- and macro-level accounts requires that nervous systems are capable of generating different possible courses of actions of which the organism may be conscious. This capacity is what empirical studies of conscious awareness propose. Crick and Koch (2003; Koch 2012), for example, specifically talk of the processes of attention and neural selection processes in their work on the physiology of consciousness. We dissent from their reductionist approach to consciousness and in particular from the tendency in reductionist accounts to create a form of mechanistic dualism by reifying a specific bit of the process, a 'seat of consciousness' somewhere in the system (see Noble et al. 2014). However, the emphasis on a process of selection is correct, as are similar approaches found elsewhere (Dehaene and Changeux 2005; 2011).

The closest example of this emphasis on selection is in the work of Gerald Edelman (1978; Edelman et al. 2011) who as long ago as 1987 proposed the theory of neuronal group selection. And even earlier, Changeux and Danchin (1976) proposed the idea of selective stabilization of developing synapses.

Furthermore, the relevant neural selection processes are extensively reviewed and extended by Ginsburg and Jablonka (2019). Chapter 3 of their book reviews previous work, while chapter 4 specifically tackles the tension (bridging) issue. Their work is a significant break with the long-standing tradition in the philosophy of mind to keep conceptual and empirical questions separate, a view stated most forcefully by (Bennett and Hacker 2003). We agree with Bennett and Hacker on the inadequacies of reductionism and the problems created by ascribing psychological attributes to parts of an animal that are intelligible only when ascribed to the animal as a whole. Nevertheless, we dissent from a hard separation of conceptual and empirical questions and argue that in a full understanding of conscious agency, there are contiguous processes. A hard separation makes it difficult to find processes by which the one can influence the other. This separation leads to a strange parallel dualism, the illusion of conscious agency, or the concept of a bit of the brain, the liaison brain, that can strangely communicate between mind and matter, the 'self' as separate from the 'body' (Popper and Eccles 1977), or a bit of the brain that sees the 'seeing' of the other.

None of this means that we cannot or should not consider conceptual issues in the abstract. Indeed, that is the essence of thinking and a fundamental ingredient of perception. Organisms, and certainly humans, can create and solve abstract problems. This paper is a testament to that. Animals can solve problems by both observation and imagination and from learning, as individuals and as social groups.

Nevertheless, what we measure and observe is, in no small degree, dependent on a conceptual framework. In biological processing, there is an interdependency between the conceptual and the empirical, although the relationship is asymmetrical. We can conceptualise with some degree of freedom from observation, but it is challenging to observe events without a conceptual framework. Arguably, thinking and anticipation are ingredients of perception.

This relationship between the conceptual and empirical brings us back to the central question addressed in this paper. If determinate molecular and other physical forces wholly and always determine the actions of organisms, then how could a rationalist explanation possibly be correct? The answer is that they do not. Organisms harness them in processes that maintain their integrity. Organisms create reasons, and abstract thinking by harnessing processes including stochasticity at the molecular, cellular, neural and social levels and are not abstract from them; they are the abstractors. By doing so, they anticipate the outcomes of their activities and that of others. They solve rational problems because they also create them. They are goal-directed. In this sense, they can generate and consider possible solutions to achieving their goals.

Even if we made what we consider to be an erroneous assumption that organisms were driven by genes to pass their genes on in a gene pool, they would have to overcome many problems in doing so. Yes, some solutions might be hard-wired, but many will require more flexible, creative solutions. As the chimpanzee learns, there may be more than one way to crack a nut (Luncz et al. 2012).

What we have shown in this article is that there is no difficulty from an empirical science viewpoint in envisaging how organisms could achieve conformity to rational actions through the process of harnessing stochasticity. Harnessing is a necessary causal process, as is evident also from its role in guiding evolution (Noble 2017; Noble and Noble 2017). Functional boundaries between organisational levels mean causation up and down are necessarily different (Noble et al. 2019), but they do not compete for primacy. They mesh together and are both enabling and creative. In setting boundaries, downward causation can be viewed more like a context, setting constraints, purpose and goals. It is then not too difficult to view reasons, ethics, laws and customs operating in

this way. They are socio-biological processes influencing predisposition states in the organism. Thinking that we need to solve how upward and downward causation 'compete' with each other is a mistake. They mesh. Reasons are not incidental or merely epiphenomenal; organisms create them as contextual logic. Thus, reasons form the contextual influences within which action occurs. As we will make clear in Sect. 7 (Can reasons and values be causes?), this does not mean or guarantee that the decision will fully conform to any particular rationalisation. It is the taking of the decision *in the context of seeking conformity to rationality* that can be influential.

Living organisms are therefore capable of at least partially reconciling logical (reasons) and physical processes. They do this by reconciling different forms of causality. Causation across the boundaries between different levels is necessarily various (Noble et al. 2019). With the macro-level contextual constraints of events at the micro-level, there is no causal closure. The process is ongoing and reiterative with an ebb and flow of predisposition states influencing decisions by the organism. This openness, or lack of causal closure, is a point that seems to be missing in the accounts of philosophers like Jaegwon Kim (2000).

6 Relevant Recent Philosophical Work

The issue of free will has been a major one in philosophy for centuries (Kenny 2006, Ch. 7, 212–245; 2007, Ch. 8, 192–219) and the debate is still very much alive, as is clear from our earlier references to Jaegwon Kim's work.

In the introduction, we noted the relevant work of Karl Popper. We can now explain why we referred to close parallels with the proposition in this paper.

In his 1986 lecture to The Royal Society, Popper distinguished between what he called "active" and passive" Darwinism. By "passive" he meant a theory of evolution that attributes all change to natural selection, which is a passive filter for degrees of fitness for survival. This idea is also a central tenet of neo-Darwinism (The Modern Synthesis). In contrast, by "active" he meant the directionality that organisms create as agents, and which was first identified by Charles Darwin through his work on sexual selection—through an active choice of mates for reproduction (Darwin 1871).

The propositions in the current article contribute to understanding why Darwin was right to distinguish between natural selection and what he called artificial selection—the *active, or purposeful* selection of varieties of species by human agency. The distinction is valid of course only if we believe that agents exist and that we can identify them. If active agency exists, then organisms are not passive prisoners of events. They can act to shape them.

Amongst modern philosophers, our proposals on the processes by which organisms make choices are closest to those of Christian List (List and Menzies 2009; List 2014, 2019).

List bases his arguments for free will on three propositions:

1. The fact that, for any macro-level situation, there will be innumerable micro-level states that correspond to the same macro-level state (List 2019, Fig. 2, 94).
2. From this, he shows that the high-level state can branch (in making a choice) while any given lower-level state does not and (in a deterministic world) cannot branch.

3. From these statements, he shows that all three of his criteria for free-will are then satisfied. These are intentional agency, the existence of alternative possibilities, and causal efficacy.

We believe that our work provides strong support for (1).

Nevertheless, we are not convinced that, on its own, (2) would answer most micro-level determinists since they could still argue that in any given case only one of the many possible lower level instantiations occurs. Which one occurs can be described by pure biochemistry. Thus, they would argue that since on any given occasion, only one of them occurs then the particular high-level state cannot branch either. One way of viewing our contribution to the debate is to show that it is not necessary to assume that only one lower-level state occurs in any given timeline. Thus events at the macro-level harness those at a micro-level in generating possible alternative actions. Our case is based precisely on the organism's ability to anticipate *many* options simultaneously. It is the subsequent selection of one or more of these that is the choice process. The organism is then able to apply an ongoing creative logic to the choices of action, which in turn continuously moulds or modifies micro-level processes. Thus, the macro-level choices influence the biochemistry.

In terms of List's timeline diagrams for macro- and micro-levels (List 2019, 94), we are therefore arguing that *the organism anticipates many of the lower-level timelines* (whether consciously or not) when making choices. Rationality can enter into that choice, but that does not necessarily mean that the organism chooses the most rational that might be possible. Organisms do not behave as rational calculators. However, they are influenced by rational considerations, particularly in choices that anticipate the actions of other organisms or the outcome of their actions. These influences have causal efficacy because they establish an ebb and flow of predisposition states. Thus, anticipation is a crucial ingredient of all perception and all choices. Senses are not merely passive receivers. Our nervous systems anticipate change. Many illusions may result from tricks on the brain, resulting from this anticipation. We might better consider thought as integral to our senses rather than being nebulously superimposed upon them. There is then less difficulty in understanding how our thoughts can influence our behaviour. Thinking is a continuous process in the ebbing and flowing of the biological anticipatory process. Humans can do something remarkable with this, which is to use language better, convey abstract ideas to others and to express it culturally in art and literature, and this can alter our perceptions transgenerationally and the choices we might make. Ideas generated socially can alter our dispositional states. We have agency at both the organism and social level. We are free agents to the extent that we can run these processes freely to generate ideas on which we may act, although our capacity to do so is constrained. Learning and training can enhance our capacities.

By adding the harnessing of stochasticity as an active ingredient to List's set of propositions, it enables many micro-level states to be available to an organism in making choices. The argument for "free will", or as we would have it "free agency", becomes more tenable, and all the consequences of List's third proposition then follow.

7 Can Reasons and Values Be Causes?

The question of whether reasons, values, and other social/ethical factors can be regarded as causes is a contested one. Philosophers like, for example, Anthony Kenny argue that reasons cannot be causes: "reasons are not causes, and the relationship between reason and

action is quite different from that between cause and effect." (Kenny 1992, 143). "This is because rules of practical inference are defeasible, whereas causal laws are not." "That is to say a conclusion which may be a reasonable one from a given set of premises may cease to be a reasonable one when further premises are added" (Kenny 1992, 145).

We agree that the rules of practical inference are defeasible. That is a necessary logical fact and not in question. Moreover, we have insisted throughout this article that actions that can be defended as reasonable are usually so defended *in retrospect*. As we wrote earlier "Organisms do not behave as rational calculators." We have also noted that, in everyday social interactions, and legal contexts, we do sometimes change the reasons we may acknowledge as to why we acted in the way we did. That also is not in question.

A choice of action may be its match to the *plausibility* of conforming to a rational construct. It is not caused in the way that micro-level events are caused by the physical dynamics of how material bodies interact, nor that it should follow the rules of practical inference. They are causal in the sense that they establish or alter states of disposition or be understood to work. Thus, the chimpanzee who has learned to crack nuts with stones will find a suitably shaped stone, which it may keep for future use if it is a useful utility for achieving its objective. We may consider the properties of a 'good' stone, such as its weight, shape and size, and how best to use it to good effect. Some chimpanzees will try different approaches; abandoning some stones on finding a "better."

What the process of choice through the harnessing of stochasticity requires is that organisms should be influenced (constrained) by considerations of rationality. Such constraint need not even be specific, or even justified, in particular cases. Furthermore, we explicitly include the fact that, often enough, in critically urgent circumstances, organisms simply do not have the time to arrive at the best response. Many actions in such cases are "on impulse". Organisms are also free to do that and will do so in crucial fight or flight situations. Being rational does not mean slavishly following a set of logic, just as we may not follow what we see or hear, touch or smell.

In these respects, the nature of the influence involved is comparable to that used by the immune system. A precisely exact fit of an antibody to the invading antigen is not necessary. What matters is that the fit is 'good enough'. If the key fits the lock, it does not have to be the perfect key.

Thus, we accept Kenny's point that reasons do not determine actions in the way that the dynamics of molecular interactions perform their role in causation. The process of harnessing stochasticity will always be, even at best, a 'good enough' response, not a logically calculated one. It involves a continuous appraisal and reappraisal of the solution and the problem.

As a specific example, consider the writing of this article. It took many iterations to arrive at the text. In highly deliberative actions, the outcome is a continuous striving towards a goal, in this case, to be as clear and logical as we can be. The logic we strove for was always defeasible. However, the striving towards that goal counts as a cause under the heading of "any factors whose alteration would result in changes in behaviour" (see Sect. 3. Causation). No specific (and defeasible) logical statement acted as *the* cause, but the general context of expressing our ideas logically was causal.

8 Future Work

Our article provides leads to future research, scientific and philosophical.

a. *Immune System* The use we have made of the immune system as a model for the process of harnessing stochasticity is firm at the molecular level, but is far from being adequately investigated at the level of the overall control processes. How exactly do cells receiving signals at their surface communicate with such pinpoint accuracy to their genomes? Cellular communication mechanisms from surface to nucleus are now known in other physiological contexts (Ma et al. 2014; Kar et al. 2016), but we need to know much more about the integrative physiology of immune responses. This further understanding could provide valuable insights into other integrative biological processes employing stochasticity functionally.
b. *Nervous System* Further work on harnessing stochasticity in labile neural networks would provide insight into the processes of creative decision making. For example, this might involve control of synaptic function in labile switching of neuronal circuits underlying ideation.
c. *Philosophy* As our Sect. 6 shows, there are already ways in which the idea of harnessing stochasticity could link up with recent work on the philosophy of free agency. We anticipate that our article may open the way to future cooperation between scientists and philosophers in this field.

9 Conclusions

The conditions for the existence of what we call agency in this article include the following three scientific factors:

9.1 The Openness of Organisms

They are not and cannot be closed systems. All organisation levels nest within others. Once that is accepted, then a constraint of micro-level processes can occur through determining the initial and boundary conditions. Those constraints do not compete with the dynamics of the micro-level. The two forms of causation mesh together in achieving goal-directed behaviour.

9.2 The Existence of Functional Boundaries, Both Between Different Levels of Organisation and Between Organisms and Their Environments

Boundaries are the locations at which downward causes (constraint of micro-level elements by the macro-level organisation) exert their effects. The social boundary is, in some respects, similar to how a human agent interacts with a computer. On its own, the computer is a determinate mechanical system. Through its interaction across the human–machine boundary, it becomes constrained to instantiate the logic imposed by the human agency in writing programs and data.

9.3 The Harnessing of Stochasticity, at Various Levels, Molecular, Cellular, Systems (e.g. Nervous System) and Social

Thus, chance is not simply experienced; but used continuously in creatively guided behaviour. Merely incorporating random number generators into an otherwise determinate AI

system does not capture this process. Harnessing of stochasticity can be either unconscious (immune system) or conscious (organisms with neural processes), even though the precise processes are different.

These conditions are interdependent. Openness automatically leads to the boundary between the organism and its environment. Harnessing stochasticity involves constraint both of micro-level chance events and socially-generated macro-level chance occurrences.

It will also be clear from this article that we regard organisms that have agency as actively anticipating their physical and social environments. Perception is an activity, not just a passive 'camera' observing the world.

But is the agency, according to these interpretations, really free? Yes, it is free of purely micro-level determination, thus giving it a degree of freedom, but it is still "determined" by the outcome of the meshing of macro-level constraints, including socially-determined constraints, with micro-level dynamics.

We, as humans, thinking we have what we call free will, would have no difficulty with someone saying of our reasons that they are what could be said to have determined what we did. That is how we answer questions such as "why did you do that?" So, in that sense, organisms with free will *are* determined, but by very complex logical, moral and other social factors meshing with the physical limitations on what organisms can do. Being open systems allows degrees of freedom in choice. Purpose and choice are determinate. It does not require a 'will' strangely posited such that it is 'free' from the material constraints. It is organisationally creative, harnessing the uncertainty of those constraints. Thus, we wrote this paper with the intention that the reader considers this point. It might bring about dispositional changes in the reader's brain that alter perception. Our rational actions are, in this sense, determined, but we determined them as purposeful agents.

In conclusion, the agency of organisms is powerful, not an illusion, and it is possible to reconcile the micro- and macro-level accounts of behaviour. Agency also endows organisms with directionality, i.e. intentional, forward-looking action. Agency gives an arrow to the interpretation of behaviour in much the same way as thermodynamics gives an arrow to time. Both do so through necessary constraints.

Acknowledgements We acknowledge valuable criticism of versions of this article from Anthony Kenny, George Ellis, Simona Ginsburg, and Eva Jablonka. We also thank the two anonymous referees who made valuable suggestions for revisions of this paper. Of course the errors that remain are our responsibility. Denis Noble is a member of the Project "Foundations of Value and Values" at the Said Business School of the University of Oxford and acknowledges valuable input from the members of that group. We declare no conflicts of interest.

References

Atmanspacher, H., & Rotter, S. (2011). On determinacy or its absence in the brain. In R. Swinburne (Ed.), *Freewill and modern science*. London: British Academy.

Bathgate, K. E., Bagley, J. R., Jo, E., Talmadge, R. J., Tobias, I. S., Brown, L. E., et al. (2018). Muscle health and performance in monozygotic twins with 30 years of discordant exercise habits. *European Journal of Applied Physiology*. https://doi.org/10.1007/s00421-018-3943-7.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell Publishing.

Boyle, E. A. L., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell, 169,* 1177–1186.

Brembs, B., & Heisenberg, M. (2018). Der Zufall als kreatives Element in Gehirn und Verhalten. In U. Herkenrath (Ed.), *Zufall in der belebten Natur*. Hennef: Verlag Roman Kovar.

Bronfman, Z. Z., Ginsburg, S., & Jablonka, E. (2016). The transition to minimal consciousness through the evolution of associative learning. *Frontiers in Psychology, 7,* 1954.

Burns, B. D. (1968). *The uncertain nervous system*. London: Arnold.

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition, 20,* 755–769.

Capra, F., & Luisi, P. L. (2014). *The systems view of life: A unifying vision*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511895555.

Changeux, J.-P., & Danchin, E. (1976). Selective stabilization of developing synapses as a machanism for the specification of neuronal networks. *Nature, 264,* 705–712.

Coyne, J. A. (2014). What scientific idea is ready for retirement? https://www.edge.org/response-detail/25381. Retrieved 12 October 2020.

Crick, F. H. C., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience, 6,* 119–126.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London: John Murray.

Dehaene, S., & Changeux, J.-P. (2005). Ongoing spontaneous activity controls access to consciousness: A neuronal model for inattentional blindness. *PLoS Biology, 3,* e141.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70,* 200–227.

Del Santo, F., & Gisin, N. (2019). Physics without Determinism: Alternative interpretations of classical physics. *Physical Review A, 100,* 062107.

Edelman, G. M. (1978). *Neural Darwinism: the theory of neuronal group selection*. New York: Basic Books.

Edelman, G. M., Gally, J. A., & Baars, B. J. (2011). Biology of consciousness. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2011.00004.

Ellis, G. F. R. (2016). *How can physics underlie the mind?*. Berlin: Springer.

Epstein, D. (2014). *The sports gene: Talent, practice and truth about success*. Hyderabad: Yellow Jersey.

Foster, Russell, & Kreitzman, Leon. (2004). *Rhythms of life*. London: Profile Books.

Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul*. Cambridge: MIT Press.

Graham, K. E., Furuichi, T., & Byrne, R. W. (2017). The gestural repertoire of the wild bonobo (*Pan paniscus*): A mutually understood communication system. *Animal Cognition, 20,* 171–177.

Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., & Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLOS Biology*. https://doi.org/10.1371/journal.pbio.2004825.

Heisenberg, M. (2009). Is free will an illusion? *Nature, 459,* 164–165.

Hille, B. (1992). *Ionic channels of excitable membranes*. Sunderland, Mass: Sinauer Associates Inc.

Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., et al. (2008). The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science, 320,* 362–365.

Hobsiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology, 24,* 1–5.

Hoffmann, P. M. (2012). *Life's Ratchet: How molecular machines extract order from chaos*. New York: Basic Books.

Kar, P., Mirams, G. R., Christian, H. C., & Parekh, A. B. (2016). Control of NFAT isoform activation and NFAT-dependent gene expression through two coincident and spatially segregated intracellular Ca^{2+} signals. *Molecular Cell, 64,* 746–759.

Kenny, A. J. P. (1992). *The metaphysics of mind*. Oxford: Oxford University Press.

Kenny, A. J. P. (2006). *A new history of western philosophy. III The rise of modern philosophy*. Oxford: Oxford University Press.

Kenny, A. J. P. (2007). *A new history of western philosophy. IV philosophy in the modern world*. Oxford: Oxford University Press.

Kim, Jaegwon. (2000). *Mind in a physical world*. Cambridge, Mass: MIT Press.

Koch, C. (2012). *The quest for consciousness—confessions of a romantic reductionist*. Cambridge: MIT Press.

Lim, D. (2008). Review of Murphy N & Brown, W.S. 2007. Did My neurons make me do it? Philosophical and neurobiological perspectives on moral responsibility and free will. *Zygon, 43,* 748–753.

List, C. (2014). Free will, determinism, and the possibility of doing otherwise. *Nous, 48,* 156–178.
List, C. (2019). *Why free will is real*. Harvard: Cambridge Mass.
List, C., & Menzies, P. (2009). Non-reductive physicalism and the limits of exclusion. *Journal of Philosophy, 106,* 475–502.
Lucas, J.R. 1970. *The freedom of the will* (Oxford University Press).
Luncz, L. V., Mundry, R., & Boesch, C. (2012). Evidence for cultural differences between neighboring chimpanzee communities. *Current Anthropology, 22,* 922–926.
Ma, H., Groth, R. D., Cohen, S. M., Emery, J. F., Li, B., Hoedt, E., et al. (2014). γCaMKII shuttles Ca^{2+}/CaM to the nucleus to trigger CREB phosphorylation and gene expression. *Cell, 159,* 281–294.
Midgley, M. (2014). *Are you an illusion?*. London: Routledge.
Murphy, N., & Brown, W. S. (2007). *Did my neurons make me do it? Philosophical and neurobiological perspectives on moral responsibility and free will*. Oxford: Oxford University Press.
Niemann, H.-J. (2014). *Karl Popper and the two new secrets of life*. Tuebingen: Mohr Siebeck.
Noble, D. (2012). A theory of biological relativity: No privileged level of causation. *Interface Focus, 2,* 55–64.
Noble, D. (2016). *Dance to the tune of life. Biological relativity*. Cambridge: Cambridge University Press.
Noble, D. (2017). Evolution viewed from physics, physiology and medicine. *Interface Focus, 7,* 20160159. https://doi.org/10.1098/rsfs.2016.0159.
Noble, D. (2018). Central dogma or central debate? *Physiology*. https://doi.org/10.1152/physiol.00017.2018.
Noble, R., & Noble, D. (2017). Was the watchmaker blind? Or was she one-eyed? *Biology, 6,* 47. https://doi.org/10.3390/biology6040047.
Noble, R., & Noble, D. (2018). 'Harnessing stochasticity. *How Organisms Make Choices', Chaos, 28,* 106309. https://doi.org/10.1063/1.5039668.
Noble, R., & Noble, D. (2019). Could artificial intelligence (AI) become a responsible agent: Atificial agency (AA)? *RUSI Journal, 164*, 130–135.
Noble, D., Noble, R., & Schwaber, J. (2014). What is it to be conscious? In J. Smythies, V. S. Ramachandran, & L. Edelstein (Eds.), *The claustrum*. Cambridge: Academic Press.
Noble, R., & Riddell, J. S. (1989). Descending influences on the cutaneous receptive fields of postsynaptic dorsal column neurones in the cat. *Journal of Physiology, 408,* 167–183.
Noble, R., & Short, A. D. (1989). Spatial spread of in-field afferent inhibition in the cat's spinocervical tract. *Journal of Physiology, 413,* 107–118.
Noble, R., Tasaki, K., Noble, P. J., & Noble, D. (2019). Biological Relativity requires circular causality but not symmetry of causation: So, where, what and when are the boundaries? *Frontiers in Physiology, 10,* 827.
Odegard, V. H., & Schatz, D. G. (2006). Targeting of somatic hypermutation. *Nature Reviews Immunology, 8,* 573–583.
Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. Bristol: Allen Lane.
Popper, K. (1945). *The open society and its enemies*. London: Routledge.
Popper, K. R. (1972). *Objective knowledge. An evolutionary approach*. Oxford: Oxford University Press.
Popper, K. R. (1973). Indeterminism is not enough. *Encounter, 40,* 20–26.
Popper, K. R., & Eccles, J. C. (1977). *The self and its brain*. New York: Springer.
Sakai, T., Tamura, T., Kitamoto, T., & Kidokoro, Y. (2004). A clock gene, period, plays a key role in long-term memory formation in Drosophila. *Proceedings of the National Academy of Sciences, 101,* 16058–16063.
Saribasak, H., & Gearhart, P. (2012). Does DNA repair occur during so0matic hypermutation? *Seminars in Immunology, 24,* 287–292. https://doi.org/10.1016/j.smim.2012.05.002.
Shapiro, James A. (2011). *Evolution: A view from the 21st century*. Upper Saddle River, NJ: Pearson Education Inc.
Smith, J. M. (1979). Letter. *The New Scientist*, 14 June 1979, 213.
Smith, J. M. (1998). The units of selection. In *The limits of reductionism in science*. Wiley.
Smith, Maynard, & John, (1964). Group and kin selection. *Nature, 201,* 1145–1147.
Tchaptchet, A., Jin, W., & Braun, H. A. (2015). Diversity and noise in neurodynamics across different functional levels. In R. Wang & X. Pan (Eds.), *Advances in cognitive neurodynamics*. Singapore: Springer.
Vohs, K. D., & Schooler, J. W. (2008). The value of believing in freewill, encouraging a belief in determinism increases cheating. *Psychological Science, 19,* 49–54.
Whiten, A. (2017). A second inheritance system: The extension of biology through culture. *Interface Focus*. https://doi.org/10.1098/rsfs.2016.0142.

To appear in Karl Popper: His Philosophy and Science. Springer-Nature 2021.

Rehabilitation of Karl Popper's ideas on evolutionary biology and the nature of biological science.

Denis Noble
Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford OX1 3PT, UK

and

Raymond Noble
Institute for Women's Health, University College London, London WC1E 6AU, UK

SUMMARY

Karl Popper's ideas on evolution in his 1986 Medawar lecture were remarkably close to Charles Darwin's original distinction between artificial and natural selection, but at odds with the Modern Synthesis in giving an active role to organisms in the process of evolution. His ideas were also compatible with recent work showing the role of the harnessing of stochasticity in enabling this active agency. He also argued against the reduction of biology to chemistry. In this article we show:

(1) That these ideas were also compatible with, and flowed from, his work on the Open Society.

(2) That organisms are necessarily open systems.

(3) That a multi-level analysis of organisms shows that there cannot be causal closure at the micro-level, as proposed by philosophers like Jaegwon Kim.

(4) While Popper's and Darwin's distinction between active and passive forms of evolution is valid, there is, in the long term, no purely passive form of selection. Organisms also create their environments.

Keywords: Karl Popper, Charles Darwin, evolution, agency, natural selection, social selection, sexual selection, open systems.

Introduction

In this article we discuss what we consider to be significant contributions Karl Popper made to evolutionary biology and to the fundamental nature of biology as a science. We explain why his ideas were controversial and why they were forgotten and neglected until recently. The key argument is on how the agency of organisms mediate change, which we develop in five parts: 1. Karl Popper and Evolutionary Biology; 2. Can Biology be reduced to Chemistry? 3. How organisms can be agents in development and evolution; 4. Organisms as open systems; 5. Conclusions: are active and passive Darwinisms entirely separate?

1. Karl Popper and Evolutionary Biology

In 1986, Popper gave a lecture to The Royal Society in London in which he laid out his "New Interpretation of Darwinism." The history of that lecture is the subject of a recent book by Hans-Joachim Niemann (2014), and a relevant article by Eva Jablonka (2017).

1.2 Passive and active Darwinism

In his lecture, Popper distinguished between what he called "passive Darwinism" and "active Darwinism." As these terms are not used today, they need an explanation, particularly because Popper referred to both of them as versions of "Darwinism" and even regarded his ideas as a refutation of Darwinism. However, in this respect Popper was not correct; it was a refutation of *neo*-darwinism, and not of the ideas proposed by Darwin. As we will show, Darwin was the first to develop both passive and active forms of his theories of evolution. In this, Darwin predated Popper.

Popper's "passive Darwinism" is more or less identical with classical neo-Darwinism: the theory that random genetic variation and natural selection are entirely sufficient (*allmacht* in Weismann's (1893) words) to explain evolution. "Passive" refers to the fact that the standard neo-Darwinist Modern Synthesis does not regard organisms as active agents in their evolution. Over the generations, they experience random variations in their genetic inheritance, but they are not viewed as *using* that randomness to generate functional variation that might be inherited. The organisms play no active part in evolution other than to pass these genetic variations on to the next generation. In this view, organisms are passive vehicles in gene transmission. In recent articles, we have outlined the opposing view that organisms themselves can, at least partially, guide evolution through the active choices that they make, and that this is achieved through the harnessing of stochasticity (Noble, 2017, Noble & Noble, 2017, 2018). The existence of such harnessing is an empirically testable process. The feedback control mechanisms determining the speed and location of genetic mutations and genome rearrangement are open to experimental investigation as dynamic biological processes. Furthermore, through such guidance (harnessing) the genetic variation is no longer random, but adaptive, even though randomness is used in the hypermutation and other cellular network processes involved. Those processes are also experimentally testable. Thus, evolution is an extension of the adaptive processes (physiology) by which organisms maintain integrity in response to environmental change.

There are therefore strong parallels between our work and that of Karl Popper. Popper wrote:

"I shall attempt to turn the tables completely on passive Darwinism . . . I shall claim that the *only* creative element in evolution is the activity of living organisms." (Niemann, 2014, p. 119). "Active Darwinism" is therefore equivalent to the theory that organisms have agency and make choices, which is the main theme of our recent papers. Those choices include selecting niches (niche selection theory) and which other organisms they interact with (social, including sexual, selection), and more recently, the discovery of aversion to cheating behaviour in populations of dogs (Essler et al, 2017) and monkeys (Brosnan & De Waal, 2003).

Popper also regarded the "metaphor of 'natural selection'" as "a theory of error elimination" (Niemann, 2014, p. 120) rather than being creative of novelty itself. He saw it as a filter eliminating errors. On its own therefore, natural selection does not involve agency.

To understand this point, we should remember that Darwin contrasted natural selection with artificial selection, which is clearly dependent on choices made by organisms (the selective breeders). His 1859 book, *The Origin of Species*, begins with a chapter on "Variation under domestication". He noted that breeders of varieties of dogs, cats, fish, and plants were actively (consciously) choosing the characteristics of the varieties they had developed. They were doing so as *active* human (= artificial in this context) agents. In introducing the term 'natural selection' Darwin was, by metaphorical extension, attributing selection also to an essentially blind *passive* process, and specifically *in contrast to* active choice in organisms. This was his great achievement in his 1859 book *The Origin of Species*. His theory of natural selection, therefore closely corresponds to what Popper called passive Darwinism.

In his later book *The Descent of Man and Selection in Relation to Sex*, Darwin (1871) showed that what he attributed to active 'artificial' selection to humans also occurs as an evolutionary process in many animals, through the process of what he called sexual selection.[1] However, the co-discoverer of the concept of natural selection, Alfred Russel Wallace, disagreed with him, and subsumed sexual selection to natural selection. That was perhaps the first great philosophical error in the subsequent narrowing down of the theory of evolution by natural selection to eventually become the Modern Synthesis. It is an error that has continued to confuse modern biology to this day. Julian Huxley in his 1942 book *Evolution: The Modern Synthesis* did not even include sexual selection as a process driving evolution. Moreover, it is a common assumption by many supporters of the more dogmatic forms of the Modern Synthesis that organisms cannot be active agents. This aversion to agency explains why they go to great lengths to understand sexual selection purely in terms of natural selection. It also explains why they often also deny agency (active choice) to humans. We will deal with these issues in section 3.

When Darwin realised that sexual selection is more like artificial selection, he realised that it is clearly an *activity* of organisms partially determining their evolution. Darwin recognized this difference as empirical since he wrote "with respect to animals very low in the scale, I shall have to give some additional facts under sexual selection, shewing that their mental powers are higher than might have been expected." (Darwin, 1871, I, pp 35-36). Sexual selection is therefore a form of active Darwinism to use Popper's terminology. Specifically, Popper wrote "sexual selection is a refutation of natural selection." (Niemann 2014, p 128)

[1] Actually, Darwin (1859) had already identified sexual selection briefly in *The Origin of Species*, pp 101-104.

Darwin distinguished very clearly between the two and we believe Popper and Darwin would have agreed on this distinction.

Popper can therefore be seen to be following in the footsteps of the *original* position of Charles Darwin, in opposition to the neo-Darwinist view that everything in evolution is attributable to blind natural selection. Since Popper persisted in regarding his lecture as a "refutation of Darwinism" he may not have appreciated that Darwin had already made the distinction to which he was drawing attention, and that it was therefore neo-darwinism that was the real target of his lecture. Else, he recognised that Darwin's position had been distorted by what had become generally regarded as 'Darwinism'. The Modern Synthesis is not the position developed by Darwin. It presents a gene-centric determinism, where physiological adaptation in the organism cannot be passed on through the germ line – genes make proteins and function, but proteins and function cannot alter the genes. The genome is viewed as a 'blueprint' for development and function, where change of that blueprint is dependent on the two ingredients of blind chance and natural selection. It established a kind of mechanistic dualism with a privileged role for the genome in function, and in doing so it removed agency from the organism. It locked the gene away in a box, free from the choices made by organisms.

1.2.Role of indeterminacy

Popper saw that a complete determinism is incompatible with viewing organisms as agents making purposeful choices. Thus, he would have seen the significance of the role of harnessing stochasticity in creative responses to change and in making choices, which we have highlighted in our recent articles. However, he also recognised that indeterminacy, or blind chance, was not alone sufficient for an open, and thus creative system. It is not the unpredictability of events that creates agency. On the contrary, as we have argued, agency requires anticipation of change and the outcome of action. Yet, organisms harness stochasticity throughout biological function as the energy of creative change. Life constrains stochasticity, moulding it in function. It is the key ingredient of all physiology – from generating membrane potentials and synaptic function to releasing hormones, the beating of the heart, moving our limbs, and thinking.

In *The Open Universe: An Argument for Indeterminism*, Popper demonstrated that indeterminism is a necessary but not sufficient condition for emergence and openness (Niemann, 2014, p 70).

In the same exposition of Popper's ideas leading up to his Royal Society lecture, Niemann presents some other points that correspond well to the ideas of our work on agency. Summing up Popper, he concluded that "all life is problem solving. Acquiring new knowledge is always purposeful activity." (Niemann, 2014, p 90). Popper insisted that "in all cases the activity comes from outside of the DNA. The former 'centre of life' is rather a dead place." (Niemann, 2014, p 96). It is the cell that divides, not only the DNA (Niemann, 2014, p 98), and that it is "The cell . . . also managing the genome." (Niemann, 2014, p 101) This insight resembles that of Barbara McClintock, the discoverer of natural genetic engineering (Shapiro, 2011) in saying that "the genome is an organ of the cell." (McClintock 1984).

Thus, Popper realised that the genome is the prisoner of the cell and organism and not the other way round. He also pointed out that "influences (on action) [are] traceable in hindsight . . . we are unpredictable but not irrational" (Niemann, 2014, p 110). Popper therefore arrived at many of the points we have made in recent articles. In solving problems life can create solutions, else it cannot solve problems. The key problem is that of maintaining the integrity of life.

It would therefore be surprising if Popper had not also seen the obvious implication, which is that organisms harness stochasticity; otherwise, creative choice in behaviour would not be possible (Noble & Noble, 2018). Thus, we are grateful to Hans-Joachim Niemann for directing us to Popper sources preceding his Royal Society lecture where he does clearly draw this conclusion. Some of the relevant texts occur in his dialogue with John Eccles *The Self and Its Brain* (Popper & Eccles, 1977). Popper writes "New ideas [*in statu nascendi*] have a striking similarity to genetic mutations" and continues "describing 'the process with respect to new ideas and to free will decisions'(Popper & Eccles, 1977, p. 540) as randomly produced proposals followed by selection based on standards coming from the world" (Niemann, 2012, pp. S510–S546). Popper arriving at this conclusion is a logical outcome of his earlier (1973) conclusion that "indeterminism is not enough." (Popper, 1945, vol 2, p. 210, 1973).

Certainly, we agree that indeterminism is not enough for creative or purposeful agency. In standard evolutionary biology, following the Modern Synthesis, stochasticity generates random genetic variations. But in the standard theory this is not directed in a functional way. These chance variations may or may not confer any advantage on the organism, although those that do will more likely be retained in the 'gene pool'. In contrast, the direction of agency comes because organisms *harness* stochasticity in functional ways. Thus, the immune system creates hypermutation in highly targeted regions of the DNA sequence of immunoglobulin proteins. Under stress, bacteria also use hypermutation to resist antibiotics and to counter other forms of genetic loss.

Nevertheless, while Popper envisaged "the cell . . . also managing the genome," (Niemann, 2014, p. 101), he does not seem to have arrived at the details of the comparison with hypermutation in the immune system. This is not surprising since the discovery of some of the detailed molecular mechanisms of somatic hypermutation occurred in 1999 after his death in 1994 (Muramatsu, et al 1999; Li et al, 2014). There may also have been a puzzle regarding the molecular mechanism of hypermutation. Increasing the natural mutation rate by a factor of up to 10^6 (a million-fold increase!) must have seemed implausible. But this is no longer so astonishing since it is also roughly the order of magnitude difference between the natural mutation rate in DNA copying before and after repair by cellular editing mechanisms (Noble, 2018). Mismatch DNA repair is indeed suppressed during somatic hypermutation. Recent research has therefore shown that there is no difficulty in accounting for hypermutation rates of up to a million times normal. All that is required is to inhibit the error-correcting process in the relevant part of the genome to bring about targeted change. Thus, the stochasticity can be released from constraint in a targeted way. We have argued that it is this targeted process of mutation that gives direction, or agency, to the organism in response to environmental change.

2. Can Biology be reduced to Chemistry?

Why was Popper's Medawar lecture never published by The Royal Society journals? A possible explanation is that, following the lecture, Popper engaged in extended discussion with Max Perutz on the question whether Biology could be reduced to Chemistry (Niemann, 2014, pp 62-66). On this question Popper and Perutz were in complete disagreement.

Perutz, a molecular biologist who shared the 1962 Nobel Prize for Chemistry with John Kendrew for their studies of the structures of haemoglobin and myoglobin, presented the reductionist case that had gained hold in biological science. Popper disagreed, as one would expect from the earlier quote:

"in all cases the activity comes from outside of the DNA. The former 'centre of life' is rather a dead place."

Popper's statement is surely correct. DNA does nothing outside a cell. A cell is required for both copying and error correction[2] and is therefore the minimal form of a living system. In order for proteins to be made, copying of the relevant DNA sequences needs to be activated by the cellular processes. A cell is also required to perform the extensive error-correction necessary to ensure faithful replication of DNA to pass on to the next generation. Else random mutation would rapidly destroy its integrity.

Perutz clearly never understood this point since, after Popper's death, he still claimed in a paper entitled *Darwin was right* that "DNA is the score of the music played by the cell" (Niemann, 2014, p. 66). However, the real question is not merely who plays the score, but who or what writes and maintains it. The tune played by the cell is not produced by the DNA, but by the processes of the cell that use the DNA to create the music of life. In this sense, the DNA is not the score; it is an instrument used by the cell.

Thus, the Central Dogma of Molecular Biology, from which Perutz like many others argued, is a viewpoint, and arguably a mistaken one. It is simply a chemical fact about DNA acting as templates for amino acid sequences. It does not alone have the capacity for genetic causation, nor does it exclude macro-level control feeding back onto the genome (Noble 2018). Indeed, such feedback is essential for living cells to work. Once it is accepted that such feedback exists, it opens the potential for agency in evolution at the cellular level.

Research on this issue has greatly progressed since Popper's time. We now understand much better the ways in which the genome is controlled by higher levels of organisation. These controls can be represented mathematically since they determine the initial and boundary conditions for any molecular level representation of biological processes. A concise way of stating this fact is what we have called the principle of biological relativity, which is the statement that, a priori, there is no privileged level of causation (Noble 2012). An important question then is what characterises the physiological properties of the boundaries between the levels of organisation. We have attempted to clarify that question with a variety of examples of experimental and modelling work on the boundaries between different levels Noble, Tasaki et al, 2019). The answer is that there are several different ways in which causation can

[2] There are many demonstrations of this fact. For a brief account see Noble (2018).

be transmitted through the boundaries. Nature seems to have been opportunistic in employing whatever mechanisms were open for evolution to exploit.

The principle of biological relativity (Noble 2012) was derived from the general principle of relativity used in physics, and was inspired by a meeting on Top-Down causation organised by the mathematician and cosmologist, George Ellis. The principle has now been extended back, by Ellis, into a form of relativity of causation in physics itself (Ellis, 2020):

> "no level is a fundamental level with priority over the others, and particularly there is not a primary one at the bottom level. This is just as well, because there is no well-established bottom-most physical level to which physics can be reduced. Every emergent level equally represents an effective theory." Murugan, Weltman & Ellis, 2012)

This means that even chemistry cannot be reduced to physics and, even more so, biology cannot be reduced to chemistry. The fundamental reason for these conclusions have also recently been explored by Stuart Kauffman:

> "the becoming of biospheres falls entirely outside the Newtonian Paradigm. The reason, as we shall see, is that the very phase space of biological evolution – which includes biological functions – persistently evolves in ways that we cannot even prestate, let alone predict. Without a prestated phase space, we can write no law of motion in the form of differential equations, hence we cannot integrate the equations we do not have. Thus, no laws at all entail the stunning unfolding of our, or any, biosphere in the universe." (Kaufmann, 2020)

The arguments in the papers of Ellis and Kauffman can also be viewed as further exploration of the consequences of the existence of open systems, to which we will return in section 4.

3. How organisms can be agents in development and evolution

Popper's argument with Perutz naturally leads to a related question: can mental processes be causal? Clearly, if conscious intention is a mere illusion with no causal power then organisms cannot be active agents in their development and evolution. Popper's and Darwin's arguments for social, including sexual, selection would then carry no weight. That is precisely what many neo-darwinists and strong reductionists believe, which is why we suspect that this issue was at the heart of the Popper-Perutz argument.[3]

The demonstration that there are several ways in which causation can be transmitted between boundaries provides an answer to the commonly held view that mental events cannot be causal. We will show that, as this view is expressed by Jaegwon Kim (perhaps the strongest advocate of the non-causal view), it is incompatible with multi-scale causation in open, such as living, systems. Far from it being the case "that physical causes exclude mental states from causally contributing to the behavior", even a rigorous mathematical analysis of physical

[3] A note for future historians: this question might be settled when the Popper archive at The Hoover Institution, Stanford, becomes open to researchers in 2029.

causation, e.g. in differential equation models, is necessarily incomplete without the contextual, including mental, processes. The forms of causation involved are complementary, they necessarily mesh with each other (Noble & Noble 2020).

We will illustrate this point by considering some of Jaegwon Kim's central arguments and how the complementarity of the forms of causation, together with the harnessing of stochasticity, can be shown to deal with those arguments.

Kim poses the dilemma very clearly:

> "The problem of determinism threatens human agency, and the challenge of scepticism threatens human knowledge. The stakes seem even higher with the problem of mental causation, for this problem threatens to take away both agency and cognition." (Kim, 2000, p 32)

One of the most powerful reasons Kim adduces to justify this threat is the problem of causal exclusion:

> "Suppose then that mental event *m*, occurring at time *t*, causes physical event *p*, and let us suppose that this causal relation holds in virtue of the fact that m is an event of mental kind *M* and *p* an event of physical kind *P*. Does *p* also have a physical cause at *t*, an event of some physical kind *N*? To acknowledge mental event *m* (occurring at *t*) as a cause of physical event *p* but deny that *p* has a physical cause at *t* would be a clear violation of the physical closure of the physical domain……the physical cause therefore threatens to exclude, and pre-empt, the mental cause…..The antireductive physicalist who wants to remain a mental realist, therefore, must give an account of how the mental cause and the physical cause of one and the same event are related to each other." (Kim, 2000, p. 37)

Our answer will be in three parts:

(i) In organisms there can be no causal closure at the micro-level.
(ii) The absence of causal closure depends on the relation of causes between and within levels of organisation.
(iii) The existence of stochasticity and its harnessing enable multiple scenarios to be anticipated by organisms from which ones that most closely instantiate the possible reasons for an action can be selected.

3.(i) No causal closure.

Organisms are composed of levels of organisation nesting within each other, and with the environment. For there to be even the possibility of causal closure we must first define the boundary of the system we wish to investigate in any causal model. At whatever level we do that, whether molecular, cellular, tissue, organs, systems or the organism as a whole, all levels within the chosen boundary will display dynamic processes since organisms are not static. They are never at equilibrium. We cannot therefore represent those processes by static algebraic equations. We note this point because many models used in evolutionary biology and in social sciences such as economics do use static algebraic equations. The problems of causal closure and the openness of the system are then hidden (Noble & Noble 2020).

Most often in non-static models the dynamics are represented by differential equations or by their equivalent. In all cases the equations by themselves do not have solutions until we add the initial and boundary conditions. Those come from the history and the organisation at all levels with which interactions can occur, including the social level at which interpersonal relationships and ideas matter. Those initial and boundary conditions cannot be restricted to events and organisation within the level we have chosen to represent. In open systems, there cannot be causal closure within any level alone. These mathematical considerations lie at the heart of the formulation of the principle of biological relativity.

3. (ii) The relation of causes between and within levels of organisation.

An important outcome of multilevel representations in biology is that the forms of causation up and down across the boundaries between the levels are not equivalent. The history and organisation at higher levels do not determine or contradict the dynamic equations used to represent activity in lower-level processes, they constrain those dynamics, in much the same way that the form and elasticity of the boundaries containing a gas constrain the movements of individual gas particles to produce the overall parameters of pressure, volume and temperature. Someone focussed on the equations for the lower-level dynamics may well not realise that and think that those equations contain all that may be required for causal closure. But they can only do that by not being aware that they will have imported factors (usually in the form of initial and boundary conditions) that are not themselves dynamic in the way in which the dynamic equations are. We consider that this is what is missing from Kim's accounts, and why causal closure is impossible. The forms of causation do not compete, they are necessarily complementary; they mesh together. To modify the relevant part of Kim's statement: the physical cause cannot therefore threaten to exclude, and pre-empt, the mental cause since the physical form of causation is itself necessarily constrained by higher level causation, which is not itself another form of dynamic causation.

3. (iii) The existence of stochasticity and its harnessing

In recent articles we have shown that organisms do not just experience stochasticity. They use it to explore options and select from those options (Noble, 2017; Noble & Noble, 2017, 2018, 2020). This is one of the essential bases of agency, as Popper also saw (section 1.2 above). The reason is that it enables organisms to explore and anticipate many options. To answer Kim again, there isn't just one series of *p*'s there can be an indefinite number. It is by selecting from amongst them that organisms can discover solutions to problems posed by the environment and in interaction with other organisms. They can select the ones that most closely instantiate the possible reasons for an action. Those possible reasons form the cultural context within which organisms with agency can act.

4. Organisms as open systems.

In his 1945 book, *The Open Society and Its Enemies*, Karl Popper contrasted closed, deterministic views of society with open and creative system views. In the open-system perspective, ideas are a vital ingredient of the dynamics of change, involving creative agency; in the closed view, the principal causes of transformation are embedded in the system, determinedly driving it forward. These opposing perspectives have profound consequences. At its extreme, the closed view makes us prisoners of our determinate existence - at best, we can only mitigate the outcome, or "lessen the birth-pangs" by

understanding the nature of that change. In the open view, we can be arbiters of our destiny, using our understanding to bring about creative evolution.

The question is not whether we are free but the extent to which we are open and how we can be free agents in our destiny. Where Marx concluded "It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness." he sees a mostly unidirectional causal chain or at least a profoundly weighted one. Yet, consciousness is a vital capacity of our material existence; it is an essential part of our social reality and not an illusory epiphenomenon. It is not merely a product of our social presence. It engages in our choices, individually and socially. We are not free from our material existence, nor do we exist apart from it; clearly, we depend upon it; but we have agency within it. As Popper puts it in countering the hard historicist Marxist view "the future depends on ourselves, and we do not depend on any historical necessity." (Popper, 1945, I p. 3)

Thus, for determinists, or historicist accounts, forces beyond our control overwhelmingly govern our destiny. It is as if we can see nature unfold, yet have no way to use that vision to alter its direction. In this view, consciousness is a product and not a player in our history. Yet, agency and awareness play vital roles in the integrity of living organisms. We are players in our history and not merely products of it.

The parallels with biology are not surprising; being social is part of our biology. A closed system has no creativity, and this was Descartes' problem. If organisms are viewed as mere machines, working like clocks, then where is the agency? If we are automata, then where is our will? We end with a body-soul, or body-mind split, and this, in turn, creates the problem of how one (immaterial) can be influenced by the other (material).

We have argued that treating organisms as closed systems leads to another, but similar kind of dualism - a materialist dualism, or something within the system that drives the system. This dualism unfolds as genetic determinism, with genes as driving agents. We become prisoners of our genes. A bit of the system that is so important that organisms become mere 'vehicles' in its transgenerational transmission, or as Dawkins put it in *The Selfish Gene*:

"We are survival machines – robot **vehicles** blindly programmed to preserve the selfish molecules known as genes. This is a truth which still fills me with astonishment." (Dawkins, 1976)

This 'truth' separates the gene (the driver) from its vehicle (the organism), thus creating an unnecessary materialist dualism. It fills us with astonishment that it is presented as 'truth'.

Popper's book was a robust defence of the open society. It is not therefore surprising that he should have championed the role of agency in biology. All life depends on the exchange of matter and energy with the environment, including social arrangements with other organisms. We can represent this openness as a matter of degree, dependent on increasing complexity of self-organisation, represented diagrammatically in Figure 1.

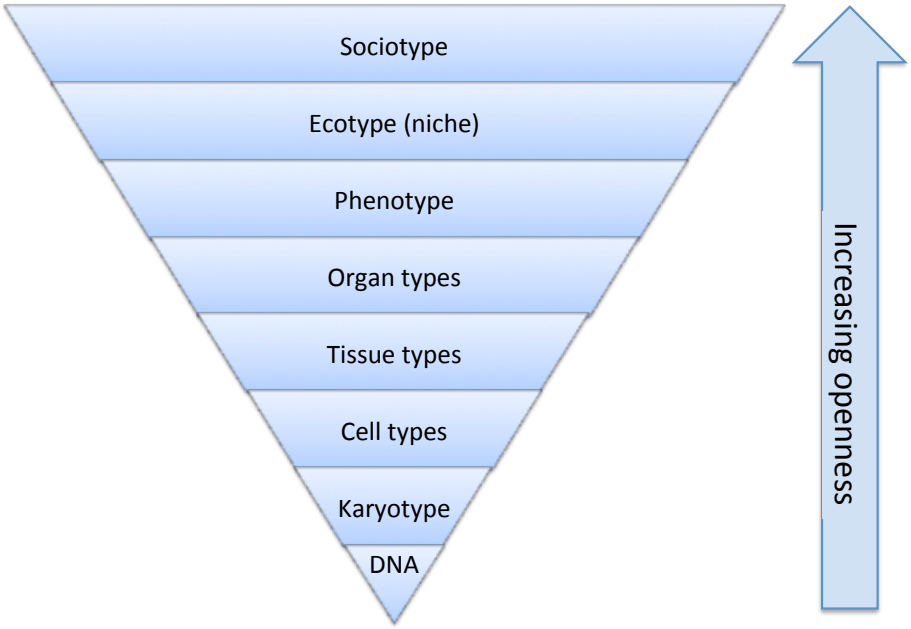


Figure 1. Levels of organisation in organisms. The width of the cone represents the degree of openness. At every level, the system is open, yet functionally constrained by the levels above, which are more variable and malleable.

The principle behind this diagram is that the sensitivity to variation dependent on the environment of each level of organisation increases as we move up from the molecular level through to the social level. At the bottom, DNA, like all molecules in isolation, cannot be said to possess agency. We only begin to encounter agency, in the sense of self-maintenance of an organised system, when we move to levels higher than the molecular.

Notice that we have specified the karyotype as a level of organisation. We have done that to recognise one of the most important discoveries in genome wide sequence studies, which is that the associations of individual genes (DNA sequences) with phenotypes, healthy or diseased, is generally very low. The hope of gene-centric biology delivering the basis of cures for common diseases has not materialised. We now know that it is the complete genome that is important to the overall genetic contribution to inheritance: the omnigenic hypothesis (Boyle et al, 2017). In his book *Genome Chaos* Heng (2019) goes even further and calls for an integrated view of the karyotype, which refers to the complete chromosomic structure as itself an organised system. The idea is not just chromosomes as a set of individual genes, but rather the activity of chromosomes in genome control and rearrangement. The karyotype is therefore a distinct level of organisation, arranged as a 4D structure rather than a 1D sequence. The karyotype is itself sensitive to control from higher levels.

The other levels are well-known already from many similar diagrams of the hierarchical organisation of organisms. But our top level, sociotype, does require some comment. The social interactions are where conscious agency is generated. The fluidity and contingency of

the sociotype is also an important source of stochasticity forming the clay out of which functional novelty can be generated.

Another way of illustrating the degrees of openness of biological systems is to represent the evolution of openness as a time series of major transitions. We attempt to do this in Figure 2.

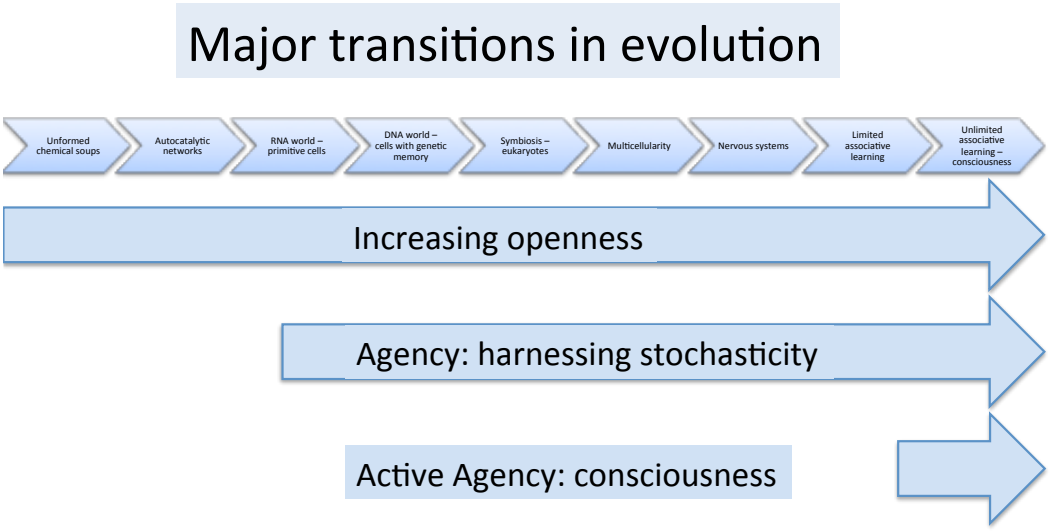


Figure 2. Evolution of organisms represented as a possible time sequence with major transitions.

The important point to note about this figure is that each transition depends on the evolution of the prior transitions. There is a ratchet effect (Hoffmann, 2012). Each time a major transition is achieved, it opens the way for the next transition. The nature of the evolutionary process also changes. Evolution itself evolves (Noble at al, 2014).

This process of emerging evolution is an old idea in evolutionary biology, explored notably in Maynard Smith and Szathmary's (1995) book *The Major Transitions in Evolution* and more recently Ginsburg and Jablonka's (2019) book *The Evolution of the Sensitive Soul*. It also goes back at least to Lamarck's idea of *Le pouvoir de la vie* (the force of life – Lamarck, 1809). Lamarck has been widely ridiculed for his ideas and this one is no exception. However, Lamarck was not a vitalist. He was firmly a materialist (Pichot 1994). The current idea of a ratchet process in the development of major transitions corresponds well to what he had in mind, which was that each stage in the evolution of organisms created a further way up the ladder of complexity. It is widely thought that Lamarck also thought that the ladder concept best represented the transformation of species with no branching. This is not correct. Lamarck himself changed his mind and replaced his ladder with a tree of life (Noble, 2020).

The specific point about such a diagram that is relevant to our article is that the transitions all mark the increasing development of openness. This is particularly true for the later stages, where we have followed Ginsburg and Jablonka in distinguishing as separate stages the development of nervous systems, on the basis of which associative learning can develop, later to become unlimited associative learning, which is their proposed marker for the development of consciousness: the last transition in Figure 2.

5. Conclusions: are active and passive Darwinisms entirely separate?

So far in this article we have expounded and interpreted Popper and Darwin on the issue of the active-passive distinction of Popper and the artificial-natural distinction of Darwin. We have shown that they correspond. We have shown that Popper's active Darwinism is Darwin's artificial selection and Popper's passive Darwinism is Darwin's natural selection. The correspondence is very close.

But we now have to row back somewhat. Like many distinctions in language and philosophy they are not so clear cut as may first appear. What, after all, is natural selection? Darwin saw the contrast as between what humans do artificially through selective breeding and what happens naturally when the environment acts as the passive filter of natural selection.

But what creates that environment?

Consider this:

Tools and language facilitate agency. We use machines and communication for reasons - to do something or to express something. Organisms use the first to do things and the second to communicate. With tools, organisms obtain food and build protection from the physical environment. Language and writing improve communication and enable ideas to be explored, transmitted and transformed across generations and between groups or individuals. Using language enables communication and understanding of intention. With tools and language, humans created civilisations and extended abstract thought through literature and art. This creativity, in turn, influences the way we perceive the world. The built environment and the psychological texture of society is the explorative embodiment of niche creation and through which selective pressure affects human evolution. We are not merely hunter-gatherers in a concrete jungle; we are evolving organisms in a created niche. What we expect of each other and ourselves affects our physiology. Sometimes we suffer as a result. If natural selection is the measure of fitness to survive and reproduce, then we must ask what it is that is doing the selecting, and what it is that is being selected. If it is the environment, then clearly, we also consciously create that environment. Humans are not alone in doing this.

Just as we have selected dogs, cats and other domesticated and farm animals, so we also choose each other, as partners, as friends, and often with whom we work. Our social being shapes us as individuals, just as we form our social being. The relationship between consciousness and nature is intertwined and not separate.

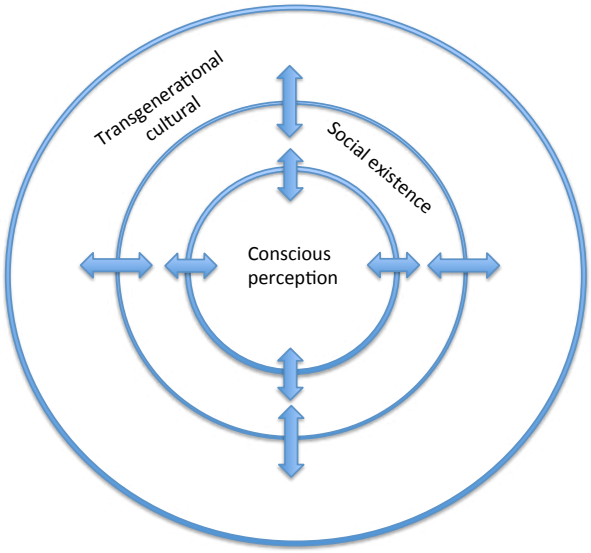


Figure 3. Diagram of functional interactions between social existence and conscious organisms with agency. Agents with conscious perception interact across functional boundaries with their social existence, which in turn facilitates interacts through the transgenerational cultural inheritance, allowing the creation of ideas, viewpoints, opinions, attitudes and actions. We have used double-headed arrows to emphasise that there is no privileged circle of interaction. There is a continuous two-way interaction. The transgenerational cultural inheritance forms what Popper (1978) referred to as World 3.

All the time, over different time scales the environment itself evolves as a consequence of the agency of organisms. We illustrate this in Figure 3. Popper understood and emphasised the two-way process:

> Our minds are the creators of world 3; but world 3 in its turn not only informs our minds, but largely creates them. The very idea of a self depends on world 3 theories, especially upon a theory of time which underlies the identity of the self, the self of yesterday, of today, and of tomorrow. The learning of a language, which is a world 3 object, is itself partly a creative act and partly a feedback effect; and the full consciousness of self is anchored in our human language (Popper, 1978).

The environment is what we, the organisms, have created, including the transgenerational cultural inheritance. Organisms have also completely changed the physical and chemical environment. The environment of the earth today is nothing like the environment of the earth when life first evolved. So, over all the more than 3 billion years since then we, organisms, have altered the environment in almost all relevant respects. Even natural selection eventually operates as it does today through the actions of organisms. The world of nature (natural) is not merely physical, it is biologically functional (selective).

Of course, there is a difference of time scale. At any one period of time we can distinguish between natural and artificial selection. Our concluding point does not invalidate the distinctions both Popper and Darwin were drawing. But there is nevertheless an important

process to add to what we have described in this article. There is a continuous circular interaction between the activity of organisms and the development of the environment. That, in turn, becomes the basis of natural selection. Even natural selection is therefore not entirely passive. Humanity itself is now the greatest driver of evolution through rapid alteration of the global environment (Corning, 2020).

Acknowledgements

No funding supported this work. We have no conflicts of interest.

References

Boyle, E.A., Li, Y., Pritchard, J.K.: An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 169, 1177-86. (2017)

Brosnan S.F., De Waal, F.B.: Monkeys reject unequal pay. Nature. 425, 297–299. (2003)

Corning, P.: Beyond the Modern Synthesis: a framework for a more inclusive biological synthesis. Progress in Biophysics and Molecular Biology. 153, 5-12. (2020)

Dawkins, R.: The Selfish Gene. Oxford, OUP. (1976)

Darwin, C.: The Origin of Species. (1859) (Macmillan Collectors Library edition, (2004))

Darwin, C.: The Descent of Man, and selection in relation to sex. London, Murray. (1871) I:35-36.

Ellis, G.F.R.: The Causal Closure of Physics in Real World Contexts. Nous, (2020) https://aeon.co/essays/heres-why-so-many-physicists-are-wrong-about-free-will

Essler, J.L., Marshall-Pescini, S., Range, F.: Domestication does not explain the presence of inequity aversion in dogs, Curr. Biol. 27, 1–5 (2017)

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

Niemann, H.-J.: Karl Popper and the Two New Secrets of Life. Mohr Siebeck, Tübingen. (2014)

Ginsburg, S., and Jablonka E.: The Evolution of the Sensitive Soul. Learning and the Origins of Consciousness. Cambridge, Mass, MIT Press. (2019)

Jablonka, E.: The evolutionary implications of epigenetic inheritance. Interface Focus. 7, 20160135. http://dx.doi.org/10.1098/rsfs.2016.0135 (2017)

Heng, H.: Genome Chaos, Rethinking Genetics, Evolution and Molecular Medicine. London, Academic Press (2019)

Hoffmann, P.M.: Life's Ratchet. How molecular machines extract order from chaos. Basic Books (2012)

Huxley, J.: Evolution. The Modern Synthesis. London, George Allen and Unwin (1942)

Kauffman, S. 2020. Eros and Logos. Journal of the Theoretical Humanities. 25, issue 3 (2020)

Kim, J.: Mind in a Physical World. An Essay on the Mind-body problem and Mental Causation. Cambridge, Mass, MIT Press (2000)

Lamarck, J-B.: Philosophie Zoologique, original edition of 1809 with introduction by André Pichot. Paris, Flammarion (1809, 1994)

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 (2014)

Maynard Smith J., Szathmary, E.: The Major Transitions in Evolution Oxford: OUP (1995)

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, K., K, Davidson, N.O. *et al.*,: Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. 274, 18470–18476. (1999)

Murugan, J., Weltman, A., Ellis, G.F.R.: (eds). Foundations of Space and Time: Reflections on Quantum Gravity. Cambridge, Cambridge University Press (2012)

Niemann, H-J.: "Nachwort des Herausgebers," in Wissen und das Leib-Seele-Problem, edited by K. R. Popper (Mohr Siebeck, Tubingen), pp. S510–S546; (afterword of the publisher; K.R. Popper, Knowledge and the mind-body problem). Sections 31-33 (2012)

Noble, D.: A theory of biological relativity: no privileged level of causation. Interface focus, 2, 55-64 (2012)

Noble, D. 2017. Evolution viewed from physics, physiology and medicine. *Interface Focus* **7**, 20160159 (2017)

Noble, D.: Central Dogma or Central Debate? Physiology, 33, 246-249.

Noble, D.: 2020. Charles Darwin, Jean-Baptiste Lamarck, and 21st century arguments on the fundamentals of biology. Progress in Biophysics and Molecular Biology. 153, 1-4. https://doi.org/10.1016/j.pbiomolbio.2020.02.005 (2020)

Noble, D., Jablonka, E., Joyner, M.J., Muller, G.B., Omholt, S.W.: Evolution evolves: physiology returns to centre stage. Journal of Physiology. 592, 2237-2244. (2014)

Noble, R., Noble, D.: Was the watchmaker blind? Or was she one-eyed? Biology **6**(4), 47 (2017)

Noble, R., Noble, D.: Harnessing stochasticity: How do organisms make choices**?** Chaos 28, 106309; https://doi.org/10.1063/1.5039668 (2018)
Noble, R., and Noble, D.: Can reasons and values influence action: how might intentional agency work physiologically? Under review Journal of the General Philosophy of Science. (2020)
Noble, R., Tasaki, K., Noble, P. J., Noble, D.: Biological Relativity requires circular causality but not symmetry of causation: so, where, what and when are the boundaries? Frontiers in Physiology 10, 827 (2019)
Pichot, A.: Introduction, in Philosophie Zoologique. Paris, Flammarion (1994)
Popper, K.R.: The Open Society and its Enemies. London: Routledge (1945)
Popper, K.R.: Indeterminism is not enough. Encounter 40(4), 20–26. (1973)
Popper, K.R.: Three Worlds. University of Michigan Tanner Lecture. https://tannerlectures.utah.edu/_documents/a-to-z/p/popper80.pdf
Popper, K.R., Eccles, J.C.: The Self and Its Brain New York, Springer International (1977)
Saribasak, H., Gearhart, P.: Does DNA repair occur during somatic hypermutation? Seminar. Immunol. 24, 287–292 (2012)
Shapiro, J.A.: Evolution: A View from the 21st Century Upper Saddle River, NJ, Pearson Education Inc (2011)
Weismann, A.: Die Allmacht der Naturzüchtung; eine Erwiderung an Herbert Spencer. Jena, Fischer, (the omnipotence of natural breeding; a reply to Herbert Spencer) (1893)

frontiers in Physiology

REVIEW
published: 18 July 2019
doi: 10.3389/fphys.2019.00827

Biological Relativity Requires Circular Causality but Not Symmetry of Causation: So, Where, What and When Are the Boundaries?

*Raymond Noble[1], Kazuyo Tasaki[2], Penelope J. Noble[2] and Denis Noble[2]**

[1] Institute for Women's Health, University College London, London, United Kingdom, [2] Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom

Since the Principle of Biological Relativity was formulated and developed there have been many implementations in a wide range of biological fields. The purpose of this article is to assess the status of the applications of the principle and to clarify some misunderstandings. The principle requires circular causality between levels of organization. But the forms of causality are also necessarily different. They contribute in asymmetric ways. Upward causation can be represented by the differential or similar equations describing the mechanics of lower level processes. Downward causation is then best represented as determining initial and boundary conditions. The questions tackled in this article are: (1) where and when do these boundaries exist? and (2) how do they convey the influences between levels? We show that not all boundary conditions arise from higher-level organization. It is important to distinguish those that do from those that don't. Both forms play functional roles in organisms, particularly in their responses to novel challenges. The forms of causation also change according to the levels concerned. These principles are illustrated with specific examples.

Keywords: biological relativity, downward causation, circular causality, entangled causation, boundaries in physiology

OPEN ACCESS

Edited by:
Marko Marhl,
University of Maribor, Slovenia

Reviewed by:
Giuseppe Longo,
Centre National de la Recherche Scientifique (CNRS), France
Leonardo Bich,
University of the Basque Country, Spain

***Correspondence:**
Denis Noble
denis.noble@dpag.ox.ac.uk

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

Received: *19 February 2019*
Accepted: *13 June 2019*
Published: *18 July 2019*

Citation:
Noble R, Tasaki K, Noble PJ and Noble D (2019) Biological Relativity Requires Circular Causality but Not Symmetry of Causation: So, Where, What and When Are the Boundaries? Front. Physiol. 10:827. doi: 10.3389/fphys.2019.00827

INTRODUCTION

The principle of Biological Relativity is that, *a priori*, i.e., before performing the relevant experiments, there is no privileged level of causality (Noble, 2012). In multi-scale networks of interactions, as found everywhere in organisms, any parts of a network at any level might affect every other part.

The principle is based on mathematical approaches to understanding biological processes. While the differential (or equivalent) equations represent the dynamics of the components of the system, the initial and boundary conditions represent the historical and contextual (environmental) factors without which no specific solutions to the equations would be possible.

The principle has found many applications in physiology and in other fields of biology. This is not surprising since the mathematical point being made is a necessary one, regardless of whether the components are molecular (genes, proteins, and metabolites), networks (at all levels), cells, tissues, organs, or any other kind of component. Moreover, in practice the principle has been applied many times in physiology even before it was formulated as a mathematical principle. All forms of feedback

between levels in biological systems inherently assume the principle. It can therefore be seen as formalizing an idea that has been inherent in physiology, at least since Claude Bernard in the 19th century (Bernard, 1878, 1984; Noble, 2008, 2013), and Walter Cannon in the 20th century (Cannon, 1932) formulated the ideas of homeostasis. Nevertheless, the principle is not limited to the usual interpretations of homeostasis as linear circularity. The regulatory systems in organisms do much more than act like sophisticated thermostats. There are no fixed set-points. There are sets of set-points each of which can vary as the organism seeks to maintain itself. Buiatti and Longo (2013) express this point by using the word homeorhesis in place of homeostasis:

> "Biological objects are, as discussed by Waddington, "homeorhetic," as opposed to homeo-static, in the sense that, during their cycles, they keep changing. Moreover, their onto-phylogenetic path is largely unpredictable, though preserving, as long as possible, the internal coherence of an organism and its relations to the ecosystem. It is unpredictable because of the random effects at each level and of the bio-resonance effects between different levels."

As our article will make clear, the various levels communicate both randomness and order between each other. We agree therefore with Rosen in *Life Itself* (Rosen, 1991, 2000), that it is the *organization of the organism itself* that constrains the component parts, not the other way round. That organization forms the basis of active agency in organisms (Noble and Noble, 2017; Noble, 2018). One of the aims of this article is to interpret the principle of biological relativity in a more radical way.

The principle also raises many other questions. The aim of this paper is to formulate those questions and attempt to resolve them. Foremost amongst those are questions concerning what is meant by a boundary.

As physiologists we might think that question has an obvious answer. Cells have membranes, tissues have surfaces, organs have shapes with anatomical boundaries, the organism has its outer structure, skin. But where are such boundaries of the great systems of the body, the immune, nervous, circulatory, digestive, respiratory, reproductive, and hormonal systems? Merely to ask the question shows that the answer is not obvious. Anatomy is not necessarily the best basis for defining a functional boundary. To varying degrees, the boundaries used in models are somewhat arbitrary. And even when we can identify an anatomical boundary it is not necessarily the mathematical computational boundary.

As an example of the kind of problem we will address consider the problem faced in modeling the electrophysiology of the heart during the 1980s when processes involving changes in ion concentrations were added to the existing equations for the gating of ionic channels (McAllister et al., 1975). Prior to the DiFrancesco-Noble equations (DiFrancesco and Noble, 1985) this had not been done in any systematic way. Yet it was necessary to incorporate changes in K^+ concentration in intercellular spaces to understand how these could make a non-specific cation channel conducting both Na^+ and K^+ behave like a pure K^+ channel. The new model was completely successful in achieving this aim. But that was not possible without changing the boundaries of the model. One of us explained this boundary problem in 2012:

> "The obvious next step was to develop the McAllister–Noble–Tsien model of 1975 to replace i_{K2} by i_f. But that was much easier said than done. It took a full 5 years of development. This was because it was not just a matter of replacing one ionic channel mechanism by another. It also involved modeling global ion concentration changes for the first time in an electrophysiological model of the heart, including the intracellular calcium signaling. Dario and I did that because it was necessary to explore fully what we had discovered. We did not know then that we would be creating the seminal model from which virtually all subsequent cardiac cell models would be developed. There are now over a hundred such models for various parts of the heart and many different species[1]."
>
> Extending biological models is often like tumbling a row of dominoes. Once one has fallen, many others do too. The reason is that all models are necessarily partial representations of reality. The influence of the parts that are not modeled must either be assumed to be negligible or to be represented, invisibly as it were, in the assumed boundary conditions and other fixed parameters of the model. Once one of those boundaries is removed, by extending out to a different boundary, other boundaries become deformed too. In this case, modeling external potassium changes required modeling of the influence of those changes not only on the ion channels already in the model, but also on exchange mechanisms, like Na-K-ATPase (sodium pump) and the Na-Ca exchanger. That, in turn, required the model to extend to modeling internal sodium concentration changes, which in turn required modeling of intracellular calcium changes, which then required modeling of the sarcoplasmic reticulum uptake and release mechanisms. For a year or two it was hard to know where to stop and where to stake out the new boundaries" (Noble et al., 2012) (Page 58).

Even more difficult is the fact that physiological boundaries can be dynamic. When and why they occur are also important questions since it is at boundaries that many of the vital functional processes occur. Recall that the nervous system develops from the embryonic "boundary," the ectoderm, and in single cell organisms the surface membrane can be regarded as its nervous system. Organisms are open systems, so their boundaries are necessarily where much of the action occurs.

DEFINITIONS

Biological Relativity

Biological relativity is the principle that there is, *a priori*, no privileged level of causation. The necessary mathematical basis of the principle was first proposed in 2012 (Noble, 2012) when it was categorized as a "theory." It is better viewed as a principle since it expresses the conceptual point that there is no empirical justification for privileging any particular level.

Upward Causation

Upward causation is the set of processes by which the lower elements in a system interact and produce changes at higher levels. In differential equation models these processes are

[1] www.cellml.org

described by the dynamics represented by the differential equations themselves.

Downward Causation

Downward causation is the set of constraints imposed by the higher levels on the dynamics at lower levels through determining many of the initial and boundary conditions. El-Hani and Queiroz (2005) use the term, "Downward Determination", but they agree that what is involved is something that "can be understood in terms of constraints that the condition of belonging to a system-token of a given kind imposes on the behavior of the components." The sense of cause we are using includes that of determination. We agree that there are different kinds of causation (Noble, 2016) (pp 176–181). Mossio et al. (2013) also emphasize the role of higher level constraints when they refer to "emergent causal powers exerted as constraints, and we claim that biological systems crucially differ from other natural systems in that they realize a closure of constraints."

Initial Conditions

Initial conditions are the initial values of each dynamic element at lower levels. They are determined by the history of development of the system, including stochastic variation as well as previous states of the system. The upward and downward forms of causation interact (**Figure 1**).

Boundary Conditions

Boundary conditions are the conditions attributable to interaction with the environment. In partial differential equation models these conditions are represented by the state of the spatial boundary of the system. In ordinary differential equation simplifications in which spatial changes are assumed to be instantaneous these conditions are represented by the constant coefficients at any moment in time.

Structure

Structure is also a condition that could be regarded as initial or boundary according to the modeling chosen.

Conditioned Causation

Conditioned causation is a state of a system where it would be misleading to attribute causation to any particular element.

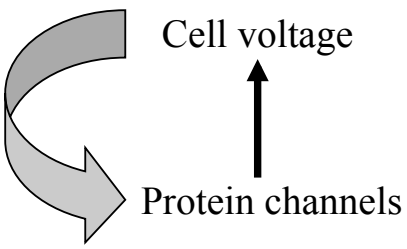


FIGURE 1 | An example of circular causality in physiology. The Hodgkin cycle represents the fact that global cell properties, such as electric potential, control molecular level properties, such as ion channel proteins, which in turn determine changes in cell properties.

MAIN SECTIONS

How Do Upward and Downward Forms of Causation Differ?

The existence of both upward and downward forms of causation is often represented as circular causality. While obviously correct in the sense that both forms exist and, in many ways, must influence each other, such diagrams hide the fact that there is an important difference. The upward and downward forms are necessarily different, just as the initial and boundary conditions of differential equation models are clearly not the differential equations themselves.

It is also important to distinguish conceptual questions about how we see things from what nature does. Nature is a continuum on which we impose somewhat arbitrary boundaries which are dependent on the models we use to understand nature. This point should be borne in mind throughout this article.

Upward Causation

Lower levels influence higher levels through the dynamic changes represented by the differential equations. These will result in global changes, for example in concentrations of ions, metabolites, proteins in cells, tissues and organs and these may in turn trigger further changes at any or all of the higher levels.

As an example, consider the processes involved in calcium movements in the various kinds of muscle in an athlete during vigorous exercise. Too much intracellular free calcium may cause maintained serious problems in the athlete's heart, skeletal muscles or smooth muscles. Any of these, such as a sudden heart attack, may cause severe pain, in turn leading the athlete to collapse. Then the influences become wider and wider as the team coach and physiotherapist enter the scene, which further leads to social interactions. This is an example of unintended effects at a lower level triggering many other events at higher and higher levels.

Downward Causation

Now let's consider how the athlete became an athlete in the first place. He spent hours a day training. This was his decision. It wasn't a decision of the calcium ions in his muscles, nor of the gene sequences in his DNA. Molecules and ions are not causes in that sense (Noble, 2016). It was a high-level choice that he made (Noble and Noble, 2018) and it resulted in many changes in his musculoskeletal, respiratory and cardiovascular systems, all becoming more powerful. Many of these changes came about through exercise influencing gene expression of the proteins in muscles, the lungs and the cardiovascular system. This in turn changes the innumerable boundary and initial conditions under which all the muscles in the athlete's body behave. The changes influence how much muscular, breathing and cardiovascular capacity the athlete has. Although the differential equations for each of his muscle fibers will still be much the same, the changed initial and boundary conditions now ensure that the athlete can do the same or even more vigorous exercise without experiencing disabling fatigue and cramp. This is an undeniable physical effect at the molecular level arising from the athlete's choice of lifestyle.

It doesn't alter the laws of molecular behavior. It alters the solution to the equations for those laws.

Identical Twin Athletes

At this point a rigorous genetic reductionist (Comfort, 2018; Plomin, 2018) might want to argue that no downward causation was involved. The athlete was simply born with the right genes to develop as an athlete. While that must be true – someone suffering from a genetic disease like muscular dystrophy, for example, could not do what the athlete does – it is far from being the complete story. Studies of identical twins who chose very different kinds of sports and exercise training show that very clearly. **Figure 2** is taken from such a study (Keul et al., 1981). The runner and the weightlifter showed completely different effects on their body physique. Bathgate et al. (2018) have recently published a more extensive study of many differences in muscle and cardiovascular health and performance in monozygotic twins. They conclude that "the cardiovascular and skeletal muscle systems exhibit greater plasticity than previously thought." Furthermore they have identified precisely which RNA levels of control are changed by the lifestyle choices.

Genome-Wide Association Studies

Genome sequence studies have failed to find just a single or a very few genes that are strongly correlated with athletic performance. A literature search on publications in the period 1997–2014 showed at least 120 genes show correlations with athletic performance, many of the correlations being very small (Ahmetov and Fedotovskaya, 2015). That number of correlated genes is likely to grow as even more extensive GWAS results appear. So much so that some GWAS scientists have come to the conclusion that virtually the whole genome may be correlated with most phenotypes, the so-called omnigenic hypothesis (Boyle et al., 2017). A study of 1520 endurance athletes and 2760 controls "did not identify a panel of genomic variants common to these elite endurance athlete groups" (Rankinen et al., 2016), and see their earlier studies (Rankinen et al., 2000, 2005). One recent study comparing the impact of genes and environment concluded "that the traditional argument of nature versus nurture is no longer relevant, as it has been clearly established that both are important factors in the road to becoming an elite athlete." (Yan et al., 2016) In a review of elite athletic performance Joyner and Coyle concluded "finding genetic markers that are strongly predictive of either success in endurance athletic performance or somehow preclude it is likely to be a daunting task because of the many cultural and environmental factors that contribute to success in sport, the many physiological factors that interact as determinants of performance, and the heroic nature of the training required" (Joyner and Coyle, 2008).

Epigenetic Control

The main reason for the failure to explain athletic performance from genetics alone is that the genome is controlled by the organism and its life-style experiences through extensive epigenetic control.

As an example, athletes have lower heart rates than non-athletes, which was once attributed to greater vagal tone. The changes have now been traced to microRNAs that downregulate expression of the HCN gene, so that the depolarizing current (I_f) produced in the sinus node cells is reduced by as much as 50% (D'Souza et al., 2017). Moreover, that changes in autonomic tone could not be the explanation was shown as long ago as 1967, but the authors could not at that time identify the mechanism (Sutton et al., 1967). The advent of modern techniques for identifying epigenetic control has transformed this field of study.

The interface between DNA and epigenetic control is therefore another important boundary. It is one of the means by which the organism controls its genome as a "highly sensitive organ of the cell" (McClintock, 1984). This boundary was first identified by Waddington (1957), who was the

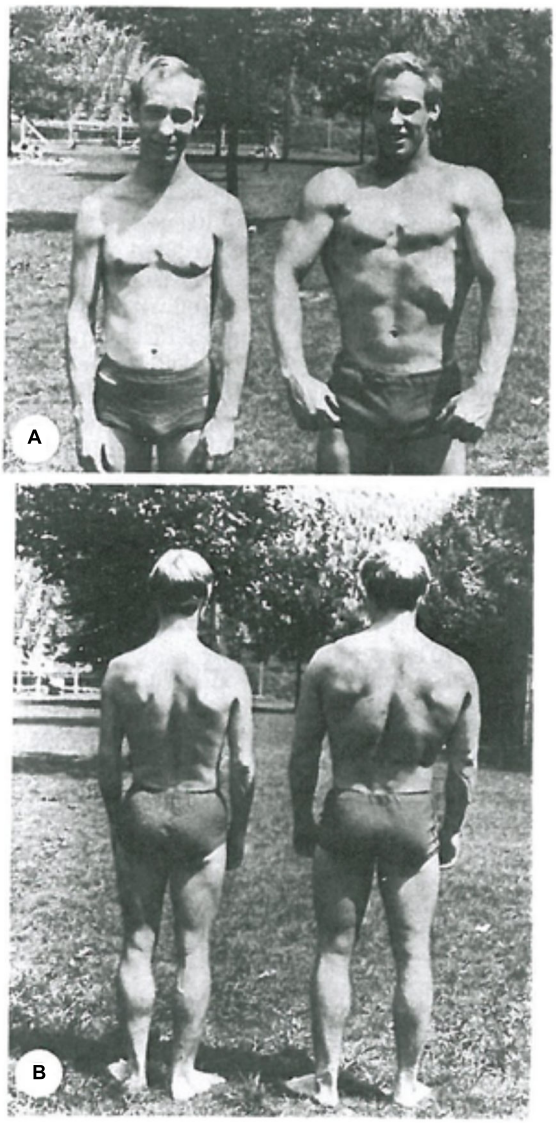


FIGURE 2 | Identical twins. **(A)** Long-distance endurance runner. **(B)** Weightlifter. Notice the highly developed calf muscles in the runner and the contrast with the highly developed arm and chest muscles of the weightlifter. Reproduced with permission from the publisher of Keul et al. (1981).

originator of the term epigenetics. Since then many forms of epigenetic control have been discovered. This control is so effective in transmitting the adaptive properties of the networks that most gene knock-outs have very little effect. The exceptions are, of course, the rare genetic diseases, such as cystic fibrosis, where the networks do not have sufficient plasticity to cope with a knock-out. But, in general, plasticity is common. In yeast, for example, 80% of gene knock-outs are silent in the sense that they produce no phenotypic effect when the yeast is well-nourished (Hillenmeyer et al., 2008). That result has been broadly confirmed by Galardini et al. (2018) who have shown the extent to which the effect of a gene deletion depends on the genetic background. They conclude that "interpretation of the impact of genetic variants on the phenotypes of individuals would likely need detailed gene-phenotype information in more genetic backgrounds than that of a model individual." We would add that the phenotype background must also be relevant. The boundary between regulatory networks and DNA is necessarily a two-way boundary. The regulatory networks can filter genetic changes, acting as what we have characterized as a "cloud" at the boundary (Noble and Noble, 2017; Galardini et al., 2018).

The downward forms of causation represented by the choices made by the individual organism and the influences of its environment must therefore be widespread and necessary.

Open Systems and Their Boundaries

One reason why boundaries are important is that all organisms are open systems. The interaction with the environment is an essential process of being alive. It is across the boundary between the organism and its environment that all the exchanges of energy and matter occur. The same principle applies within the organism. There are boundaries between cell components, between cells, tissues, organs, . . .all the way up. Downward causation can be seen to be traversing a cascade of boundaries. Each level of organization provides the boundary and initial conditions for solutions to the dynamic equations for the level below.

Are All Forms of Downward Causation Functional?

So far, we have established why downward causation is effective and that its necessary effectiveness is mathematically demonstrable. Now let's look at those initial and boundary conditions more carefully. When we inspect the most complete of the mathematical models of skeletal, cardiac and smooth muscles we can identify more than 100 constants in the equations (DiFrancesco and Noble, 1985; Yang et al., 2003; Shorten et al., 2007). Each of those, alone or in combination, reflects an initial or boundary condition. So, there are at least that many parameters that might be sensitive to causative action from higher levels. These parameters are determined by the state of the boundaries between higher and lower levels. In reality there will be many more. The model is just a partial abstraction of reality.

Could all parameter changes in the initial and boundary conditions be attributable to downward causation? There are several reasons why that cannot be true.

The lowest boundary: molecular stochasticity

As Robert Brown showed in 1827, fine particles suspended in water show stochastic movement which was eventually shown by Einstein to be produced by random bombardment by individual water molecules. The molecules in cells are an aqueous suspension and must also be subject to Brownian motion. Water, and all molecules, will also be subject to quantum mechanical randomness. On some interpretations of quantum mechanics, all objects are subject to such randomness (Becker, 2018), although it becomes negligible at a large enough scale.

This is a boundary *within* the system. In a sense it is a boundary between levels or scales. Later in this article we will discuss how organisms use this and other boundaries between levels. But here it is sufficient to note that the boundary is fuzzy. There is no precise cut-off scale at which molecular stochasticity, whether quantal or not, becomes negligible. This is a major issue in the interpretation of quantum mechanics (Becker, 2018), but it need not detain us here. We note that it is a good example of a boundary that cannot be given a precise anatomical location. In a sense the boundary is everywhere. It is a boundary between levels of organization.

Functional and non-functional initial and boundary conditions

Influences on a system from its environment and higher scales can be of at least two kinds. Some will be contingent and even apparently random. These will provide opportunities for novelty in the organism's behavior, in much the same way as we have described in related articles (Noble and Noble, 2017, 2018). Stochasticity can be used by organisms to generate novelty. That can happen whatever the origin of the stochasticity, whether molecular within the organism or environmental without the organism.

But what is usually meant by downward causation are influences that arise from the regulatory *organization* at higher levels. Organization is what defines a level as distinct from a scale. Cellular organization defines the level of cells, organ organization defines the level of organs, and so on through the levels.

What do we mean here by organization? What precisely is homeostasis? Yet again, the common diagrams of upward and downward causation can be misleading. Regulatory processes in the body are rarely simple feedback loops maintaining a specific parameter, like blood pressure or temperature, constant. Nor is the circularity a simple feedback loop that can be described as a linear sequence of causation: A leads to B which leads to C and so on. This way of thinking leads to the need to specify the *direction* of the causation, in turn leading to the idea of emergence, usually interpreted to mean that the higher-level organization emerges *from* the lower level activity. But how can that be? At the lower level we can't even *see* the organization. Low-levels do not possess such organization. The constraints of higher-level organization will be represented by a seemingly disorganized set of initial and boundary conditions. We don't for example "see" the organization of bird haemoglobins as they vary according to different altitudes by sequencing their genomes. At that level, the different species have used different molecular level solutions to evolve haemoglobins for high and low altitudes. At

the functional level, the haemoglobins can be characterized as functional for the altitude at which they live so that all high-altitude birds show higher affinity for oxygen even though the DNA sequences are different (Natarajan et al., 2016). Only at the higher level of organization is the function of the genome changes evident.

We have elaborated on this problem in a previous article (Noble and Noble, 2017). From the molecular level of DNA, RNA, proteins, metabolites, ions etc., we will not be able to see the organization. As we noted earlier, it was not the athlete's calcium ions that caused his decision to be an athlete.

Emergence – a-mergence?

For these reasons, we have argued elsewhere for replacing the term e-mergence (suggesting privileging upward causation) with the neutral term a-mergence (Noble and Noble, 2019). In terms of causation, this requires replacing the linear sequence A causes B which causes C etc., with the existence of the state X, the occurrence of which means that A, B, and C etc., will also occur. This is the characteristic of high-level attractors. Once they occur, they take over the organization of the system. This fact becomes hidden when we insist on a linear causation viewpoint. Yet it is implicit when we solve model differential equations numerically since all factors are taken into account at each integration step. In a cell model we don't, for example, first calculate the influence of all the global cell parameters (such as potentials and concentrations) and then calculate the influence of the microscopic elements (such as transporter and enzyme states) separately.

This issue of simultaneity of action is fundamental. Another way of expressing it is to ask whether circular causality can be said to have a direction. Diagrams often strongly imply that they do, by giving the impression that, if one could be a nano-level observer, one would see one stream of causation running upward and another flowing downward. That picture is far from the reality. This is where the mathematical interpretation of circular causality is so useful in providing a totally different picture of the situation, since the integration procedures must proceed simultaneously (Noble, 2012). A nano-level observer would surely see something more like a cloud of happenings, which would not be resolvable into separate streams of happenings[2].

In this respect, the Biological Relativity interpretation of multi-level causality resembles wave theories of quantum mechanics. Electrons circling a nucleus, for example, are referred to as a cloud because the wave interpretation does not, and cannot, identify where any particular electron may be. The cloud exists as a quantum mechanical state that is precisely and quantitatively described by quantum mechanical wave equations. What matters is the existence of that state, not where any particular electron may be.

Similarly, it is the *state* of a multi-level biological system that matters, not just its breakdown into any particular separate sequences of causation. In any case everything else depends on the existence of the combined state of the system, which is unresolvable into two streams of causation. Not only would there not be two separate streams of causation, what is happening would not be evident in a single slice in time. The attractor or any other organizational property would only be apparent in a phase space representation within which the organizational pattern can be appreciated in an extended time period.

Purely reductionist thinking tends to avoid such language, which is usually criticized as being somehow fuzzy. But it is no more so than quantum mechanics. The analogy is quite close, since the breakdown of an attractor state can be viewed in much the same way as the collapse of a QM wave function. The same criterion for success is also applicable: is the resulting theory empirically predictive? Multi-scale physiological modeling is increasingly successful by this criterion. Vecchi et al. (2018) have introduced the term Entangled Causation to represent their conclusion that "there is no biological rationale for assuming that every switch point should be regulated by a single causal factor and that development generally involves interactive causation in the form of multiple simultaneously contributing difference-making causes to the regulation of the threshold mechanism at every switch point." The resemblance of their conclusions to ours is clear.

Representing organisms as high-level attractors and similarly organized states therefore corresponds much better to what we know experimentally. Most changes at the level of DNA are buffered by the high-level attractors. As Baverstock and Rönkko have shown, the phenotype can best be "represented by high dimensional attractors, evolutionarily conditioned for stability and robustness" (Annila and Baverstock, 2014; Baverstock and Rönkkö, 2014).

Further Physiological examples

We have already used a specific example, that of muscular exercise, to illustrate some of the main points of this article. We will now give further physiological examples. These will illustrate the variety of the forms of boundaries in physiology. It will be through understanding this variety that we will be able to summarize some general principles in See Sections "Delayed differential equations" and "Boundaries between levels: how do they differ?"

Anatomical and functional boundaries in the heart. The heart as an organ has many anatomical boundaries within it since the cells from the sinus node, the atrium, the AV node, the ventricular conducting system, and the ventricle all have different electrophysiological properties, which reflect different protein expression patterns. These in turn are susceptible to different dynamic states within the regulatory networks. The anatomical boundaries between these parts of the heart will therefore experience different magnitudes and direction of ion current flow between them.

These differences also occur within each area. Ventricular cells, for example, differ between epicardial cells and endocardial cells and between the base and apex of the ventricle. These

[2]In any programming of the integration procedure the precise algorithm used depends on the integration formula used. Usually this consists in successive iterations until a preset level of accuracy is achieved. It would not make sense to divide the integration step up into parts. The step itself is just an approximation to an infinitesimally small step. From the viewpoint of this article everything computed in each step can be regarded as an approximation to true simultaneity.

differences are very important in the interpretation of the electrocardiogram. Cells within the sinus node also differ in a graded way. Cells from the periphery have a higher natural frequency than cells near the center.

These differences led to a surprising result when multicellular models of the sinus node became possible, as a result of the increase in computer power offered by the first parallel computers in the 1990s. Using a 64,000 parallel array with each computer processor representing a single cell model, it was found that the origin of the heartbeat, defined as the first cells to depolarize, occurred at the periphery of the model node, creating a wave that spreads inward toward the center (Winslow et al., 1993). This is surprising since in a real heart the beat originates near the center and spreads outward toward the periphery.

The solution to this puzzle was given by the experimental work of Boyett et al. (2003). When the sinus node is carefully separated from the atrium by surgical dissection, the node does indeed behave like the computer model. The sinus-node/atrium boundary is therefore functionally important in creating the conditions in which the beat begins toward the center of the SA node. The high negative resting potentials of the atrial cells together with their high membrane conductance due to high expression of inwardly rectifying potassium channels create the functionality of the complete structure.

Furthermore, the shape of the boundary involved here is not a simple circle or ellipse. The regions of atrial and sinus cells interdigitate in a pattern that enables the weak sinus cells to succeed in depolarizing the stronger atrial cells by almost entirely surrounding cells at the tips of the interdigitations. The impedance-matching process at this boundary is critical in enabling the SA node signal to succeed in spreading through every part and so exciting the whole heart in a functionally important sequence. This functionality is clearly constrained by the high-level geometric structures (Boyett et al., 2003).

Intercellular potassium waves generate oscillatory growth patterns in bacterial films and in vertebrate circulations. Not all bacteria are free swimming single cell organisms. Many form multicellular colonies in the form of films, strings and various matted structures. In their patterns of growth these colonies can behave as intercommunicating networks resembling those of multicellular organisms. Thus, a bacterial film may not grow at a constant speed. It may instead display oscillations in growth rate. These oscillations have been shown to be produced by communications between the cells involving intercellular potassium waves. In effect the cells at the center of the colony are informing those at the periphery when to divide since the release of potassium ions is linked to metabolic activity which in turn enables division to occur (Prindle et al., 2015). Prindle et al. (2015) conclude: "The ensuing "bucket brigade" of potassium release allows cells to rapidly communicate their metabolic state, taking advantage of a link between membrane potential and metabolic activity. This form of electrical communication can thus enhance the previously described long-range metabolic co-dependence in biofilms" (Liu et al., 2015).

Intercellular communication is widespread even in nominally single cell organisms. Potassium wave communication occurs in many organisms, particularly in the circulation in vertebrates, where it is responsible for functionally important phenomena like retrograde vasodilation (Longden et al., 2017). The evolutionary origin of such communication between cells and tissues is clearly very ancient.

Such boundaries can be maladaptive. In the brain, the phenomenon known as spreading depression is due to the generation of a wave of potassium efflux arising principally from glial cells that leads to the depolarization of neurons, resulting in their refractoriness to the nerve impulse with consequent loss of neural activity.

In such forms of communication, the boundaries are fuzzy and distributed. What is a component from some levels may be a boundary at others. Functional boundaries can come and go according to the state of the whole system. Boundaries are themselves therefore interactive. Thus, in the life history of Amoeba *Dictostylium* (?), intercellular boundaries exist at some phases of the cycle and not at others since the organism can function either as an integrated well-ordered colony or as single cells or spores.

Cancer formation and suppression. The standard theory of cancer formation is the somatic mutation theory according to which the accumulation of mutations cause some cells to proliferate abnormally to develop the cancerous tissue. A competing theory is the tissue organization field theory which attributes the cause of cancerous development to properties at a tissue rather than cell or genetic level (Soto and Sonnenschein, 2011). This theory locates the main action at the boundaries between individual cells and the state of the surrounding tissue. A key prediction of this explanation of cancer is that cancers may be "normalized" by changing the boundary, i.e., by transplanting the cancerous or precancerous tissue into normal tissue. This has been shown to happen (Mintz and Ilmensee, 1975; McCullough et al., 1997; Maffini et al., 2005; Kasemeier-Kulesa et al., 2008).

Sponges. All multicellular organisms and colonies of unicellular organisms face the problem of the open boundary requiring exchange with the environment. If the cells are packed too close together some will not be able to exchange nutrients and waste rapidly enough. In See Section ("Intercellular Potassium Waves Generate Oscillatory Growth Patterns in Bacterial Films and in Vertebrate Circulations") above we saw that bacterial colonies solve this problem by signaling when parts of the colony experience metabolic stress. Sponges solve this problem in a different way: the organism is structured using collagen forming open networks of spaces through which freshwater or seawater can flow. Water is wafted through the channels by flagella on the lining of cells, so enabling all cells to exchange freely with the environment. This movement of fluid is the sponge's equivalent of a circulation. There is experimental evidence that this slow-moving aqueous boundary enabled the earliest animal sponges to survive in very low oxygen levels and therefore to evolve before the general oxygenation of the environment around 580 million years ago (Mills et al., 2014).

Delayed differential equations

Equations of this form are sometimes used to represent situations in which there is a significant delay in the action of a part or level of the system on its components (Bocharov and Rihan, 2000). These are important because they also show that chaotic behavior can arise from deterministic equations (Ikeda and Matsumoto, 1987). This form of mathematical representation may seem to contradict our earlier claim of simultaneity of upward and downward causation. That this is not so can be understood by noting that such equations represent an *ordinary* differential equation simplification of any real system, where a full representation would require *partial* differential equations in which the delay would be modeled as a diffusion process in space. This more complete representation would then satisfy the simultaneity condition, with the delay being properly computed in time at each point in space. At each point in space there would be no delay.

Boundaries between levels: how do they differ?

Figure 3 shows the original diagram of multi-level causation (Noble, 2006). The downward arrows were drawn as large and as separate arrows to emphasize the importance of downward causation [see also (Tasaki, 2013)]. These are the forms of causation that constrain the lower levels and which are necessary for an organism to be alive.

However, there are two aspects of this diagram that could be misleading.

First, both the upward and downward forms of causation differ in their details as we move between the levels. We have discussed examples of these differences in the present paper. An important difference that we will highlight here is the difference between the downward forms of causation onto the genome. The arrow between *Protein and DNA Networks* and *Genes* (the smaller left downward arrow) will consist of molecular details concerning the set of transcription factors, regulatory RNAs and methylation by which molecular events at the network level control gene expression. The higher level causation of the same process (right downward arrow) will include properties at the highest levels of the organism that would enable these controls of the genome to be understood functionally, for example why some cells are constrained to produce the patterns of expression for bone cells while others are constrained to become heart cells, albeit from the same genome. Comparable differences occur between the upward arrows. The arrow from *Genes* to *Proteins and RNAs* consists in the transcription and translation machinery of cells. That between *Cells* and *Tissues* consists in the processes that bind cells together to form tissues. The causation at the different levels depends on all the other forms of causation between lower and higher levels. There is a form of nesting of causation, both upward and downward.

Second, as we have already shown, it would be a mistake to think of the upward and downward causations between any levels as sequential, with one occurring before the other. The lesson we learn from representing these forms of causation in mathematical models is that they are necessarily simultaneous.

Figure 4 gives a different representation in which double-headed arrows are used on the left to indicate the simultaneity of action between the different levels. Yet it is still formally correct to say that each of these consists of different kinds of causation. Some will be stochastic, others are ordered constraints. We can therefore imagine these as formally separate lines, as illustrated on the right hand diagram.

The brown colored arrow between DNA and the level of proteins and RNAs is special. The upward influence is a kind of template: genes as DNA sequences act as a template for amino acid sequences in proteins. The downward influences are twofold:

Normal. Influence on expression levels of proteins and RNAs with no change in DNA sequence.

Special. Creation of new DNA by, e.g., the immune system, and other forms of targeted mutations and natural genetic engineering.

Boundaries beyond the organism

Figure 4 also illustrates the fact that, since organisms are open systems, there are necessarily levels above that of the whole organism, extending into the various forms of social interactions and, in the case of humans, the constraints of laws and ethics. Here we simply note that they also introduce different forms of causation, including constraints on behavior exerted by reasons and habits. The blue arrow at the top therefore represents the very different forms of causation that depend on reasons and contextual logic The relations and distinctions between reasons and causes are deep philosophical issues which we do not deal with here. This is part of the reason why we have represented the social and cultural factors involved all together as a single cloud. The diagram does not imply fuzziness or "ghostliness" in the actions on organisms. On the contrary, there is nothing ghostly about the fact that choice of lifestyle affected the muscles of the identical twins in **Figure 2** so differently, nor in the fact that Bathgate et al. (2018) have now identified the specific RNA changes involved at the molecular level.

This is a suitable point to comment on Craver and Bechtel (2007) case against the use of "causation" in top-down influences. Their case is that "the notion of top-down causation is incoherent or that it involves spooky forces exerted by wholes upon their components." We see nothing incoherent in the expression of top-down influences in terms of boundary and initial conditions. Open systems necessarily have boundaries. The forms of causation across those boundaries differ in the two directions, as we have shown and acknowledged throughout this article, but they are nonetheless real. Both forms are mathematically rigorous. As differential equation models show, they are both also necessary. An important clue to the substantial difference between our viewpoints is their statement that "both phrases describe *mechanistically mediated effects*" (their emphasis). We agree that setting boundary conditions is not "mechanistic" in the same sense as the dynamic role of upward causation represented in the differential terms in model equations. Moreover, processes that harness stochasticity are not well represented by the term "mechanistic." It is precisely their non-mechanistic nature that is important.

We are not the first to draw attention to the fact that the causal effects of organization at higher levels are exerted through

the boundary conditions at lower levels. The physical chemist Michael Polanyi made exactly this point as long ago as 1968 (Polanyi, 1968):

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

Polanyi's article is remarkably close to our use of differential equation models to illustrate the different forms of causation in multi-level interactions. The only aspect of his work that has dated is his complete acceptance of Watson and Crick's Central Dogma. He wrote "the morphogenetic process is explained in principle by the transmission of information stored in DNA." He did not know that organisms can influence DNA sequences (the downward aspect of the brown arrow in **Figure 4**) and that much more than DNA is involved in the morphogenetic process.

It is difficult to represent all of these important theoretical distinctions in a single diagram. **Figures 1**, **3**, **4** in our article should therefore each be taken as partial guides to understanding. They each have their limits in representing the conceptual distinctions we are making.

DISCUSSION

The Questions in Our Title: What, Where and When Are Boundaries?

What?

Our paper shows that there are many kinds of boundaries in and around living organisms. Furthermore they are not usually, or ever, passive. They are an essential ingredient of functionality. The reason is that organisms are open systems, operating far from equilibrium. Boundaries are where many of those non-equilibrium processes take place. We cannot therefore understand the behavior of organisms or their parts from their composition alone, and certainly not from the genome alone. The consequences for physiological research are profound. Isolated components of organisms, whether molecules, cells, tissues or organs, do not necessarily behave in the same was as those components *in situ*. This fact is evident even at the molecular level. Proteins, for example, assume different forms in different environments (Balchin et al., 2016) and so do the processes in which they take part (Garcia-Contreras et al., 2012).

Where?

In answering this question we need to remember that it is we who decide what to study in physiological research, whether whole organisms or their components. The way in which we divide nature up determines where the boundaries lie in modeling systems. Where a boundary exists therefore depends on our choice (see the example of the DiFrancesco-Noble equations cited in the Introduction). These choices are not arbitrary, they depend on what has already been discovered. As an example, before the discovery of the variety of epigenetic controls of the genome, the idea of a boundary between the genome and its control by cellular and higher level processes would not have been conceivable. The discovery of these processes and the relevant boundary has far-reaching consequences for physiological research, including interpretations of the Central Dogma of Molecular Biology and of the Weismann Barrier (Noble, 2018).

Choice of boundary also plays a major role in the way in which multi-scale physiology discovers the relative importance of different molecular components. Examples in this article include how the extensions of heart muscle modeling in the 1980s led to the discovery of the quantitative importance of the sodium-calcium exchanger, and how the importance of this exchanger and its regulation has now been discovered using a similar shift from cell to tissue level modeling in skeletal muscle.

When?

Organisms develop, so many boundaries do not exist in the same way at the earliest, single cell, stages. Furthermore, they may differ in their ingredients from system to system even though achieving similar objectives. Boundaries between levels can obviously only arise when those levels develop.

Clarifications of the Principle of Biological Relativity

Our article clarifies several aspects of the Principle of Biological Relativity.

(1) The forms of causation involved in downward and upward causation are fundamentally different. Downward causation consists in constraints exerted by higher levels on the initial and boundary conditions within which the dynamics of lower level elements operate. By contrast, upward causation is the way in which those dynamics influence higher level states.
(2) These two forms of causation do not form a temporal sequence. They occur simultaneously.
(3) It is the state of organization of a higher level that can constrain lower levels. Causation by a state means that it does not make sense to separate out causation by any one element of the state.
(4) Conditioned causation exists in attractors since any perturbation of the state will be resisted. The strength of an attractor can be measured by the speed with which it re-establishes itself (Kaneko, 1998). The strength of downward causation in organisms is generally high since organisms are very effective at resisting changes in phenotype in response to changes at the molecular level, including changes in DNA sequences. Some authors describe conditioned causation as entangled causation (Vecchi et al., 2018). This is a term borrowed from quantum mechanical theory. The analogy is correct to the extent that the causal states involved should not be separated and the entanglement involved resembles that in quantum mechanical states. But there is also an important difference, which is that entangled states in quantum mechanics are very fragile, collapsing in a fraction of a second, whereas the attractor states in biology are often very robust.

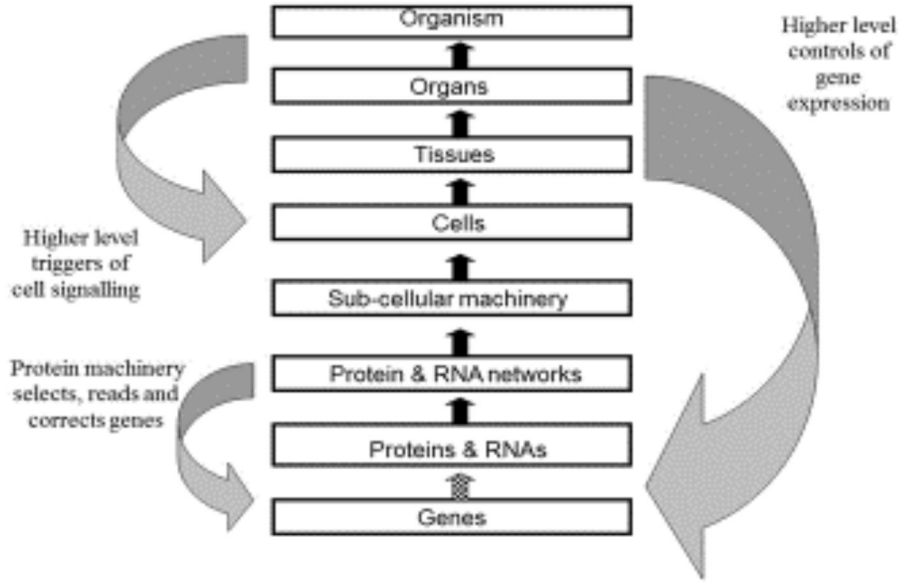


FIGURE 3 | Original multi-level causation diagram illustrating some of the forms of downward causation. Redrawn from Noble (2006), **Figure 2**.

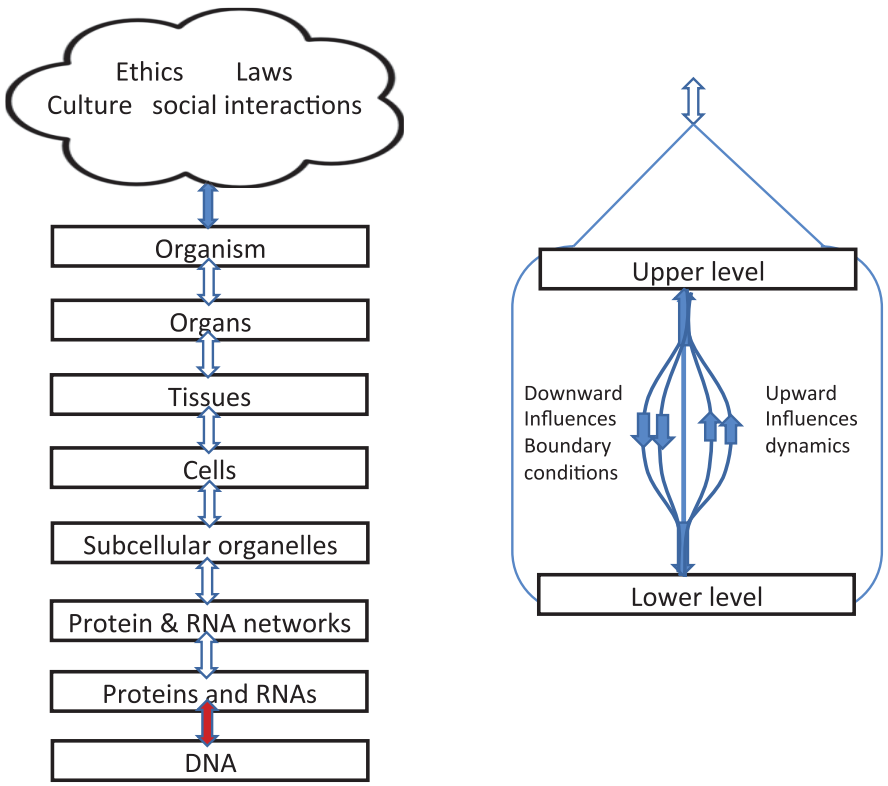


FIGURE 4 | Left: Representation of levels of interaction emphasizing that upward and downward causation operate simultaneously and are shown as double arrows. Right: diagram showing that, within each bidirectional causal arrow, there are different forms of causation, up and down.

Consequences for the Foundations of Physiology

(5) By clarifying the principle of biological relativity, and the nature of the boundaries, multi-level physiology gains rigor. We have not used specific mathematics in this article, nor are many of the points we have discussed primarily mathematical. They are points about the fundamentals of *physiology*. Expressing those fundamentals in terms of arguments drawn from mathematics simply shows that they can, in principle, be as rigorous as any form of science.

(6) What have we not explained? We believe our article opens up many further questions concerning the nature of multi-level physiology. In See Section "Boundaries Beyond the Organism" we have drawn attention to the fact that the causal relations between different levels differ in important ways. One of the most important of these is the increasing role of logic and reasons as we move up to and beyond the level of the whole organism. This is one of the most intractable problems in philosophy and clearly requires more research.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We thank George Ellis, Michael Joyner, and Juan Pascual, for valuable comments on the manuscript.

REFERENCES

Ahmetov, I. I., and Fedotovskaya, O. N. (2015). Current progress in sports genomics. *Adv. Clin. Chem.* 70, 247–314. doi: 10.1016/bs.acc.2015.03.003

Annila, A., and Baverstock, K. (2014). Genes without prominence: a reappraisal of the foundations of biology. *J. R. Soc. Interface* 11:20131017. doi: 10.1098/rsif.2013.1017

Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2016). In vivo aspects of protein folding and quality control. *Science* 353:aac4354. doi: 10.1126/science.aac4354

Bathgate, K. E., Bagley, J. R., Jo, E., Talmadge, R. J., Tobias, I. S., Brown, L. E., et al. (2018). Muscle health and performance in monozygotic twins with 30 years of discordant exercise habits. *Eur. J. Appl. Physiol.* 118, 2097–2110. doi: 10.1007/s00421-018-3943-7

Baverstock, K., and Rönkkö, M. (2014). The evolutionary origin of form and function. *J. Physiol.* 592, 2261–2265. doi: 10.1113/jphysiol.2014.271775

Becker, A. (2018). What is real?: the unfinished quest for the meaning of quantum physics. *Am. J. Phys.* 86:957.

Bernard, C. (1878). *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*. Paris: Baillière.

Bernard, C. (1984). *Introduction à l'étude de la médecine expérimentale 1865*. Paris: Flammarion.

Bocharov, A. B., and Rihan, F. A. (2000). Numerical modelling in biosciences using delay differential equations. *J. Comput. Appl. Math.* 125, 183–119.

Boyett, M. R., Dobrzynski, H., Lancaster, M. K., Jones, S. A., Honjo, H., Kodama, I., et al. (2003). Sophisticated architecture is required for the sinoatrial node to perform its normal pacemaker function. *J. Cardiovasc. Electrophysiol.* 14, 104–106.

Boyle, E. A., Li, Y., and Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186. doi: 10.1016/j.cell.2017.05.038

Buiatti, M., and Longo, G. (2013). Randomness and multilevel interactions in biology. *Theory Biosci.* 132, 139–158. doi: 10.1007/s12064-013-0179-2

Cannon, W. B. (1932). *The Wisdom of the Body*. New York, NY: W W Norton & Co.

Comfort, N. (2018). Genetic determinism rides again. *Nature* 561, 461–463. doi: 10.1038/d41586-018-06784-5

Craver, C. F., and Bechtel, W. (2007). Top-Down causation without top-down causes. *Biol. Philos.* 22, 547–563.

DiFrancesco, D., and Noble, D. (1985). A model of cardiac electrical activity incorporating ionic pumps and concentration changes. *Philos. Trans. R. Soc. B Biol. Sci.* 307, 353–398.

D'Souza, A., Pearman, C. M., Wang, Y., Nakao, S., Logantha, S. J. R. J., Cox, C., et al. (2017). Targeting miR-423-5p reverses exercise training-induced HCN4 channel remodelling and sinus bradycardia. *Circ. Res.* 121, 1058–1068. doi: 10.1161/CIRCRESAHA.117.311607

El-Hani, C., and Queiroz, J. (2005). Downward determination. *Abstracta* 1, 1621–1692.

Galardini, M., Busby, B. P., Vieitez, C., Dunham, A. S., Typas, A., Beltrao, P., et al. (2018). The impact of the genetic background on gene deletion phenotypes in Saccharomyces cerevisiae. *bioRxiv* [Preprint]. doi: 10.1101/487439 (accessed February 12, 2019).

Garcia-Contreras, R., Vos, P., Westerhoff, H. V., and Boogerd, F. C. (2012). Why in vivo may not equal in vitro – new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions. *FEBS J.* 279, 4145–4159.

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

Ikeda, K., and Matsumoto, K. (1987). High-dimensional chaotic behavior in systems with time-delayed feedback. *Physica D* 29, 223–235.

Joyner, M. J., and Coyle, E. F. (2008). Endurance exercise performance: the physiology of champions. *J. Physiol.* 586, 35–44.

Kaneko, K. (1998). On the strength of attractors in a high-dimensional system: milnor attractor network. Robust global attraction, and noise-induced selection. *Physica D.* 124, 332–344.

Kasemeier-Kulesa, J. C., Teddy, J. M., Postovit, L. M., Seftor, E. A., Seftor, R. E., Hendrix, M. J., et al. (2008). Reprogramming multipotent tumor cells with the embryonic neural crest microenvironment. *Dev. Dyn.* 237, 2657–2666. doi: 10.1002/dvdy.21613

Keul, J., Dickhuth, H. H., Simon, G., and Lehmann, M. (1981). Effect of static and dynamic exercise on heart volume, contractility, and left ventricular dimensions. *Circ. Res.* 48, 1162–1170.

Liu, J., Prindle, A., Humphries, J., Gabalda-Sagarra, M., Asally, M., Lee, D. Y., et al. (2015). Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature* 523, 550–554. doi: 10.1038/nature14660

Longden, T. A., Dabertrand, F., Koide, M., Gonzales, A. L., Tykocki, N. R., Brayden, J. E., et al. (2017). Capillary K+-sensing initiates retrograde hyperpolarization to locally increase cerebral blood flow. *Nat. Neurosci.* 20, 717–726. doi: 10.1038/nn.4533

Maffini, M. V., Calabro, J. M., Soto, A. M., and Sonnenschein, C. (2005). Stromal regulation of neoplastic development: age-dependent normalization of neoplastic mammary cells by mammary stroma. *Am. J. Pathol.* 67, 1405–1410.

McAllister, R. E., Noble, D., and Tsien, R. W. (1975). Reconstruction of the electrical activity of cardiac Purkinje fibres. *J. Physiol.* 251, 1–59.

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

McCullough, A. R., Coleman, W. B., Smith, G. J., and Grisham, J. W. (1997). Age-dependent induction of hepatic tumor regression by the tissue microenvironment after transplantation of noplastically transformed rat liver epithelial cells into the liver. *Cancer Res.* 57, 1807–1873.

Mills, D. B., Ward, L. M., Jones, C., Sweeten, B., Forth, M., Treusch, A. H., et al. (2014). Oxygen requirements of the earliest animals. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4168–4172. doi: 10.1073/pnas.1400547111

Mintz, B., and Ilmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc. Natl. Acad. Sci. U.S.A.* 72, 3583–3589.

Mossio, M., Bich, L., and Moreno, A. E. (2013). Emergence, closure and inter-level causation in biological systems. *Erkenntnis* 78, 153–178. doi: 10.1007/s10670-013-9507-7

Natarajan, C., Hoffmann, F. G., Weber, R. E., Fago, A., Witt, C. C., Storz, J. F., et al. (2016). Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* 354, 336–339.

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

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

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus* 2, 55–64. doi: 10.1098/rsfs.2011.0067

Noble, D. (2013). Claude Bernard: un Precurseur de la biologe systemique? in *Claude Bernard. La methode de la physiologie, F. Duchesneau*, eds J.-J. Kupiec, and M. Morange Paris: Editions Rue d'Ulm, 105–114.

Noble, D. (2016). *Dance to the Tune of Life. Biological Relativity*. Cambridge: Cambridge University Press.

Noble, D. (2018). Central dogma or central debate? *Physiology* 33, 246–249. doi: 10.1152/physiol.00017.2018

Noble, D., Chen, Z., Werner, E., and Auffray, C. (2012). *The Selected Papers of Professor Denis Noble CBE FRS. A Journey in Physiology toward Enlightenment*. London: Imperial College Press.

Noble, R., and Noble, D. (2017). Was the watchmaker blind? Or was she one-eyed? *Biology* 6:47. doi: 10.3390/biology6040047

Noble, R., and Noble, D. (2018). Harnessing stochasticity. how organisms make choices. *Chaos* 28:106309. doi: 10.1063/1.5039668

Noble, R., and Noble, D. (2019). "A-mergence," in *The Routledge Handbook of Emergence*, eds S. Gibb and R. Hendry (London: Routledge), 387–399.

Plomin, R. (2018). *Blueprint: How DNA Makes us Who we are*. Hartland: Allen Lane.

Polanyi, M. (1968). Life's irreducible structure. *Science* 160, 1308–1312.

Prindle, A., Liu, J., Asally, M., Ly, S., Garcia-Ojalvo, J., Süel, G. M., et al. (2015). Ion channels enable electrical communication in bacterial communities. *Nature* 527, 59–63. doi: 10.1038/nature15709

Rankinen, T., Fuku, N., Wolfarth, B., Wang, G., Sarzynski, M. A., Alexeev, D. G., et al. (2016). No evidence of a common dna variant profile specific to world class endurance athletes. *PLoS One* 11:e0147330. doi: 10.1371/journal.pone.0147330

Rankinen, T., Pérusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., Bouchard, C., et al. (2005). The human gene map for performance and health-related fitness phenotypes: the update. *Med. Sci. Sports Exerc.* 38, 1863–1888.

Rankinen, T., Wolfarth, B., Simoneau, J. A., Maier-Lenz, D., Rauramaa, R., Rivera, M. A., et al. (2000). No association between the angiotensin-converting enzyme ID polymorphism and elite endurance athlete status. *J. Appl. Physiol.* 88, 1571–1575.

Rosen, R. (1991). *Life Itself. A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. New York, NY: Columbia University Press.

Rosen, R. (2000). *Essays on Life Itself*. New York, NY: Columbia University Press.

Shorten, P. R., O'Callaghan, P., Davidson, J. B., and Soboleva, T. K. (2007). A mathematical model of fatigue in skeletal muscle force contraction. *J. Muscle Res. Cell Motili.* 28, 293–313.

Soto, A. M., and Sonnenschein, C. (2011). The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. *Bioessays* 33, 332–340. doi: 10.1002/bies.201100025

Sutton, J. R., Cole, A., Gunning, J., Hickie, J. B., and Seldon, W. A. (1967). Control of heart-rate in healthy young men. *Lancet* 1967, 1398–1400.

Tasaki, K. M. (2013). Circular causality in integrative multi-scale systems biology and its interaction with traditional medicine. *Prog. Biophys. Mol. Biol.* 111, 144–146. doi: 10.1016/j.pbiomolbio.2012.09.005

Vecchi, D., Miquel, P.-A., and Hernandez, I. (2018). From biological determination to entangled causation. *Acta Biotheor.* 67, 19–46. doi: 10.1007/s10441-018-

Waddington, C. H. (1957). *The Strategy of the Genes*. London: Allen and Unwin.

Winslow, R. L., Kimball, A. L., Varghese, A., and Noble, D. (1993). Simulating cardiac sinus and atrial network dynamics on the connection machine. *Physica D* 64, 281–298. doi: 10.1016/0167-2789(93)90260-8

Yan, X., Papadimitriou, I., Lidor, R., and Eynon, N. (2016). Nature versus nurture in determining athletic ability. *Med. Sport Sci.* 61, 15–28. doi: 10.1159/000445238

Yang, J., Clark, J. W. Jr., Bryan, R. M., and Robertson, C. (2003). The myogenic response in isolated rat cerebrovascular arteries: smooth muscle cell model. *Med. Eng. Phys.* 25, 691–709.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

In press. Handbook of the Philosophy of Emergence (Routledge 2019)

A-mergence of biological systems

Raymond Noble
Institute for Women's Health,
University College London,
Gower Street, London, UK

Denis Noble
Department of Physiology, Anatomy & Genetics
Sherrington Building, Parks Road
Oxford OX1 3PT, UK

Summary

We argue that (1) emergent phenomena are real and important, (2) for many of these, causality in their development and maintenance is necessarily circular, (3) the circularity occurs between levels of organization, (4) although the forms of causation can be different at different levels, there is no privileged level of causation a priori: the forms and roles of causation are open to experimental investigation, (5) the upward and downward forms of causation do not occur in sequence, they occur in parallel, i.e. simultaneously, (6) there is therefore no privileged direction of emergence, the upper levels constrain the events at the lower levels just as much as the lower levels are necessary for those upper level constraints to exist, (7) to emphasise this point, we introduce the concept of a-mergence, which expresses the lack of causal directionality. We illustrate these points with a major test case: Schrödinger's distinction between physics and biology in which he proposed that physics is the generation of order from molecular disorder, while biology is the generation of order from molecular order. This characterization of biology is physically impossible. Modern biology has confirmed both that this is impossible and that, on the contrary, organisms harness stochasticity at low levels to generate their functionality. This example shows in fine detail why higher level causality can, in many cases, be seen to be more important than lower level processes. The chapter highlights a number of further examples where a-mergence seems to be a more appropriate way of describing what is happening than emergence.

(1) *Emergent phenomena are real and important*

Biological reductionism can be seen to have originated with Descartes in the seventeenth century, while relying heavily on Newtonian mechanics later in the century, and in later centuries on the mathematical genius of Pierre-Simon Laplace. Descartes laid the foundation by arguing that animals could be regarded as machines in some way comparable to the ingenious hydrostatic robots that had become popular amongst the aristocracy in their gardens. Newtonian mechanics cemented the foundation with the laws of mechanical motion, and Laplace systematised the ideas with his famous statement that a supreme intelligence could use mathematics to predict the future completely, and retrodict the past as well. Everything that has or will happen would be clear to such a being. Descartes even foresaw one of the central ideas of Neo-Darwinism:

> "If one had a proper knowledge of all the parts of the semen of some species of animal in particular, for example of man, one might be able to deduce the whole form and configuration of each of its members from this alone, by means of entirely mathematical and certain arguments, the complete figure and the conformation of its members." (*On the formation of the fetus*)[1]

which is essentially the idea that there is a complete mathematical 'program' there in the semen, prefiguring Jacob and Monod's 'genetic program'. Complete because he writes "from this alone". The causation, on this view, is entirely one way.

It is therefore significant that the first clear statement of the opposite view can be traced back to Descartes' main philosophical opponent. In 1665, just two years after the foundation of The Royal Society, Benedict de Spinoza, working in Holland, was in extensive correspondence with the first Secretary of that Society, Henry Oldenburg, working in London.

Oldenburg had just returned from meeting Spinoza in Holland and had been fascinated by his discussions with him on "the principles of the Cartesian and Baconian philosophies". Spinoza was opposed to the dualism of mind and body espoused by Descartes. This was necessary in Descartes' view of animals as automata since he wished to exclude humans from this view and so attributed their free will to a separate substance, the soul, which could interact with the body. Spinoza was in the process of seeking to publish his great work (The Ethics: *Ethica ordine geometrico demonstrata*) in which he proposes an alternative philosophy. Spinoza did not publish in *Philosophical Transactions*, but this correspondence includes an important letter from Spinoza which could form a text for the systems approach and the concept of Biological Relativity (Noble 2012, Noble 2016). The original letter in Latin is still kept in the Royal Society library. He writes: "every part of nature agrees with the whole, and is associated with all other parts" and "by the association of parts, then, I merely mean that the laws or nature of one part adapt themselves to the laws or nature

[1] The French text reads « Si on connoissoit quelles sont toutes les parties de la semence de quelque espece d'Animal en particulier, par exemple de l'homme, on pourroit déduire de la seul, par des raisons entierement Mathematiques et certaines, toute la figure & conformation de ses membres ; » (de la formation du fœtus, para LXVI p 146)

of another part, so as to cause the least possible inconsistency." He realised therefore some of the problems faced in trying to understand what, today, we would characterise as an open system. An open system is one that freely exchanges energy and matter with its surroundings. By definition, in a closed system each part must be influenced only by rules governing the behaviour of the parts within it. If those parts behave deterministically then the whole must also do so. But when parts of wholes are considered as sub-systems, they are necessarily open in the context of the whole. As we will explain in diagram 1 below, even the equations used to describe the behavior of parts require initial and boundary conditions provided from outside the system. Thus, biological systems are open in relation to their environment.

Spinoza therefore appreciated the difficulty in working from knowledge of minute components to an understanding of the whole:

> "Let us imagine, with your permission, a little worm, living in the blood, able to distinguish by sight the particles of blood, lymph etc, and to reflect on the manner in which each particle, on meeting with another particle, either is repulsed, or communicates a portion of its own motion. This little worm would live in the blood, in the same way as we live in a part of the universe, and would consider each particle of blood, not as a part, but as a whole. He would be unable to determine, how all the parts are modified by the general nature of blood, and are compelled by it to adapt themselves, so as to stand in a fixed relation to one another"[2]

This paragraph could stand even today as a succinct statement of one of the main ideas of Biological Relativity. He doesn't use a mathematical medium to express his idea, but this could be so expressed as the aim to understand how the initial and boundary conditions of a system constrain the parts to produce a particular solution to the differential equations describing their motions. We need then to move to the complete system (with whatever boundary we choose to use to define that) in order even to understand the behavior of the parts.

The essence of Spinoza's argument, to use modern language, is that organisms are open systems. This must be so since they can survive only by exchanging matter and energy with their environment. If an organism, or a part of an organism, is treated as a closed system by experimentally preventing those exchanges, it will become dead. The great majority of biochemical and molecular biological experiments are performed on dead and dying organisms, or their parts, such as cells and molecules. To understand how they operate as complete organisms it is completely necessary to take into account the exchanges of matter and energy with their environment. It is through those interactions that organisms can be alive.

---

[2] The Latin text is « Concipiamus iam, si placet, vermiculum in hoc fluido, nempe in sanguine, vivere, qui visu ad discernendas particulas lymphae et chyli etc. valeret, et ratione ad observandum, quomodo unaquaequae particula ex alterius occursu vel resilit, vel partem sui motus alteri communicat. Ille quidem in sanguine, ut nos in hoc universi parte viveret, et unamquamque sanguinis particulam ut totum non vero ut partem consideraret nec scire posset, quomodo partes omnes ab universali natura sanguinis moderantur, et invicem prout universalis natura sanguinis exigit accomodari coguntur ut certa ratione inter se consentiant. »

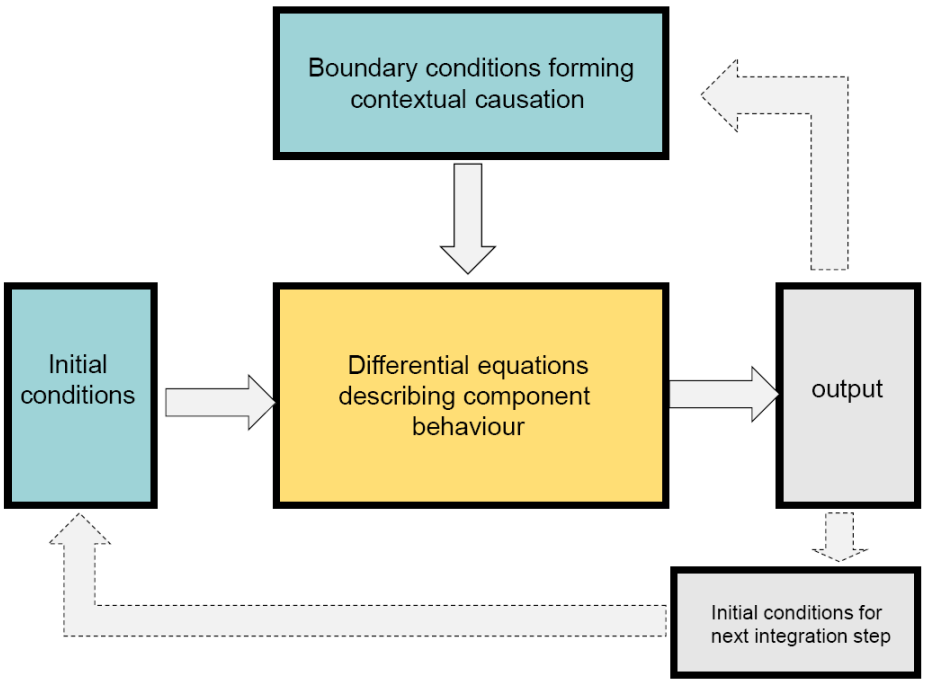


Figure 1 Diagram of causal sequences involved in integrating differential equation models. Description in text. (from Noble(Noble 2012)).

(2) *Causality in the development and maintenance of emergent processes is necessarily circular*

Since the environmental influences arise from a higher level, circular causality must occur, downwards as well as upwards. 'Down' and 'up' here are metaphors and should be treated carefully. The essential point is the more neutral statement: there is no privileged scale of causality, which is the a priori principle of Biological Relativity. One of the consequences of the relativistic view is that genes, defined as DNA sequences, cease to be represented as active causes. They are templates and are passive causes, used when needed to make more proteins or RNAs. Active causation resides in the networks which include many components for which there are no DNA templates. It is the interactive relationships of those dynamic networks which determine what happens. No single component or single mechanism can do so.

This view of organisms can be formalized mathematically as shown in Figure 1. 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.

Several important conclusions follow from this analysis. First, the equations used in modeling biology cannot even be solved if we do not specify the boundary and initial conditions. Second, those conditions necessarily require causal information about the environment of the system we are modeling. Third, this conclusion is true irrespective of whether we consider the world to be determinate. Even a Laplacian determinist would have to accept this. Recall that Spinoza also was a determinist. We can of course introduce stochasticity into the modeling to produce a non-determinate model. In fact this is necessary to formulate the complete principles of biological relativity (Noble 2016)(chapter 6), but this does not change the essential need for input from the environment of any open system.

(3) *The circularity occurs between levels of organization*

Consider as a concrete example the regularity of the normal heartbeat and how it is disturbed in life-threatening arrhythmias. The normal heartbeat is an attractor caused by a circular form of causality in which both the cell potential and the individual proteins are entrained by their interaction. Once the rhythm begins it can continue indefinitely. Even large perturbations in the individual proteins or their genes can be resisted (Noble 2011). This is precisely what is meant by an attractor. If you represent the parameters as a multidimensional space, there are large volumes within this space representing possible parameter sets, from which the system will automatically move towards the attractor.

Now consider what happens when a different kind of attractor is established. This happens in the heart when abnormal spiral waves of excitation arise at the level of the whole heart. The individual molecules in each cell are now constrained to dance to a different and more chaotic rhythm. Viewed from the level of the individual molecules both of these influences from the higher levels of the cell or the whole organ will seem inexplicable. The molecules are like boats tossed around in a storm beyond their own control. Yet, the storm also depends on their activity. Indeed it can be modeled using the equations for that activity (Carro, Rodríguez et al. 2011). But as explained in the previous section, those equations will necessarily represent the circularity between the causal levels. Each of the three views, the molecular, the cellular and the organ, are valid, but only from the higher levels can we provide a full account of what is happening, including the lower levels whose behavior is in need of explanation. .

(4) *Although the forms of causation can be different at different levels, there is no privileged level of causation a priori: the forms and roles of causation are open to experimental investigation.*

The principle that there is no privileged scale of causality can easily be misinterpreted. It is important now to introduce some clarifications.

First, we must distinguish between its conceptual status and its practical implications. It is an *a priori* statement, i.e. a statement about what we should or should not assume in advance of doing the experiments. We should not assume that causation *necessarily* resides at a particular, e.g. molecular, scale. That is the mistake made by naïve reductionism in biology. The reduction to molecular level events is treated as a

methodological necessity, whereas it should emerge, if at all, from the experiments themselves. Before we do those experiments, we cannot know which parts of a system are involved in its behaviour, nor attribute any privileged position to them.
But that does not mean that all scales must be involved in any given example. The circles of causal networks may span particular ranges of scales, which may be more or less limited in extent. And there may be particular levels that act as important hubs. Those facts are for us to discover as empirical observations. For example, many biologists regard the cell as a central level of integration in much of biology. That conclusion is a result of extensive experiments on cells showing their functional integrity and that many physiological functions cannot be ascribed to entities lower than the cell. Cells contain the main metabolic networks, circadian and various other rhythm networks, cell cycle networks, and so on. Moreover the great majority of living organisms are single cells.

The genome also has a unique position. But it is not the one most often ascribed to it as a program dictating life. As the American cell biochemist Franklin Harold puts it in his book *In Search of Cell History* "The genome is not the cell's command center but a highly privileged databank, something like a recipe or a musical score, yet for the purpose of parsing evolution, genes have a rightful claim to center stage."(Harold 2014). Parsing is the analysis of strings of symbols, usually with guidance from some rules of grammar. In the case of DNA, the start and stop sequences and those for binding transcription factors, amongst other features, provide those guidelines. Analysis of this kind has indeed been exceptionally useful in the inter-species DNA sequence comparisons that now form the basis of much of our understanding of evolutionary history.

The genome sequences are therefore comparable to a formal cause[3] in Aristotle's classification of the forms of causation, while the causation from the networks operating at higher levels than the genome can be regarded as an efficient cause (Noble 2016)(pp. 176-181). The sequences are a formal cause since they form templates to enable ribosomes to construct the proteins specified by those sequences, while those proteins form part of the dynamic networks that form the efficient cause[4] necessary for the attractor to exist. This distinction is particularly clear in the example of cardiac rhythm discussed above. The attractor doesn't even require the involvement of DNA or RNA sequences until the cell requires more proteins to be made. Some other rhythmic attractors do involve DNA sequences in the cycle. Circadian rhythm is a good example. One form of the attractor includes feedback from the level of the protein involved to inhibit the formation of the protein (Hardin, Hall et al. 1990, Foster and Kreitzman 2004). But even in this case, the genome is not completely necessary. So-called 'clock' genes in mice can be knocked out without affecting circadian rhythm (Debruyne, Noton et al. 2006).

[3] A formal cause exists when it is the geometrical arrangement of something that influences the outcome. It is the formal arrangement of nucleotides in sequences that gives the genome the power to determine amino acid sequences in proteins and nucleotide sequences in RNAs.
[4] An efficient cause exists when it is the motion of something that affects the outcome, as in billiard balls colliding. The billiard balls may all have the same form, but their movements are different.

Each feature of organisms at the various levels may therefore have unique causal properties. The principle of Biological Relativity should not be taken to require that all forms of causation involved are equivalent.

(5) *The upward and downward forms of causation do not occur in sequence, they occur in parallel, i.e. simultaneously*

It is important to understand that the processes represented in Figure 1 all occur as a process. It is merely a convenience of representation that the integration step is represented as coming after setting the initial conditions, which then precedes the formation of the output. In a computer program representing the sequence, we do indeed write the code in precisely this sequence. But this is a mathematical fiction arising from the fact that we solve the equations in finite steps. Differential equations themselves do not express finite steps. On the contrary, the differential symbol 'd/dt' represents a vanishingly small step. In reality also, all the processes represented in the equations proceed simultaneously. Our 'difference' equations actually solved by the computer are simply approximations. The test we use for whether they are accurate enough is precisely to reduce the integral step length until the solution converges to an arbitrarily high degree of accuracy. In principle, for infinitely good accuracy, we would have to reduce the step length to zero, which is exactly what differential equations themselves represent. In this respect the equations are better representations of what we are modeling than any particular computer simulation. In the rare cases in non-linear differential equation models where we can solve the equations analytically (Hunter, McNaughton et al. 1975, Jack, Noble et al. 1975), the solution is revealed as a complete solution as a function of time, with no sequence of causation. This is an important reason for which we will introduce the concept of *a*-mergence at the end of this chapter.

(6) *There is therefore no privileged direction of emergence, the upper levels constrain the events at the lower levels just as much as the lower levels are necessary for those upper level constraints to exist.*

It follows that it is simply a matter of convenience that we often talk of the higher level functions arising *from* the interactions of the components. It would be just as correct to say that the constraints on the lower level components arise *from* the existence of the higher level function. Best of all, we should conclude that they necessarily co-arise.

(7) *To emphasise this point, we introduce the concept of a-mergence, which expresses the lack of causal directionality.*

We have developed our argument with examples from cardiac and circadian rhythms. We will now illustrate all these points with a central test case: Schrödinger's distinction between physics and biology in which he proposed that physics is the generation of order from molecular disorder, while biology is the generation of order from molecular order (Schrödinger 1944). This is a central test case because, as both Watson and Crick acknowledged, the formulation of the central dogma of molecular biology was greatly influenced by Schrödinger's ideas. It is also hard to think of a more concrete example where the directionality of causation is widely accepted to be

one way. The Central Dogma has been interpreted to mean that the genome sequences cause the organism but are not themselves affected by the organism. This view has been taken to deny the existence of emergent properties, and it is implicit in versions of evolutionary theory that equate the Central Dogma to the Weismann Barrier or, at least, claim that the Weismann Barrier is now 'embodied by' the Central Dogma. We develop this final section of our argument in four stages.

(a) *It is a mistake to interpret Crick's statement to mean one-way causation*. The relevant statement is:

> "The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that *such information* cannot be transferred back *from protein* to either protein or nucleic acid."(Crick 1970)

We have italicised '*such information'* and '*from protein*' since it is evident that the statement does not say that *no* information can pass from the *organism* to the genome. In fact, it is obvious that it must do so to produce many different patterns of gene expression, which enable many different phenotypes (e.g many different cell types in the same body) to be generated from the same genome.

This information from organisms is conveyed to their genomes by patterns of transcription factors, genome marking, histone marking, and many RNAs, which in turn control the patterns of gene expression. These controls are exerted through preferential targeted binding to the genome or histone proteins. For example, methylation of cytosines preferentially occurs at CpG sites. Binding to histones preferentially occurs at the histone tails. Even though these are the targeted molecular mechanisms by which the functional control is exerted, there is no guarantee that the functionality will be evident at the molecular level. It would require many correlations between the *patterns* of binding and the functional processes at a higher level to identify the functionality involved. Without that correlation the binding patterns will appear random.[5] Yet it is those patterns that control the expression of individual genes. Those patterns are phenotypes, not themselves genotypes. A good example comes from the study of the evolution of hemoglobins in many avian species to adapt to altitude. Natarajan et al show that "predictable convergence in haemoglobin function has unpredictable molecular underpinnings" (Natarajan, Hoffmann et al. 2016).

That first point establishes that the same genotype can be used to create an effectively unlimited number of phenotypes (Feytmans, Noble et al. 2005). That demonstration does not, however, exhaust the role of the phenotype in determining the functioning of the genome. Not only is it true that the same genotype can be used to generate

[5] The stochasticity is therefore epistemological. In principle, once the higher level constraints are known a bottom-up computation becomes conceivable. However, given the effectively unlimited combinations and the associated requirement for unlimited computer power, it is extremely unlikely that such computations could be successful. And they almost certainly could not be performed without the insights provided by the higher level functional information. As we emphasized earlier, the initial and boundary conditions are essential for the computation to be performed.

many different phenotypes, *it is the phenotype that enables it to be the same (or even a different) genotype*.

(b) *It is the phenotype that enables the genome to be the same genotype.*

If correct, this statement would completely reverse the direction of causality assumed in reductionist explanations of living organisms. How, then, was the currently accepted one way genotype —› phenotype explanation ever thought to be correct? The answer lies in Schrödinger's book *What is Life?*(Schrödinger 1944). That book makes one spectacularly correct prediction and a second necessarily incorrect prediction. The correct prediction was that the genetic material would be found to be what he called a non-periodic crystal. Remember that this was in 1942 before it had been shown that genetic information is found in DNA sequences. If one thinks of a linear polymer as a crystal that does not endlessly repeat itself, then non-periodic (or a-periodic) crystal is quite a good description of what molecular biology subsequently discovered. Remember too that the book was written at a time when X-ray crystallography had come into use to 'read' the molecular structure of organic molecules. This enabled Dorothy Hodgkin to determine the structure of cholesterol in 1937, penicillin in 1946, and vitamin B12 in 1956. These were spectacular achievements. What was more natural than to conclude that if the genetic material is a form of crystal it could also be 'read' in a determinate way? That was indeed the conclusion Schrödinger drew in his book.

But he was too good a physicist not to notice, initially at least, that this conclusion was 'ridiculous':

> "We seem to arrive at the ridiculous conclusion that the clue to understanding of life is that it is based on a pure mechanism, a 'clock-work'…."

'Ridiculous', because how could biological molecules not show the extensive stochasticity that would arise from their possession of kinetic energy?[6] That was precisely why he had, earlier in his book, concluded that physics was the generation of order, e.g. the laws of thermodynamics, from disorder, i.e. molecular level stochasticity.

But he had difficulty harmonizing the two insights. Confusingly, he wrote:

> "The conclusion is not ridiculous and is, in my opinion, not entirely wrong, but it has to be taken 'with a very big grain of salt'".

He then explains the 'big grain of salt' by stating that even clock-work is, 'after all statistical' (p. 103).

---

[6] The only way known to modern physics would be for the molecules to form a Bose-Einstein condensate. But molecules can only do this at extremely low temperatures near absolute zero (Whitfield, J. (2003). "Molecules form new state of matter." Nature **doi:10.1038/news031110-16**. Nevertheless, some biologists have speculated along these lines (Ho, M.-w. (2008). The Rainbow and the Worm. The physics of organisms. London, World Scientific Publishing.). Whether or not this happens, it is not needed as an explanation since we already know that the stochasticity is present, even in copying DNA.

Schrödinger clearly realised that something is far from right but was struggling to identify what it might be.

(c) *It is the phenotype that enabled the genome to be a different genotype.*

We would now say that the molecules involved (DNA) *are* subject to frequent statistical variations (copying errors, chemical and radiation damage, etc.), which are then corrected by the cell's protein and lipid machinery that enables DNA to become a highly reproducible molecule. This is a three-stage process that reduces the copy error rate from 1 in 10^4 to around 1 in 10^{10}, which is an astonishing degree of accuracy. In a genome of 3 billion base pairs this works out as less than 1 error in copying a complete genome, compared to millions of errors without error correction. The order at the molecular scale is therefore actually created *by the system as a whole*, including lipid components that are not encoded by DNA sequences. This requires energy, of course, which Schrödinger called negative entropy. Perhaps therefore this is what Schrödinger was struggling towards, but we can only see this clearly in retrospect. He could not have known how much the genetic molecular material experiences stochasticity and is constrained to be highly reproducible *by the organism itself*. The order at the molecular (DNA) level is actually imposed by higher level constraints. If we ever do synthesise from scratch the complete genome of a living organism, we would have to give it this cellular environment in which to function accurately. Otherwise, any genome longer than about 10,000 bases would fail to be preserved reliably at the first copying process.

The Central Dogma was originally formulated by Crick in 1958 (Crick 1958) in a very hard form: DNA —› RNA —› protein. This formulation would have completely protected the genome from alteration of its sequence by the organism. No changes in proteins or their relative expression patterns could conceivably have altered a genome that was isolated in such a way. By 1970 however, the Dogma had to be modified after the discovery of reverse transcriptase (Temin and Mizutani 1970) to become DNA ‹—› RNA —› protein, and even to become:

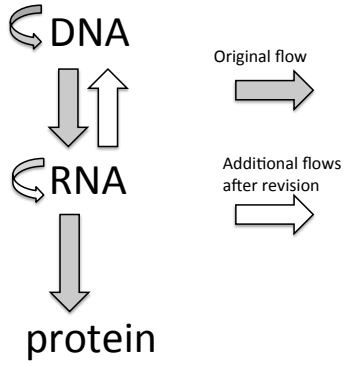


Figure 2. The central Dogma of Molecular Biology

Reverse transcription enables the white upward arrow to occur to allow RNAs to be back-transcribed into DNA, while the upper curved arrow enables DNA sequences to be pasted directly into the genome. The two together completely counter the Central Dogma since they enable sequences of any length to be moved around the genome,

either directly as DNA or indirectly via RNA. Way back in the 1930s and 1940s Barbara McClintock had shown that such transfers do occur in plants in response to environmental stress. This was why, on winning the Nobel prize for mobile genetic elements in 1983, she referred to the genome as a

> "highly sensitive organ of the cell, monitoring genomic activities and correcting common errors, sensing the unusual and unexpected events, and responding to them, often by restructuring the genome." (McClintock 1984)

Did this happen in the evolution of genomes? The answer is yes, it must have done. The evidence comes from the comparative sequencing of genomes from many different species ranging from yeast to man reported in the 2001 *Nature* report on the first full draft of the human genome sequence (International Human Genome Mapping Consortium 2001). The gene sequences for both transcription factor proteins and chromatins show precisely this kind of massive genome re-organisation (Shapiro 2011). Add Shapiro 2014.

This process has also been recorded in real time experiments performed on bacteria evolving in a nutrient medium that does not provide what was an essential metabolite. By periodically allowing conjugation with bacteria that metabolise the new chemical and gradually removing the usual essential metabolite the bacteria succeeded in weaning themselves completely off their usual nutrient. Sequence analysis showed that conjugation had allowed the shuffling of gene domains during the periodic conjugations. Significantly, the authors entitle their article with reference to "directed evolution" (REF Crameri et al 1998)). In a recent article we have shown why this kind of process should be characterised as "directed" since it arises from circular causation that represents a form of organism intelligence (REF). Hosseini et al (2016) have confirmed such findings, which they characterise as "phenotypic innovation through recombination".

(d) *It is the phenotype that <u>enables</u> the genome to be a different genotype*

Notice the small difference in tense compared to statement (c). In this section we ask whether the phenotype can be observed to alter the genome in real time observations on the evolution of cells and organisms.

It is in fact well-known already that cells can harness stochasticity to generate specific function since cells of the immune system show the phenomenon of highly targeted somatic hyper-mutation. Figure 3 summarizes what we know. Faced with a new antigen challenge, the mutation rate in the variable part of the genome can be accelerated by as much as 1 million times. So far as we know, the mutations occur randomly. But the location in the genome is certainly not random. The functionality in this case lies in the precise targeting of the relevant part of the genome. The mechanism is directed, because the binding of the antigen to the antibody itself activates the proliferation process.

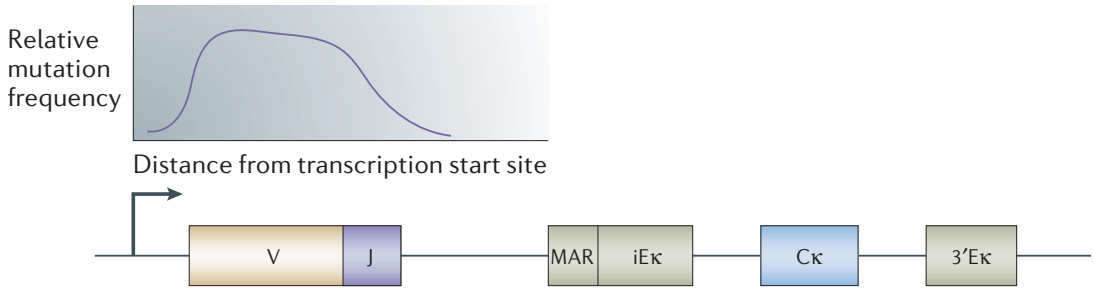


Figure 3. Schematic diagram of gene-specific targeted hyper-mutation in immunoglobulin gene loci. The mutation rate is greatly increased only in the variable part of the genome, which is a ~1.5 kilobase region in each of the three immunoglobulin loci. In this figure, the graph above the rearranged variable (V) and joining (J) gene segments that form the variable region of Igκ depicts the mutation domain in the κ-light chain (Igκ) locus. 3′Eκ, Igκ 3′ enhancer; Cκ, Igκ constant; iEκ, Igκ intronic enhancer; MAR, matrix attachment region (Odegard and Schatz 2006).

This example from the immune system shows that functionally significant targeted hyper-mutation can occur in the lifetime of an individual organism. There is no reason why this kind of mechanism should not be used in evolutionary change, and it is.

A well-known functionally-driven form of genome change is the response to starvation in bacteria. Starvation can increase the targeted reorganizations of the genome by five orders of magnitude, i.e. by a factor of over 100,000. This is one of the mechanisms by which bacteria can evolve very rapidly and in a functional way in response to environmental stress. It would be important to determine whether such targeted reorganization occurs in experiments on conjugating bacteria discussed above. The question would be whether bacteria sensing deprivation trigger higher frequencies of conjugation and shuffling of domains. This “sensing” of the environment, as in the immune system, is precisely what constitutes the feedback necessary for the process to be characterized as directed (see Noble & Noble 2017 for the relevant definitions of agency and directionality in the evolution of organisms and their populations).

A similar targeting of location where genomic change can occur has been found in experiments on genetically modified fruit flies. One of the common ways in which genetic modification is achieved is to use a particular kind of mobile genetic element that can move around the genome using a cut-and-paste mechanism that does not require an RNA intermediate. Most often the insertions occur in a random way. But when DNA sequences from certain regulatory regions are used, they get inserted preferentially near the gene from which the sequence was derived (Bender. W and Hudson 2000). This process targets the changes in a way that is clearly not random with respect to possible function.

There are many more examples in bacterial evolution (Noble 2017).

Conclusions

It will be evident that we have a concern about the word 'emergence'. The 'e' naturally leads to us asking "what emerges *from* what?" Our position makes it obvious that this is often the wrong question. Of course, there may be "emergence" in the sense that, *on a temporal time scale*, atoms emerged from fundamental particles, stars emerged from condensations of matter, life emerged from the formation of suitable planets, and so on. But at each stage a new level of organization takes over. Once an attractor has formed, the components are constrained *by the attractor itself*. The direction of causality then changes. The term 'a-mergence' tries to make that clear. There is no privileged level of causality, in the sense that all levels can be causal. But it is clear from what we have written that the nature of causality changes with the level, and that the higher levels can be said to be directed functionally, and in that sense they are privileged.

Moreover we doubt whether any directionality of causation should be assumed, whether sequential or in parallel. Once an attractor has formed, the best description would be to say that this condition of the system is followed by that condition. There is no need to isolate any components, at any level, as the primary cause. The condition is the a-mergent state and this condition is causative. Moreover, the organization of the state is precisely what defines the level at which it can be said to occur. Thus we refer to atomic, molecular, cellular, tissue, organ, organism, niche, habitat etc, each with a dynamic of causative, functional (goal-directed) organization.

Does the concept of goal-directed processes lead to a better understanding? It is difficult to understand causality without knowing this logic. We understand 'the function' of a thermostat better by understanding that it operates TO maintain temperature within a certain range - it is the logic of the thermostatic system. We understand better the function of baroreceptors in the circulatory system by knowing they are part of a system to maintain pressure within a given range and facilitate blood flow round the body - it is the logic of the system. Knowing or understanding the logic of a system is as important as any reductionist detail about the system. It is an organisational logic - no one part of the system has primacy in that logic. How each part behaves is influenced by its arrangement within the system. It is the situational logic of the system.

The thoughts and discussion we have had in writing this chapter are part of the dynamic organisation at a social level, where action can be identified not only as purposive but also intentional. It is only at this level that behaviour can be described, for example as 'selfish' or 'altruistic', as only at this level can there be reasoned choice, and reasoned logic, or where we can talk of motivation and emotion, hopes, desires, fears and anxieties.

Science by method will hold variables constant to study the effect of changing a given one – clamping a voltage for example. What we know is that this is artificial and establishes an artificially fixed sequence of events. The a-mergent state is in continuous flux, but biological processes maintain such states within functional range. Life as an a-mergent state is in this sense autopoietic, or self-maintaining (Maturana and Varela 1980, Luisi 2016)

The conclusions we have drawn in this chapter are firmly based on experimental findings on the ways in which organisms harness stochasticity to generate functionality. Our reinterpretation of Schrödinger's ideas to produce a conclusion diametrically opposite to the one he himself drew, and which has dominated biological science ever since, is clearly based on experimental findings on the mechanism of reproducibility of DNA copying in whole cells, which could not occur without the integrative activity of cells as a whole. While the cardiac rhythm and targeted hypermutation examples are also firmly based on experimental findings, requiring the integrative action of whole cells.

Our conclusions also strongly support the philosophical approaches developed by, for example, Nancy Cartwright and John Dupré. In his book *The Disorder of Things* ((Dupré 1993), p 101) Dupré writes "…causal completeness at one particular level is wholly incredible. By contrast with even the weakest versions of reductionism, the pluralism I have in mind precludes the privileging of any particular level." This statement accurately reflects the metaphysical position adopted in our work, and its empirical basis. Dupré's work focuses on biology. Cartwright ((Cartwright 1999), p 31) comes to very similar conclusions in her study of causality in physics and economics: "….nature is governed in different domains by different systems of laws not necessarily related to each other in any systematic or uniform way; by a patchwork of laws." This nicely expresses our analysis that the constraints exerted by higher level systems on lower level components depend on the rules being followed by the system, not the highly stochastic behavior of the molecular components.

Modern philosophers of science arrived at these conclusions at least twenty years ago on the basis of careful recognition of the significance of experimental work already achieved at that time. But the silo-isation of disciplines has meant that there has been very little cross fertilization back from these philosophical works and the scientific community. The veritable flood of experimental work now appearing (Noble 2017) that makes the conclusions even more convincing may, we hope, now have its impact in the strategy of experimental biological science. It is high time to escape the limited metaphysical straightjackets of purely gene-centric interpretations.

Acknowledgements. Some parts of the text have been developed from our previous articles and books. We thank Nancy Cartwright and John Dupré for valuable criticisms.

References

Bender. W and A. Hudson (2000). "P element homing to the Drosophila bithorax complex." Development **127**: 3981-3992.
Carro, J., J. Rodríguez, P. Laguna and E. Pueyo (2011). "A human ventricular cell model for investigation of cardiac arrhythmias under hyperkalaemic conditions." Philos Transact A Math Phys Eng Sci **369**(1954): 4205-4232.
Cartwright, N. (1999). The Dappled World. A study of the boundaries of science. Cambridge, Cambridge University Press.

Crick, F. H. C. (1958). "On protein synthesis " Symposia of the Society for Experimental Biology **12**: 138-163.
Crick, F. H. C. (1970). "Central Dogma of Molecular Biology." Nature **227**: 561-563.
Debruyne, J. P., E. Noton, C. M. Lambert, E. S. Maywood, D. R. Weaver and S. M. Reppert (2006). "A Clock Shock: Mouse CLOCK is not required for circadian oscillator function." Neuron **50(3)**: 465-477.
Dupré, J. (1993). The disorder of things. Cambridge, Mass, Harvard.
Feytmans, E., D. Noble and M. Peitsch (2005). "Genome size and numbers of biological functions." Transactions on Computational Systems Biology **1**: 44-49.
Foster, R. and L. Kreitzman (2004). Rhythms of Life. London, Profile Books.
Hardin, P. E., J. C. Hall and M. Rosbash (1990). "Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels." Nature **343**: 536-540.
Harold, F. (2014). In search of cell history. Chicago, University of Chicago Press.
Ho, M.-w. (2008). The Rainbow and the Worm. The physics of organisms. London, World Scientific Publishing.
Hunter, P. J., P. A. McNaughton and D. Noble (1975). "Analytical models of propagation in excitable cells." Progress in Biophysics and Molecular Biology **30**: 99-144.
International Human Genome Mapping Consortium (2001). "A physical map of the human genome." Nature **409**: 934-941.
Jack, J. J. B., D. Noble and R. W. Tsien (1975). Electric Current Flow in Excitable Cells. Oxford, OUP.
Luisi, P. L. (2016). The emergence of life. From chemical origins to synthetic biology. Cambridge, CUP.
Maturana, H. and F. Varela (1980). Autopoiesis and Cognition. Dordrecht, Reidl.
McClintock, B. (1984). "The significance of responses of the genome to challenge " Science **226**: 792-801.
Natarajan, C., F. G. Hoffmann, R. E. Weber, A. Fago, C. C. Witt and J. F. Storz (2016). "Predictable convergence in hemoglobin function has unpredictable molecular underpinnings." Science **354**: 336-339.
Noble, D. (2011). "Differential and integral views of genetics in computational systems biology." Interface Focus **1**: 7-15.
Noble, D. (2012). "A Theory of Biological Relativity: no privileged level of causation." Interface Focus **2**: 55-64.
Noble, D. (2016). Dance to the Tune of Life. Biological Relativity. Cambridge, Cambridge University Press.
Noble, D. (2017). "Evolution viewed from physics, physiology and medicine." Interface Focus **in press**.
Odegard, V. H. and D. G. Schatz (2006). "Targeting of somatic hypermutation." Nature Reviews Immunology **8**: 573-583.
Schrödinger, E. (1944). What is Life? Cambridge, Cambridge University Press.
Shapiro, J. A. (2011). Evolution: a view from the 21st century. Upper Saddle River, NJ, Pearson Education Inc.
Temin, H. M. and S. Mizutani (1970). "RNA-dependent DNA polymerase in virions of Rous sarcoma virus." Nature **226**: 1211-1213.
Whitfield, J. (2003). "Molecules form new state of matter." Nature **doi:10.1038/news031110-16**.

Harnessing stochasticity: How do organisms make choices?

Raymond Noble[1,a)] and Denis Noble[2,b)]
[1]*Institute for Women's Health, University College London, London WC1E 6AU, UK*
[2]*Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford OX1 3PT, UK*

(Received 10 May 2018; accepted 20 July 2018; published online 9 October 2018)

Choice in the behavior of organisms involves novelty, which may be unpredictable. Yet in retrospect, we can usually provide a rationale for the choice. A deterministic view of life cannot explain this. The solution to this paradox is that organisms can harness stochasticity through which they can generate many possible solutions to environmental challenges. They must then employ a comparator to find the solution that fits the challenge. What therefore is unpredictable in prospect can become comprehensible in retrospect. Harnessing stochastic and/or chaotic processes is essential to the ability of organisms to have agency and to make choices. https://doi.org/10.1063/1.5039668

Faced with unusual challenges in their environments, organisms have to make new choices to survive. The question addressed in this paper is how such choices can be creative and non-deterministic. We argue by analogy with the immune system, which faces a similar difficulty when a new antigen invades the organism, and for which it does not have the relevant DNA sequence to make an antibody with the correct shape. The immune system responds by rapidly mutating the variable part of the immunoglobulin sequence until, by chance, a cell evolves which does have the DNA sequence for an immunoglobulin with the correct shape. Stochasticity is therefore used to generate novelty. We speculate that by harnessing stochasticity in their nervous and other systems, organisms can similarly generate novel behavioral responses to meet the unusual challenge.

I. INTRODUCTION

How do organisms make choices? One very simple answer to this question would be that they do not. Following Descartes (1665),[1] the assumption would be that organisms are determinate machines. Despite their fiendish complexity, if we knew enough about the mechanisms involved, we would be able to predict their behavior to any arbitrary degree of accuracy.

Descartes actually excluded humans from this viewpoint, but that requires an assumption either that a *non-material entity* somehow intervenes in the case of humans, or that some *non-determinate (stochastic) material* process operates. Descartes chose the first option, which creates the difficulty that we have no way of representing how a material body could be so influenced. For example, would such an influence necessarily appear to be stochastic to scientific investigation, precisely because it would not be caused by any measurable physical events, and would have to appear to be stochastic in order to be indeterminate? Without making metaphysical assumptions beyond the possibility of scientific investigation, what we would find in this case simply collapses to the second possibility, at least insofaras we can investigate it objectively.[2]

In this article, we will conclude that stochastic material processes are involved. Moreover, there is no reason to suppose that such processes do not operate in organisms other than humans. Since humans evolved from other organisms, we should expect both of these conclusions.

Moreover, at the micro-level, we now know that the material universe is fundamentally stochastic, whether it be by virtue of random kinetic energy producing the form of stochasticity observed in the Brownian motion of molecules or by virtue of quantum mechanical behavior at the level of particles. Organisms must be affected by such stochasticity. Neither animals nor humans can be fully determinate. But that leaves open the question how stochasticity is involved or used in living processes.

In recent articles, we have addressed the following issues which can be seen to be introductory to the focus of the present article.

1. Can stochastic and/or chaotic processes be *used* in organisms, rather than organisms being arbitrarily subject to them, i.e., can such processes be harnessed so that they become part of the necessary functional repertoire of organisms? This issue was addressed in Noble[3(p.1)] and the answer is yes, organisms necessarily harness stochasticity.
2. Can we know whether organisms have agency, and does their behavior generate a form of directionality both in individual organisms, and at the level of populations so that the behavior can in turn influence the direction of evolution? This issue was addressed in Noble and Noble[4] and the answer again is yes, organisms do have agency. Harnessing stochasticity is an essential part of the means by which they do so. As we will show later in this article, a fully determinate process (meaning completely predictable) would not satisfy the conditions for agency.

Those articles leave open the question how the harnessing of stochasticity and the possession of agency may be represented in empirical (i.e., experimentally testable) terms. As a test of

a)r.noble@ucl.ac.uk
b)Denis.noble@dpag.ox.ac.uk

what we propose, we will also ask the question whether any such representation can show why we cannot predict what we call free choice, yet can often account for it in rational terms in retrospect.

II. DEFINITIONS

A. Agency

An agent acts, it does not just react in the way, for example, in which a billiard ball is caused by another ball to move. There are many levels of agency (Ref. 5, p. 32–40). Organisms are agents to the extent that they can interact socially with other organisms to choose particular forms of behavior in response to environmental challenges. Agency requires causal independence.[6] It also requires intentionality, i.e., the sense of purpose, in order to be causally effective as a driving force.[7]

B. Information

Inanimate objects can contain information. But it requires interpretation by an organism to become knowledge of what the information means. For example, rocks contain information, and that only becomes knowledge when organisms interpret it, e.g., to work out dates of events in the history of the earth. By this definition, DNA is also inanimate. It contains sequence information, but it does not contain knowledge. Until they are interpreted, DNA sequences are like uninterpreted hieroglyphics.

C. Interpreter

DNA information is interpreted by organisms. Outside a living cell, DNA is inert. A complete cell therefore is a minimal interpreter of DNA.

D. Knowledge

Knowledge about the world arises through organisms being creative in finding new solutions to environmental challenges. We can distinguish two types of knowledge:

E. Objective knowledge

This can be verified by those other than the organism that has the knowledge. In this sense, plants and bacteria have knowledge. Plants possess functional processes enabling them to use sunlight to create oxygen, and nutrients like sugars. We do not yet have that knowledge but wish we did! Note that this definition is not identical with Popper's use of "objective knowledge."[8]

F. Subjective knowledge

Organisms that "know that they know" have this kind of knowledge. They can communicate this kind of knowledge to others through behavior and language.

Note. Many philosophers do not attribute knowledge to organisms unless they are conscious, e.g., Anthony Kenny,[5] who refers to "capacity or ability" rather than "objective knowledge." We acknowledge the difference of usage of "knowledge" but do not think that the conclusions of our article depend very much on which view one takes. Here, we simply note that resolving this question would depend on one's view of animal consciousness; see, e.g., Ref. 9. In this article, we are not primarily concerned with this kind of knowledge, and we do not address questions of self-awareness and consciousness.

G. Rational choice

In this article, we refer to accounting for choice behavior in retrospect as being rational. What is meant is that it is possible to answer the question why an organism did what it did using the common sense meaning of rational, e.g., in terms of the organism's presumed goals. This does not mean that the organism's choice would be predicted by any particular version of Rational Choice Theory (https://en.wikipedia.org/wiki/Rational_choice_theory). Nor does it mean that the "rationality" does not contain an element of delusion. We will return to this question in the Discussion.

H. Stochasticity

Interpreted as the inability to predict, stochasticity is a level-dependent property. Thus, molecular level stochasticity is compatible with higher level predictability, as is obvious from the predictability of thermodynamics. Stochasticity is therefore a relativistic concept. Whether underlying stochasticity can influence the overall behavior of a system must depend on whether the higher level is organized to enable it to do so. Organisms are high-level systems. In this article, we show that molecular stochasticity does not only cancel itself out at higher levels, as in the case of thermodynamics, it also becomes used in goal-directed feedback control processes. Higher-level organization can make that possible.

I. Chaos

As many readers, particularly of this journal, will be aware, stochasticity and chaos are not identical. Chaotic sequences can be produced by determinate algorithms as first shown by Lorenz.[10,11] The difference is important because the variations in determinate chaos are constrained by an attractor, whereas genuine stochasticity is not. The difference can be made clear in phase plots. However, we doubt whether the difference between determinate chaos and stochasticity is relevant to the process we ascribe to choice behavior. If the attractor constraining a chaotic sequence is not itself an integral part of the organism's control networks, the variations will appear random to the choice process.

III. MULTI-LEVEL CAUSATION

An important basis for our paper is that organisms are open systems in which causation operates between multiple levels. That they are open systems is obvious: they exchange matter and energy with their environment and engage in social interactions with other organisms. Multi-level causation is not, however, universally accepted in biology. We follow the argument that causation *must* be multi-level. The demonstration that this is the case is mathematical. Even if we try to imagine that only molecular level mechanisms are causative,

we are faced with a fundamental difficulty when we try to solve the differential equations for those mechanisms. There is no solution unless we introduce boundary conditions that represent the causative action of higher levels and scales. This is the mathematical basis of the principle of biological relativity.[12,13] The principle states that there is no privileged level of causation. But it is important to note that the upward and downward forms of causation are not necessarily of the same form. Causation by setting the boundary conditions for lower level processes is more like a constraint on the forms of organisation that the lower level elements may take.[14] These causal interactions can occur between any of the levels of organization and are the reason why downward is causally effective. Indeed, in purposive behavior, it is primary since it will only be at the higher levels that the purposive organization may be evident.[4] This is the general causal basis for the choice process that we will now present.

IV. THE CHOICE PROCESS

For an empirically testable theory of choice to be possible, we need to know at which stages in the process experimental interventions could test its validity. At first sight, that may seem impossible. How can we specify a process that is necessarily *unpredictable* but which can be given an at least apparently *rational* justification once it has happened? Our previous work provides a clue to that problem. In Ref. 4, we analyzed agency by comparing it to the purposive behavior of the immune system. The immune system solves what we can best characterize as a template puzzle: given a new invader with an unknown chemical profile (shape of template), what is the best way to find the key (an anti-template, i.e., the antibody) to lock onto and neutralize the invader? The answer in the case of the immune system is one of the most remarkable forms of the harnessing of stochasticity. In response to the new environmental challenge, a feedback loop activates a massive increase in mutation rate in a highly targeted region of the immunoglobulin DNA sequence.[15]

The process of choice in organisms can be viewed as analogous to the immune system. The process can be represented as follows:

1. Influences from the environment (boundary conditions) and the organism's history (initial conditions) lead to defining the problem facing the organism. This will be the state of the organism in which the environmental challenge has occurred but not yet a solution in the organism's reactions. We conjecture that such a problem can be viewed as a puzzle analogous to the form of a template for which a match is needed. The configuration of these conditions might be a routine one, in which case what we normally characterize as a reflex response may be adequate. But it is precisely such responses that we would *not* characterize as involving a choice. We say that a choice occurs when there is no automatic reflex response possible. The challenge facing the organism then is what could fit the puzzle template?
2. Instead of an automatic response therefore, the organism must search amongst existing stored possible fits to the problem template. By analogy with the immune system, this is equivalent to finding that the DNA sequence for the correct immunoglobulin shape already exists. It is precisely when no such solution exists that hypermutation is triggered. We hypothesise that a comparable process occurs generally in choice situations in organisms.
3. In which case the organism can spin (i.e., activate) stochastic processes within itself to generate further possible new solutions. This is where novelty arises. These processes can be of any biological kind. For cognitive problems in organisms with highly developed nervous systems, these will be primarily neural. Note also two important characteristics of this stage of the process. First, the organism *triggers* the resort to stochasticity but no longer *controls* it, just as the immune system does not directly control which mutations occur. Second, the options at this stage are effectively infinite. In the case of the immune system, the number of possible sequences for the variable part of the immunoglobulin must be larger than the total number of particles in the universe. That is also true for the number of interactions between the 20 000 or so genes in a human.[16] Stochasticity and/or chaos in the nervous system must make even more options available.[17]
   Neural processes are extensively stochastic—at all functional levels, from the opening and closing of ion channels via action potential generation, spontaneously or through synaptic transmission in neuronal networks, up to cognitive functions including decision making (8 chapter 22).[18–22] As pointed out in Braun,[23] the reason may be found in the functional organization of living systems composed of a manifold of nonlinear feedback loops that often are adjusted to operate in the neighbourhood of bifurcations where it can essentially depend on random effects of what will happen next, e.g., whether an ion channel is opened or remain closed or whether an action potential is generated or not—what even may decide the choice between leaving the bar and going home or having another drink.
4. The organism returns to direct control at the next stage, which is to compare what is thrown up by the stochastic process with the problem template to determine what fits. "Template" and "fit" here are used metaphorically, in much the same sense in which a logical answer can be said to "fit" (i.e., answer to) the problem posed by a question. This is the essential choice process, needing a comparator. The comparator therefore forms part of what we call the interpreter (see definitions). This is the stage at which we can say that the organism knows that it has found a possible solution.
5. The final stage is to implement the discovered action to solve the problem.
   This is an idealized process, but it clearly helps one to explain an apparent paradox regarding the predictability or otherwise of what we call a free choice. Step 4 ensures that, in retrospect (and only in retrospect), the choice may be what in the case of humans we call rational. There may be a complete logic to why it was made. The logic lies in the fit between the problem template and the solution template. But step 3 ensures that the choice was

unpredictable since we cannot predict what stochasticity will throw up. So, free choice is both rational and novel. (See also Parallels in the work of Karl Popper below.)
This hypothetical process is open to empirical tests at all stages since it makes significant assumptions about what is actually happening within the organism. The kind of knowledge the organism has is what Popper characterised as objective knowledge (as distinct from subjective knowledge) and is fully open to observational test.

For example, the existence of stage 3 naturally explains why problems leading to the necessity for making a choice may lead to what we can call the puzzled state. Before stage 3, there is no solution in sight. Only after stage 3 might there be a solution that can lead to rational action. There will therefore be a period during which the organism does not know the solution. In the case of humans, we can communicate such states in language ("I havn't a clue"). Other organisms can communicate by behavior: frustration, depression, displacement activity, etc.

V. ACKNOWLEDGMENT OF PREVIOUS WORK

We are far from being the first to favor active agency as an explanation of the behavior of organisms and to favor the role of choice in the direction of evolutionary change. The arguments about the active role of organisms have their origins in a long tradition in which deterministic and non-deterministic views of life have been pitted against each other. As noted in our Introduction, the two threads were present in the same philosopher in the case of Descartes who in the seventeenth century struggled to reconcile his determinist interpretation of animal behavior with his conviction that this could not be true of humans. How else could he have written his great works? It would have taken a monkey billions of years to manage by chance to type just a single sentence of Descartes' work (Ref. 4, p. 1). (The relevant combinatorial mathematics is given in Ref. 16; see https://en.wikipedia.org/wiki/Weasel_program.)

The existence of creativity shouts out loud and clear that the universe cannot be simply deterministic, and since the early 20th century revolutions in physics, we have the proof that it is not. Yet, this revolution had surprisingly little effect on biology, which continued with deterministic interpretations of life and its evolution throughout the century. It was thought that indeterminacy at microphysical scales could hardly be relevant to processes at physiological scales. The proof that it is relevant came with the discovery of the hypermutation mechanism in the immune system. As we have shown in previous articles, the harnessing of stochasticity at a molecular level is precisely what enables organisms to be creative. The immune system serves as a model, which can be generalized (Ref. 4, p. 4–5). Given the nature of the universe, uncertainty is inevitable. Choice necessarily involves dealing with uncertainty. Low-level stochasticity is the clay from which high-level novelty can develop.

We wish to credit two more recent predecessors for major influences on our ideas: Patrick Bateson and Karl Popper. Patrick Bateson's work on the active role of behavior in evolution[24–28] was pursued throughout his career and has been summarized in a book published just before his death in 2017.[27] He was a careful historian as well as a great biological scientist. He documented the development of the ideas of active agency through from Darwin, through Spalding and Baldwin to his phrase the "adaptability driver" to describe the active nature of organism agency.[26] His phrase captures the directionality of agency in organisms.

VI. PARALLELS IN THE WORK OF KARL POPPER

Amongst fore-runners of the ideas explored in this article is the outstanding work of Karl Popper. In 1986, Popper gave a lecture to The Royal Society in London in which he laid out his "New Interpretation of Darwinism."[29] In that lecture, he distinguished between "passive Darwinism" and what he called "active Darwinism." His "passive Darwinism" is more or less identical with classical neo-Darwinism: the theory that random genetic variation and natural selection are entirely sufficient (*allmacht* in Weismann's words)[30] to explain evolution. Popper wrote: "I shall attempt to turn the tables completely on passive Darwinism … I shall claim that the *only* creative element in evolution is the activity of living organisms."[29(p.119)] "Active Darwinism" is therefore equivalent to the theory that organisms have agency and make choices, which is the main theme of our paper. Those choices include choosing niches (niche selection theory) and which other organisms they interact with (including sexual selection), and more recently, the discovery of aversion to cheating behavior in populations of dogs[31] and monkeys.[32]

Popper regarded the "metaphor of 'natural selection'" as "a theory of error elimination"[29(p.120)] rather than being creative of novelty itself. He saw it as a filter eliminating errors. To understand this point, we should remember that Darwin contrasted natural selection with artificial selection, which is clearly choices made by organisms (the selective breeders). When Darwin realised that sexual selection is more like artificial selection, he therefore faced a problem. Sexual selection is clearly an *activity* of organisms determining their evolution. The problem is that this blurs the distinction he was drawing. Sexual selection is therefore a form of active Darwinism to use Popper's terminology. Specifically, he wrote "sexual selection is a refutation of natural selection."[29(p.128)]

Popper saw that complete determinism was incompatible with viewing organisms as agents making choices. He would therefore have seen the importance of the role of stochasticity in our paper. In *The Open Universe*, Popper demonstrated that indeterminism is a necessary but not sufficient condition for emergence and openness.[29(p.70)]

In the same exposition of Popper's ideas leading up to his Royal Society lecture, Niemann[29] presents some other points that correspond well to the ideas of our paper. Summing up Popper, he repeated that "all life is problem solving. Acquiring new knowledge is always purposeful activity."[29(p.90)] He insisted that "in all cases the activity comes from outside of the DNA. The former 'centre of life' is rather a dead place."[29(p.96)] That it is the cell that divides, not only the DNA.[29(p.98)] And that it is "The cell … also managing the genome."[29(p.101)] This insight resembles that of Barbara

McClintock, the discoverer of natural genetic engineering[33] in saying that "the genome is an organ of the cell."[34]

Finally, there is his point that "influences (on action) [are] traceable in hindsight… we are unpredictable but not irrational" (Ref. 29, p. 110). Popper therefore arrived at many of the points we are making here.

It would therefore be surprising if he had not also seen the obvious implication, which is that organisms harness stochasticity; otherwise, choice behavior would not be possible. We are grateful to Hans-Joachim Niemann for directing us to Popper sources preceding his Royal Society lecture where he does clearly draw the correct conclusion. Some of the relevant texts occur in his dialogue with John Eccles *The Self and Its Brain*.[35] Popper writes "New ideas [*in statu nascendi*] have a striking similarity to genetic mutations" and continues "describing 'the process with respect to new ideas and to free will decisions' (Ref. 35, p. 540). As randomly produced proposals followed by selection based on standards coming from the world" (cf. Ref. 36, Secs. 31–33). Popper arriving at this conclusion is all the more remarkable for the fact that it required him to abandon his earlier (1973) conclusion that "indeterminism is not enough."[37,38]

The main difference is that while he envisaged "the cell… also managing the genome," (Ref. 29, p. 101). He does not seem to have arrived at the details of the comparison with hypermutation in the immune system. Perhaps, this is attributable to the fact that the discovery of some of the detailed molecular mechanisms of somatic hypermutation occurred in 1999 after his death in 1994.[39,40] There may also have been a puzzle regarding the molecular mechanism of hypermutation. Increasing the natural mutation rate by a factor of up to 10^6 must have seemed implausible. But this is also roughly the order of magnitude difference between the natural mutation rate in DNA copying before and after repair by cellular editing mechanisms. Mismatch DNA repair is indeed suppressed during somatic hypermutation.[41]

VII. DISCUSSION

Our main conclusion is that it is possible to construct an account of choice behavior using stochastic processes by analogy with the way in which the immune system harnesses stochasticity to discover novel solutions to new challenges. There are several predictions and implications.

A. Psychological experiments on primates

One of the implications is that it could be important in investigations of choice behavior in animals to include tests for signs of delay or other behavioral signs attributable to stage 3 in our choice process. These could include hesitation (time taken to decide), displacement activity, or other signs of puzzlement. Just as an example, we could take from many good and interesting studies of animal choice; a study of risk-taking behavior in primates[42] was successful in showing varying degrees of risk-taking in the different primate species but did not include any parameter that would answer this question. Most studies on choice in animal behavior seem to be assuming that animals behave as though they solve a calculation of probability. Thus, in the cited paper, we find: "Any agent, in order to successfully navigate a world of possibilities, needs to strike the right balance between these factors, utilizing mechanisms that when confronted with risky choices, lead to decisions, which optimally combine the probability of receiving a reward multiplied by the amount of the reward." Animals may not actually be "calculating" in quite the way this quote implies. If we are correct, no calculation or its equivalent, using, e.g., forms of Rational Choice Theory, could represent all of what is happening. That is particularly true when extrapolation to human behavior is involved. To quote the same source: "Based on our findings, we propose that decision-making in the great apes provides a promising context for the interpretation of decision-making in humans, the fifth great ape species." We agree with this conclusion, but note that it will be particularly important to consider the role of stochasticity in both animals and humans.

Krupenye *et al*. have in any case shown that humans and animals display departures from Rational Choice Theory which they characterize as biases in choice behavior dependent on whether decisions involve losses or gains.[43] The involvement of stochastic processes does not of course exclude biases.

Rosati and Hare have shown that chimpanzees and bonobos can distinguish between risk and ambiguity in choices presented to them.[44] They write "Importantly, apes' divergent preferences for risk and ambiguity diminished with time: although apes chose the risky option more frequently than the ambiguous option in the first session; by session two they showed no difference. One possibility is thus that the apes are able to rapidly incorporate new information about previously ambiguous options into their decision strategies: after choosing the ambiguity option and receiving some feedback about what it provided, they may have treated the ambiguity and risk option as equivalent because the functional outcome was the same." The stochastic choice process we describe here would account for this form of learning. By analogy with the immune system model, once a novel challenge has been met, it becomes part of the standard repertoire.

Santos and Rosati have written a valuable review of this field.[45] They write "we now know that human choice is often not as rational as one might expect." We see two ways in which this statement can be interpreted. First, within the context of our Choice Process, there is obviously no guarantee that a stochastic process will throw up a fully rational solution. Partial success is what would be expected most of the time. The same is true of the immune system. All it needs to do is to come up with a "good enough" template match. It does not have to be the perfect match. If a key fits the lock, it does not really matter whether it is an exact fit.

Second, that leaves the question how it happens that, nevertheless, most of the time, we and others can give a "good enough" rational explanation of a choice, at least in retrospect. That seems to be true however partial the "fit" seems to be to the problem. A possible solution to that problem could be what Santos and Rosati call the endowment effect. Animals and humans privilege retaining what they already own. Could the same effect operate in the case of decisions? Do we and perhaps other animals "own" decisions. It seems plausible at least.

B. Observations on primates in the wild

Observations of primates in their natural environments have extended our knowledge of choice behavior in ways that enable us to obtain important insights into *subjective* knowledge.

That organisms may know that others have subjective knowledge is itself an important factor in objective knowledge and is part of situational logic or behavioral cognition. Such knowledge necessarily carries with it a great deal of uncertainty. An animal must predict that the other knows and how they might act on such presumed knowledge. This is manifest in both human and non-human animal behavior.

Spinning the wheel as a creative process therefore occurs not solely at a physiological level, but also at a social and cultural level. The evolution of language allows sophisticated and abstract problem solving. Language allows a cultural spinning of the wheel. Thus, chimpanzees use communication that distinguishes private from public interaction. What they know that others may not know is a part of their objective knowledge. Chimpanzees employ signals with a sensitivity to the public/private nature of information, by adjusting their use of signal types according to social context and by taking into account potential out-of-sight audiences.[46]

The written and recorded word, together with artistic representation, allows problem solving across many generations—a repository of social wheel spinning, and to "see" the world in different ways. Solutions to problems can differ from group to group depending on context and cultural history. This is evident in the use of tools by chimpanzees to crack nuts. The use of stones to crack nuts has to be "introduced" to the group and is learned by others in the group. Furthermore, the stones are modified to better crack the nuts. Tools may be shared or hidden and kept for later use. This demonstrates creative decision making in practice.

C. Observations on *Drosophila* short-term memory mutants

A further prediction is that choice behavior should depend on the processes of plasticity since the ability to store and retrieve the results of stochastic variation requires such plasticity. Tang and Guo[47] and van Swinderen[48,49] showed that choice behavior in *Drosophila* is strongly affected by mutations that lead to defective short-term memory. The behavior that remains is then rigid optomotor responses. As van Swinderen expresses it, "a strong and non-distractable optomotor response, as seen in the *dnc* and *rut* mutants, may reflect failure of an interacting attention-like mechanism designed to periodically alternate among competing percepts of variable salience." Alternating between competing outcomes of stochastic processes is precisely what must be involved in the choice process.

ACKNOWLEDGMENTS

D.N. would like to thank Professor Michael Joyner for hosting his visit to the Mayo Clinic in Rochester, MN, while this article was being written. R.N. is an Honorary Senior Lecturer at the Institute of Women's Health at University College London. We thank Hans Braun, Michael Joyner, Anthony Kenny, Hans-Joachim Niemann, and the two journal referees for valuable comments on early drafts of this article.

[1] In his Treatise on the formation of the foetus, Descartes wrote: "If one had a proper knowledge of all the parts of the semen of some species of animal in particular, for example of man, one might be able to deduce the whole form and configuration of each of its members from this alone, by means of entirely mathematical and certain arguments, the complete figure and the conformation of its members."

[2] A Cartesian dualist would argue that, nevertheless, it would not appear to be stochastic to the individual concerned since, subjectively, he/she would have willed the action.

[3] D. Noble, "Evolution viewed from physics, physiology and medicine," Interface Focus **7**, 20160159 (2017).

[4] R. Noble and D. Noble, "Was the watchmaker blind? Or was she one-eyed?," Biology **6**(4), 47 (2017).

[5] A. J. P. Kenny, *The Metaphysics of Mind* (Oxford University Press, Oxford, 1992).

[6] K. D. Farnsworth, "How organisms gained causal independence and how it might be quantified," Biology **7**, 38 (2018)

[7] H. Liljenstrom, "Intentionality as a driving force," J. Conscious. Stud. **25**, 206–229 (2018).

[8] K. R. Popper, *Objective Knowledge. An Evolutionary Approach* (Oxford University Press, Oxford, 1972).

[9] D. R. Griffin, *Animal Minds. Beyond Cognition to Consciousness* (University of Chicago Press, Chicago, 1992).

[10] E. N. Lorenz, "Deterministic non-periodic flow," J. Atmos. Sci. **20**, 130–141 (1963).

[11] D. P. Feldman, *Chaos and Fractals: An Elementary Introduction* (Oxford University Press, Oxford, 2012).

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

[13] D. Noble, *Dance to the Tune of Life. Biological Relativity* (Cambridge University Press, Cambridge, 2016).

[14] More precisely, boundary conditions always arise from higher scales: that is why they are "boundary" conditions (outside the boundaries of the defined system), but not all such conditions arise from higher-level organization. The difference between scales and levels is important. Levels are defined by their organization (as cells, tissues, etc.). Scale is a neutral matter of size.

[15] V. H. Odegard and D. G. Schatz, "Targeting of somatic hypermutation," Nat. Rev. Immunol. **8**, 573–583 (2006).

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

[17] Since the number of possible interactions between around 25 000 genes is already vastly larger than the total number of particles in the known universe, a brain containing 100×10^9 nerve cells will be capable of a number of possible interactions that effectively ensure that any particular one will never recur again.

[18] B. Hille, *Ionic Channels of Excitable Membranes* (Sinauer Associates Inc., Sunderland, MA, 1992).

[19] B. D. Burns, *The Uncertain Nervous System* (Arnold, London, 1968).

[20] M. Heisenberg, "Is free will an illusion?," Nature **459**, 164–165 (2009).

[21] A. Tchaptchet, W. Jin, and H. A. Braun, "Diversity and noise in neurodynamics across different functional levels," in *Advances in Cognitive Neurodynamics*, edited by R. Wang, X. Pan (Springer, Singapore, 2015), vol. 5, pp. 681–687.

[22] B. Brembs and M. Heisenberg, "Der Zufall als kreatives Element in Gehirn und Verhalten," in *Zufall in der belebten Natur*, edited by U. Herkenrath (Verlag Roman Kovar, Hennef, 2018), pp. 80–94; (chance as a creative element in the brain and behavior).

[23] H. A. Braun, *Der Zufall in der Neurobiologie - von Ionenkanälen zur Frage des freien Willens. Zufall in der belebten Natur* (Verlag Roman Kovar, Hennef, 2018), pp. 109–137; (chance in neurobiology—from ion channels to the question of free will).

[24] P. Bateson, "The active role of behaviour in evolution," in *Evolutionary Processes and Metaphors*, edited by M.-W. Ho and S. W. Fox (Wiley, Chichester, 1988), pp. 191–207.

[25] P. Bateson, "The active role of behaviour in evolution," Biol. Philos. **19**, 283–298 (2004).

[26] P. Bateson, "The adaptability driver: Links between behaviour and evolution," Biol. Theory **1**, 342–345 (2006).

[27]P. Bateson, *Behaviour, Development and Evolution* (Open Book Publishers, London, 2017).
[28]P. Bateson, "Adaptability and evolution," Interface Focus **7**, 20160126 (2017).
[29]H.-J. Niemann, *Karl Popper and the Two New Secrets of Life* (Mohr Siebeck, Tuebingen, 2014).
[30]A. Weismann. *Die Allmacht der Naturzüchtung; eine Erwiderung an Herbert Spencer* (Fischer, Jena, 1893); (the omnipotence of natural breeding; a reply to Herbert Spencer).
[31]J. L. Essler, S. Marshall-Pescini, and F. Range, "Domestication does not explain the presence of inequity aversion in dogs," Curr. Biol. **27**, 1–5 (2017).
[32]S. F. Brosnan and F. B. De Waal, "Monkeys reject unequal pay," Nature **425**, 297–299 (2003).
[33]J. A. Shapiro, *Evolution: A View from the 21st Century* (Pearson Education Inc, Upper Saddle River, NJ, 2011).
[34]B. McClintock, "The significance of responses of the genome to challenge," Science **226**, 792–801 (1984).
[35]K. R. Popper and J. C. Eccles, *The Self and Its Brain* (Springer International, New York, 1977).
[36]H.-J. Niemann, "Nachwort des Herausgebers," in *Wissen und das Leib-Seele-Problem*, edited by K. R. Popper (Mohr Siebeck, Tubingen, 2012), pp. S510–S546; (afterword of the publisher; K.R. Popper, Knowledge and the mind-body problem).
[37]K. R. Popper, "Indeterminism is not enough," Encounter **40**(4), 20–26 (1973).
[38]There are also echoes of Popper's and our view as far back as *The Open Society and Its Enemies* (Popper, 1945, Routledge), e.g., Vol. 2, p. 210, where he refers to "accidental experiences" as one of the determinants of novelty, using Beethoven as an example.
[39]M. Muramatsu, V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson *et al.*, "Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells," J. Biol. Chem. **274**, 18470–18476 (1999).
[40]Z. Li, C. J. Woo, M. D. Iglesias-Ussel, D. Ronai, and M. D. Scharff, "The generation of antibody diversity through somatic hypermutation and class switch recombination," Genes Dev. **18**, 1–11 (2014).
[41]H. Saribasak and P. Gearhart, "Does DNA repair occur during somatic hypermutation?," Seminar. Immunol. **24**, 287–292 (2012).
[42]D. B. M. Haun, C. Nawroth, and J. Call, "Great apes' risk-taking strategies in a decision making task," PLoS ONE **6**, e28801 (2011).
[43]C. Krupenye, A. G. Rosati, and B. Hare, "Bonobos and chimpanzees exhibit human-like framing effects," Biol. Lett. **1**, 20140527 (2015).
[44]A. G. Rosati and B. Hare, "Chimpanzees and bonobos distinguish between risk and ambiguity," Biol. Lett. **7**, 15–18 (2011).
[45]L. R. Santos and A. G. Rosati, "The evolutionary roots of human decision making," Ann. Rev. Psychol. **66**, 321–347 (2015).
[46]C. Hobaiter, R. W. Byrne, and K. Zuberbühler, "Wild chimpanzees' use of single and combined vocal and gestural signals," Behav. Ecol. Sociobiol. **71**, 96 (2017).
[47]S. Tang and A. Guo, "Choice behavior of *Drosophila* facing contradictory visual cues," Science **294**, 1543–1547 (2001).
[48]B. van Swinderen and K. A. Flores, "Attention-like processes underlying optomotor performance in a *Drosophila* choice maze," Dev. Neurobiol. **67**, 129–145 (2007).
[49]B. van Swinderen, "Attention-like processes in *Drosophila* require short-term memory genes," Science **315**, 1590–1593 (2007).

biology

MDPI

Review

Was the Watchmaker Blind? Or Was She One-Eyed?

Raymond Noble [1] and Denis Noble [2,*]

[1] Institute for Women's Health, University College London, Gower Street, London WC1E 6BT, UK; r.noble@ucl.ac.uk
[2] Department of Physiology, Anatomy & Genetics University of Oxford, S Parks Rd, Oxford OX1 3QX, UK
* Correspondence: Denis.noble@dpag.ox.ac.uk

Academic Editors: Edward L. Braun and Jukka Finne
Received: 6 July 2017; Accepted: 14 December 2017; Published: 20 December 2017

Abstract: The question whether evolution is blind is usually presented as a choice between no goals at all ('the blind watchmaker') and long-term goals which would be external to the organism, for example in the form of special creation or intelligent design. The arguments either way do not address the question whether there are short-term goals within rather than external to organisms. Organisms and their interacting populations have evolved mechanisms by which they can harness blind stochasticity and so generate rapid functional responses to environmental challenges. They can achieve this by re-organising their genomes and/or their regulatory networks. Epigenetic as well as DNA changes are involved. Evolution may have no foresight, but it is at least partially directed by organisms themselves and by the populations of which they form part. Similar arguments support partial direction in the evolution of behavior.

Keywords: blind chance; harnessing stochasticity; hypermutation; evolutionary hold mechanisms; adaptability driver; internal and external goals

1. Introduction

We again use our computer monkey, but with a crucial difference in its program. It again begins by choosing a random sequence of 28 letters, just as before ... it duplicates it repeatedly, but with a certain chance of random error—'mutation'—in the copying. The computer examines the mutant nonsense phrases, the 'progeny' of the original phrase, and chooses the one which, *however slightly*, most resembles the target phrase, METHINKS IT IS LIKE A WEASEL. (Richard Dawkins, *The Blind Watchmaker*).

1.1. Background

In chapter 3 of his book, *The Blind Watchmaker* [1], Richard Dawkins produces his famous Weasel program. He shows that a monkey writing out 28 characters randomly on a typewriter would require much more than the whole lifetime of the universe to arrive by pure chance at a correct 28 letter sequence to match Shakespeare's text [2]. But if each correct character were to be held constant between successive generations of random typing, it would require only a modest number (43) of iterations to achieve a correct result. The program resembles the operation of an old-fashioned three-wheel fruit (slot) machine. If the target for a reward is, say, three lemons and a spin of the wheels produces two, the best strategy might be to hold the wheels with the two lemons and spin the remaining wheel until that also shows a lemon. The number of 'wheels' (28 in the Weasel example) doesn't change the principle of this mechanism.

The example shows that, however unlikely a pattern might be, it might evolve in a reasonable amount of time by using such an incremental strategy.

Dawkins acknowledges that the original program does not truly represent the process of blind variation followed by natural selection assumed in neo-Darwinist evolutionary models since it uses a long-term goal set by the computer program writer. The program only 'knows' when to hold a character constant between generations by comparing it with the long-term goal. As Dawkins writes:

> "Life isn't like that. Evolution has no long-term goal. There is no long-distance target, no final perfection to serve as a criterion for selection, although human vanity cherishes the absurd notion that our species is the final goal of evolution. In real life, the criterion for selection is always short-term, either simple survival or, more generally, reproductive success."

It is also important to acknowledge that more complex versions of the Weasel program have been produced that do not require that a correct character should be completely fixed. In those cases back-mutation is also possible. But all these programs still require various kinds of comparison, in the selection process, with the long-term goal in order to succeed. When we refer to the 'hold' metaphor in this article it is important to note that this does not completely exclude mutations. It refers to the ability to preserve existing functionality sufficiently well for subsequent generations to inherit.

A further important deficiency in the Weasel program as originally formulated is that it assumes that the goal would be correctly represented by a particular genome sequence. From the viewpoint of organisms and their functionality, that is not correct. Genomes and phenotypes are far from being equivalent. The mismatch works both ways. The same genome can be used to generate many different phenotypes and the same phenotype can evolve through many different genome variations, to such an extent that the sequence variations may even be unpredictable [3]. It is of course the functional phenotype that is 'seen' by natural selection. DNA sequences are not directly available for selection other than through their functional consequences in the production of RNAs and proteins, and even most of those variations are effectively buffered by the regulatory networks and so may also be invisible to natural selection. For example, 80% of DNA knockouts in yeast are 'silent' in controlled experimental conditions [4]. We will return to this important point later (see Section 4.2) since it is the fundamental reason why gene-centric views of evolution are incorrect. Evolution is a high-level *forming* process, not simply a matter of genome informatics.

1.2. Purpose and Organization of This Article

In this article we will agree with Dawkins that (a) completely stochastic processes with no 'hold' or similar 'guiding' mechanism would require impossibly long periods of time for successful evolution to occur, and (b) there is no need to assume that evolution has a long-term goal. This is where both he and we part company with Intelligent Design (ID) and creationist theories.

But we will nevertheless show that organisms and populations of organisms do have identifiable and empirically testable goals, and that variations on the theme of the Weasel program found experimentally in nature, show this to be true. The key to understanding why we differ from neo-Darwinists on this matter lies in multi-level feedback processes that have been shown to exist which enable organisms and populations to direct their evolution in response to stimuli from the environment and so achieve the inheritance of acquired characteristics. These feedback processes require analysis of function at a high (systems) level, e.g., networks, cells, tissues, organs and the organism as a whole in interaction with the environment, including other organisms. Multi-level feedback is a requirement of goal-directed behaviour. A purely gene-centric view will not necessarily 'see' such feedback. Empirical tests used routinely in physiology and engineering do so readily.

Our article is not intended to be a systematic review. It is rather the development of a conceptual interpretation of the process of evolution that differs from neo-Darwinism in implementing the principle that there is no privileged level of causation [5,6]. We believe this is a novel conceptual advance. It differs radically from views of evolution that privilege the role of DNA sequences in the intergenerational transmission of inheritance [7–9], and is therefore more sympathetic to

views of evolution that emphasise the active role of functional regulatory networks and behavior in organisms [6,10]. These are the processes that endow organisms with what we will call natural purposiveness. DNA alone cannot do that. Outside the regulatory network environment of the complete cell it is inactive.

We develop our case in stages: first, to show how multicellular organisms use targeted evolution of their cells to respond to environmental challenge; second, to show how populations of microorganisms achieve similar targeted responses; third, to show how epigenetic inheritance occurs in multicellular organisms with separate germ-lines; fourth; to show how the evolution of behavior can use similar processes that have developed agency in their evolution. In all these cases, variation is not random with respect to genome location and/or organism functionality. The targeting of variation and the preservation of existing functionality ensure that evolution is not entirely blind. Where relevant we reference alternative viewpoints, including standard neo-Darwinist interpretations. But this article does not analyse where we believe those viewpoints are deficient. That was the purpose of a recent related article [11].

2. Definitions

Agency: an agent acts, it does not just react in the way, for example, in which a billiard ball is caused by another ball to move. Organisms are agents to the extent that they can interact socially with other organisms to choose particular forms of behavior in response to environmental challenges. This definition of agency can therefore apply to microorganisms, such as bacterial films and eukaryotic slime moulds, that form interacting communities [12,13] as well as to multicellular organisms. In principle, it can also apply to the subcellular networks responsible for buffering organisms against many forms of DNA variation.

Goals: Goals can be ascribed to agents since choice of action involves directionality in their actions. A goal in this sense is the situation towards which the agent's action leads. Goals arise naturally from within the agent's cognitive behavior, albeit in interaction with other agents. This kind of behavior can be called natural purposiveness. Goals can therefore be ascribed empirically on the basis of observation of the behavior of organisms.

Teleology: The possession of goals is what defines teleology. Some biologists prefer the word teleonomy [14] to emphasise the view that goals in organisms (sometimes with the qualification 'other than humans') are only apparent. Since our use of the word 'goal' enables empirical physiological tests for the presence of the required natural purposiveness we see no need to avoid the word teleology.

Natural purposiveness: Natural purposiveness is an emergent property of multi-level evolved systems. It is easier to understand and appreciate its significance within the principle of biological relativity, i.e., no privileged level of causation [6].

Neo-Darwinism: Classical neo-Darwinism was formulated by August Weismann [15,16] and others in the late nineteenth century to expunge the inheritance of acquired characteristics from Darwin's theory. Blind variation followed by natural selection was claimed to be *entirely sufficient* (Weismann's *allmacht*). This is clear from his extensive argument with Herbert Spencer [17–19]. Many biologists today redefine neo-Darwinism in various ways (see e.g., the on-line dialogue between one of us and David Sloan Wilson (https://thebestschools.org/dialogues/evolution-denis-noble-david-sloan-wilson/) and the relevant entry in the Encyclopedia of Evolution [20]). Redefining a term does not however change the fact that the original theory using that term is no longer the complete story. Our position can therefore, to some degree, be seen to return to Darwin's multi-mechanism viewpoint, though with vastly extended empirical evidence (see Figure 6 in reference [11]).

Gene-centrism: We will refer to gene-centric views of evolution several times in this article. There are two senses in which we view neo-Darwinist theories as gene-centric. The first is the view that the genome is "the Book of Life" [21], i.e., that the development of an organism is essentially a read out of the DNA sequences, in interaction with the environment. The hidden assumption here is that inheritance depends on DNA alone. Sometimes this is spelt out, as in the distinction between the 'replicator' (DNA) and the 'vehicle' (the rest of the 'disposable' organism). The second sense is that, even though it is the phenotype that is selected in evolution, only those aspects of the phenotype that are represented in DNA are inherited.

These definitions are conceptual, as are all definitions, but they endow the theory we develop here with empirically testable predictions.

3. Goals within Organisms

3.1. Regulated (Directed) Hypermutation Processes

The Weasel program example shows that the monkey at the keyboard needs some kind of guidance to have any chance at all of reaching the goal. In the evolutionary process the 'monkey at the keyboard' is blind chance mutations, the process assumed in neo-Darwinism to be the only process producing genetic variation. The assumption that all mutations are produced by blind chance is central to the theory. This is the assumption that appears to exclude goal-directed behavior [6].

Yet, as we will show in this paper, organisms have demonstrably evolved *guided* random mutation mechanisms that can respond rapidly and correctly to environmental challenges. These mechanisms allow organisms and populations to harness stochasticity to evolve a solution to such challenges at high speed compared to what could be achieved by blind chance alone. It is the harnessing of stochasticity in guided response to environmental challenges that achieves what blind chance alone could not possibly do [11].

One way in which the guidance can occur is through the process of natural selection. Progressively, through the generations, selection acts as a filter. Neo-Darwinism assumes that this is the only guide. We disagree with that view because it is demonstrably insufficient: nature also uses other faster guidance processes.

How can that be achieved? The answer is already implicit in our fruit machine analogy. The quickest way to achieve the fruit machine target is to hold correct wheels while spinning the others to let chance find the target. By analogy, this is precisely what the immune system does within our bodies.

Figure 1 summarizes how this is achieved. Faced with a new antigen challenge, the mutation rate in the variable part of the genome can be accelerated by as much as 1 million times. So far as is known, those mutations all occur stochastically. But the location in the genome is certainly not a matter of chance. The functionality in this case lies precisely in the specific targeting at the relevant part of the genome. The mechanism is directed, because the arrival of the antigen itself activates the hypermutation process, and its binding to a successful antibody triggers proliferation of those cells that make it. What this mechanism achieves is that all the other 'wheels' in the DNA sequence forming a template for the immunoglobulin protein are held sufficiently constant for functionality to be retained. Even more remarkably, all the functionality in the rest of the genome is also maintained. Considering the huge size of the complete genome, this is pin-point targeting requiring highly specific feedback processes to be successful.

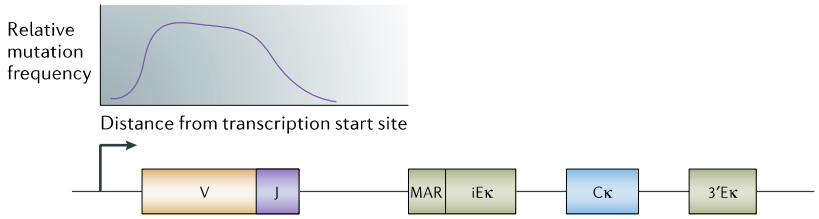


Figure 1. Schematic diagram of gene-specific targeted hyper-mutation in immunoglobulin gene loci. The mutation rate is greatly increased only in the variable part of the genome, which is a ~1.5 kilobase region in each of the three immunoglobulin loci. In this figure, the rectangular elements (V, J, MAR, iEκ, Cκ, 3′Eκ) represent different functional parts of the DNA sequence for the immunoglobulin protein. V is the variable part, subject to hypermutation, while the other parts are fixed. For further details on the functions of the parts see Odegard and Schatz [22]. Those details are not important for the purposes of this article.

3.2. Is the System Purposive?

Holding correct parts of the immunoglobulin sequence constant is the way rapid mutation can then be restricted to only very small *and relevant* parts of the whole genome. Hyper-mutation of all the immunoglobulin sequence, and even more so everywhere in the genome, would not work. As Odegard and Schatz say:

> "Somatic hypermutation (SHM) introduces mutations in the variable region of immunoglobulin genes at a rate of ~10^{-3} mutations per base pair per cell division, which is 10^6-fold higher than the spontaneous mutation rate in somatic cells. To ensure genomic integrity, SHM needs to be targeted specifically to immunoglobulin genes."

What this example shows is that the basic idea in Dawkins' Weasel program is actually broadly correct. Imagine that the monkey already has XYZHINKS IT IS LIKE A WEASEL. Then the best strategy is to treat only the XYZ sequence with stochastic mutation until MET turns up. Within the Weasel program analogy, it would be essential to hold the sequence HINKS IT IS LIKE A WEASEL constant.

At this point it is important to recall what we emphasized in the INTRODUCTION: Evolution is a high-level forming process, not simply a matter of genome informatics. The more correct way to look at the process therefore is that it is the high-level functionality that corresponds to HINKS IT IS LIKE A WEASEL and to any equivalently functional sequence that needs to be maintained. Any low-level sequence changes that are neutral with respect to phenotype functionality would not matter. The targeting may therefore be attributable to higher-level buffering by regulatory networks in addition to differential genome mutation rates. This point is important since not all the examples we discuss later in this article necessarily involve differential rates of mutation.

This is also why it is misleading to talk of the 'language of the genes' [23] or the 'book of life' [21]. In a language, the sequence *is* the written language's 'phenotype'. That is even more obvious in languages employing idiograms. By contrast, the genome is a template resource used by the organism, and is far from identical with or simply translatable into the phenotype.

The targeted mechanism in the immune system has been known and intensely studied for many years [24]. So, how did many people not realise that it is a physiologically guided process? The answer is that the guidance does not lie at the genome level. At the genome level the process appears blind. It depends on stochastic mutation. The functionality enabling the process to be described as guided lies in the system as a whole.

The system includes: (a) sensing the environmental challenge, i.e., the antigen invasion, (b) transmitting this signal to the nuclei of immune system cells to trigger hyper-mutation in just a tiny fraction of the genome. (c) Then sensing of the correctness or otherwise of the outcome, followed by the "reproduce or die" signal: cells that do not produce an antibody that fits the antigen do not survive. At this stage, natural selection occurs amongst the population of immune system cells [25]. This is a complete finely-tuned physiological feedback system that rapidly generates an acquired characteristic in response to an environmental challenge, which is then inherited in the surviving population of cells. This is what is *meant* by a goal-oriented system. By all the usual criteria this is a teleological, i.e., goal-directed, process (see Section 2).

It may not be perfect; it doesn't have to be. Not all keys have to be perfect to open a lock. The system feels its way forward, harnessing stochasticity to create novelty while using targeted preservation of what already works. The targeted preservation is what gives the system its purpose: to maintain its own integrity. It uses stochasticity to change what it must change, precisely because that is the part that doesn't work.

It is important moreover to see that the goal, the directionality, exists *within* the organisms and their populations. The goals of organisms have developed during the evolutionary process. Our position does not therefore require the ideas of Intelligent Design. In agreement with this aspect of Dawkins' position, we do not have to assume there is a long-term goal.

At this stage it is also important to clarify that we partly agree with alternative (such as neo-Darwinian) views of hypermutation mechanisms, to the extent of saying that such differential mutation rates must have evolved, and that the neo-Darwinian mechanism of stochastic variation combined with natural selection has operated [26–31]. The point to understand is that, once hypermutation has evolved and is linked to environmental feedback that endows the organism with natural purposiveness, subsequent evolution is not purely neo-Darwinian. Natural purposiveness evolves and then changes the nature of subsequent evolution. There is a transition, one of many transitions in evolution [32], the most spectacular of which has been the transition to enable cultural evolution leading to the development of humans, to which we will return in Section 6.

3.3. Natural Genetic Engineering

Such physiologically functional feedback leading to genomic change in response to an environmental challenge is not restricted to the immune system. In fact, responsiveness of the genome generally to environmental stress was discovered by the Nobel laureate, Barbara McClintock, more than 70 years ago. Working on Indian corn, she showed that in response to stress genetic material can move around even between different chromosomes [33]. She was therefore the discoverer of what are now called mobile genetic elements, known more colloquially as 'jumping genes'. In her 1983 Nobel Prize lecture she wrote:

> "In the future attention undoubtedly will be centered on the genome, and with greater appreciation of its significance as a *highly sensitive organ of the cell*, monitoring genomic activities and correcting common errors, sensing the unusual and unexpected events, and *responding to them, often by restructuring the genome*. We know about the components of genomes that could be made available for such restructuring. We know nothing, however, about how the cell senses danger and instigates responses to it that often are truly remarkable" (our italics). [34]

This was highly perceptive since it was written before whole genome sequencing. By 2001 with the publication of the first complete draft of the human genome, it became possible to compare genome sequences in different organisms. The results show that movements of whole domains of sequences corresponding to functional domains of transcription factor proteins and chromatin proteins must have occurred as evolution diverged to produce organisms as different as worms, yeast, flies, mouse and human [6,24,35].

Movement and rearrangement of functional domains of proteins can also function as a mechanism for speeding up evolutionary change. Like targeted hypermutation it also avoids having to wait for very slow accumulation of small (point) mutations. To appreciate this in less technical language, imagine two children given Lego bricks to construct a model bridge. To the first child we give a pile of the original small Lego bricks which have to be laboriously pieced together to form an architectural feature like an arch. To the second child we give some preformed Lego structures. It is obvious that the second child will construct a realistic bridge much faster than the first.

Moving complete functional domains around the genome is therefore a bit like the mirror image of hypermutation since it recombines already functional parts of proteins. In terms of the Weasel program, imagine already having METHINKS IT IS and LIKE A WEASEL. Joining them up is worth trying. Of course, not all joined up sequences will produce new functionality. What the mechanism gives is a much improved chance of obtaining new functionality. There is a bias in the process, which is precisely the extent to which it is not blind. It plays with existing functionality. As we will show later in this paper, behavioural evolution can use comparable mechanisms in which existing functionality is preserved and rearranged.

4. Goals within Populations

The cells of the immune system can evolve extremely rapidly to achieve the goal of the system, where the goal is the protection of the organism, and the system is the organism itself. But each organism does not transmit all of this information to its progeny. In this section we will look at ways in which populations of organisms can use stochastic mutation to evolve inheritable functional responses to environmental challenges very rapidly.

4.1. Contingency Loci in Bacteria

A comparable mechanism to that employed by the immune system has been extensively investigated by Richard Moxon and his colleagues who use the term 'contingency locus' to characterise the targeted loci of hypermutable DNA [36]. In bacteria, these loci are simple sequence repeats in which the repeating unit is one to several nucleotides in length. In eukaryotes these loci are called microsatellites and often consist of hundreds of repeats. In both kinds of organism these microsatellites are prone to high rates of mutation through slippage during strand repairing, leading to either increases or decreases in the number of repeat units. When these mutations occur within functional gene sequences, they can therefore produce high frequency reversible switching of genotype. Since "mutation rates vary significantly at different locations within the genome" they propose that "it is precisely in the details of these differences and how they are distributed that major contributions to fitness are determined." In an earlier article, Moxon and Thaler write

> "This phenotypic variation, which is stochastic with respect to the timing of switching but has a programmed genomic location, allows a large repertoire of phenotypic solutions to be explored, while minimizing deleterious effects on fitness." [37].

Moxon and Thaler's conclusion is correct. If the hyper-mutation were not restricted to a small subset of the genes, the results would certainly be deleterious, just as non-targeted hypermutation of immunoglobulin genes would rapidly destroy the functional proteins of the immune system.

The phenotype effects can also be combinatorial. The example given by Moxon et al. [36] is that switching in just seven independent loci to produce two genotypes in each case could generate up to 128 phenotypes. Many of these phenotype changes occur significantly in bacterial cell surface structure, which is the structure through which organisms detect changes in the environment and foreign invaders. They can also generate switching between metabolic and regulatory cell networks: Ritz et al. discovered a triplet repeat enabling *E. coli* to switch between adaptation to reducing and oxidizing environments [38].

Can such processes be demonstrated in actual evolutionary time? An example of organisms making use of this ability to reorganise their genomes is the study of Bos et al., who have observed the emergence of antibiotic resistance from multinucleated bacterial filaments. They write:

> "The strategy of generating multiple mutant chromosomes within a single cell may represent a widespread and conserved mechanism for the rapid evolution of genome change in response to unfavorable environments (i.e., chemo-therapy drugs and antibiotics)" [39].

Similarly, Jack et al. (2015) have shown that

> "Signaling pathways that sense environmental nutrients control genome change at the ribosomal DNA. This demonstrates that not all genome changes occur at random and that *cells possess specific mechanisms to optimize their genome in response to the environment*." (our italics) [40].

It is important at this stage in the argument to note that, in addition to the functional feedback, an essential property in purposive genome adaptation in response to environmental stress is the 'hold'

mechanism, by which existing functionality is preserved. This mechanism can operate whether or not hypermutation comparable to that in the immune system and many bacteria occurs. Hypermutation is simply an extreme example of the non-random location of mutations. Any differential mutation rate in genomes *might* be exploited by organisms and so improve their chances of generating new functionality, though equally clearly differential mutation rates alone do not *necessarily* indicate functionality. The relevant feedback loops with environmental interaction must also exist. That is an essential part of how a goal-oriented system is defined (see Section 2). The evolution of such links is a major transition.

4.2. Genetic Buffering by Regulatory Networks

Buffering of genome variation by regulatory networks may also be involved. This is a further important part of our argument so we have represented it in a development of Waddington's famous landscape diagram shown in Figure 2. As in the original diagram, genes (as DNA sequences) are represented by the pegs at the bottom. The regulatory networks are represented as lying between the genes and the phenotype. Genes can only influence the phenotype through the networks; they do not do so directly. They themselves do not exhibit agency in the sense in which we define it. This is one of the reasons why we believe the 'selfish gene' metaphor is misleading. Moreover, from a physiological perspective, selfish gene theory is not testable [41].

We have added two new features to the diagram. The first is the inclusion of environmental interactions above the phenotype landscape. The second is a cloud which we have placed to represent buffering of genomic change by the networks. Inspired by Hillenmeyer's work on yeast [4], we have represented the cloud as covering as much as 80% of the genome, which is the proportion of silent knockouts in Hillenmeyer's experiments, to which we have already referred in the INTRODUCTION. Similar robustness has been found in the networks involved in the natural pacemaker of the heart, where multiple mechanisms exist that can maintain rhythm if one is disabled either by genetic change or by pharmacological blockers [42].

Of course, that percentage will vary with different species, cell types and many other parameters. The buffering cloud may cover different proportions of genomic change under different environmental and experimental conditions. Thus, Hillenmeyer et al.'s experiments to which we referred earlier [4] also showed 97% of genes to be functional by varying metabolic conditions. The controlled experimental conditions in which 80% were silent are unlikely to match actual wild population conditions. The cloud is therefore dynamically variable, as its name suggests. Note also that the cloud is not a separate structure from the networks. It represents the filtering (buffering) action of the networks: a process of the networks, not an object in its own right.

The idea of the cloud mechanism was already implicit in Waddington's early work on what he called epigenetics [43] and has been confirmed by many physiological and developmental studies since then on the robustness of organisms in accommodating genomic variations, the most recent studies being the comparative failure of genome-wide association studies to reveal very much about the genetic origins of health and disease [44,45]. This is one of the most important empirical findings arising from genome sequencing. But its implications for evolutionary biology have not yet been sufficiently well appreciated.

The main implication that is relevant to this article is that the hold mechanism need not require targeted differential mutation rates. In the current stage of our knowledge we do not believe it is possible to estimate how often evolutionary change might depend on targeted differential mutation compared to the operation of regulatory networks protecting themselves via 'cloud' mechanisms. Those mechanisms are of course a further aspect of the robustness of regulatory networks in the face of genetic changes. Nor are the mechanisms only epigenetic. The phenomenon of epistasis by which different genes can influence each other's effects can play a similar role, particularly when the interactions, mediated through the networks, are to cancel each other's effects [46].

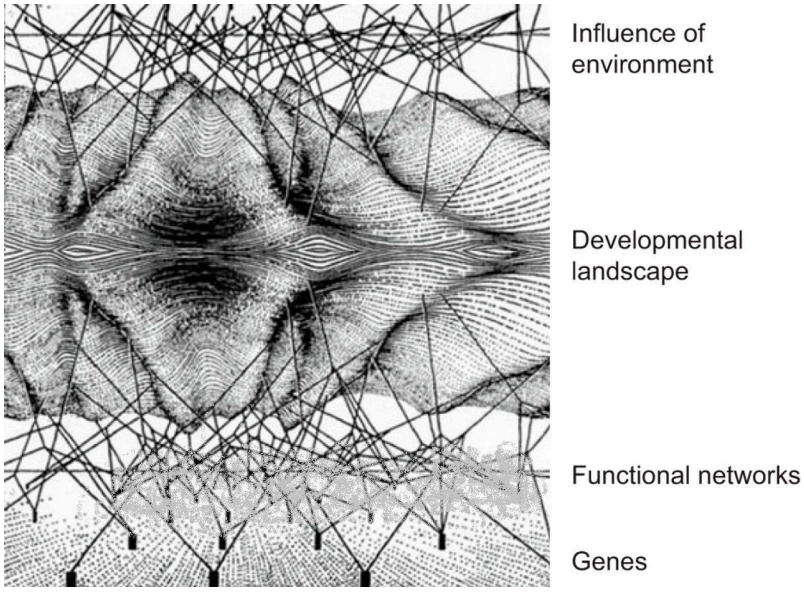


Figure 2. Development of Waddington's (1957) landscape diagram [43]. The original diagram was simply the lower half of this diagram, which Waddington used to indicate that the developmental phenotype (the landscape) is not directly dependent on the genes (the pegs at the bottom) but also depends on the regulatory networks represented as lying in between genes and the phenotype. Our version of the diagram incorporates two new features. First, organisms are open systems sensitive to the environment. This is represented by the top half of the diagram. Second, the regulatory networks in the lower half (the original half) of the diagram act to buffer genetic variation. This is represented by a 'cloud' covering a large fraction of the genome, corresponding to the fact that many mutations at the genome level are silent functionally. The regulatory networks can buffer many variations at the genome level. The filtering action of the 'cloud' performs a function similar to that of the 'hold' mechanism in this article. Differential mutation rates are not therefore essential to enable organisms to guide their own evolution.

4.3. Switch of Function in Regulatory Networks

Networks can not only act as buffers, they can also switch function. Our next example is from the work of Taylor et al. [47] who have shown that bacteria that have lost their flagella through deletion of the relevant DNA sequence can evolve the regulatory networks required to restore flagella and so restore motility in response to a stressful environment within just four days.

That ability is a property of the bacterium regulatory networks and of the ability of the organism to signal the environment pressure to those physiological networks to enable them to adapt. It is that feedback that makes such a rapid and clearly functional response possible. Two mutations were involved:

> "Step one mutations increase intracellular levels of phosphorylated NtrC, a distant homolog of FleQ, which begins to commandeer control of the FleQ regulon at the cost of disrupting nitrogen uptake and assimilation. Step two is a switch-of-function mutation that redirects NtrC away from nitrogen uptake and toward its novel function as a flagellar regulator. Our results demonstrate that natural selection can rapidly rewire regulatory networks in very few, repeatable mutational steps".

Viewed from the level of the genome site(s) where the mutations are occurring, there is a process of Darwinian selection amongst the results of stochastic mutation, as the authors themselves say. But the response clearly exploits the physiological regulatory properties of *existing* cell regulatory networks, resulting in a switch of function at the regulatory network level.

4.4. Roles of Stochasticity and Natural Selection

This is also a suitable point at which to emphasise that we are not denying the essential contribution of a neo-Darwinian process, i.e., mutation followed by natural selection. Natural selection necessarily operates within the context of these evolved functional characteristics. The key to our argument lies in the way in which organisms maintain and develop existing functionality *either* through exploiting differential mutation rates *or* through buffering the effects of mutations in many parts of the genome, *or* combinations of the two processes. It requires a multi-level systems-level approach to see that the neo-Darwinist process is harnessed, so that it is not sufficient in itself to explain the functionality of what is happening.

This point reflects once again our insistence that evolution is a high-level forming process. It may help to clarify that there are two senses in which this point has force. The first is conceptual. Even when a process is entirely neo-Darwinian when viewed from the genome sequence level, that viewpoint is not necessarily the most productive way to view it. Much more than the genome is inherited. The roles of regulatory networks, which are relatively well buffered against sequence changes, and the physiological feedback processes that ensure that the process leads to the inheritance of a characteristic in response to an environmental change, are better characterized from a physiological systems perspective. They also function within a cellular structural environment involving lipids and other components not coded for by the genome.

The second sense is empirical. It may, as a matter of fact, be the case that at the molecular level targeted differential mutation rates form an essential part of the process that explains the speed with which the adaptation occurs. In this article we have given examples of both of these senses.

4.5. Communication to the Genome

How can genomes know about what is happening at the cell surface? The physiological mechanisms by which events in tiny micro-domains near the cell surface signal to the nucleus, and so control specific gene expression levels, have now been studied in fine detail [48,49]. There is no longer any mystery in understanding the highly specific transmission of information to the nucleus that can control gene expression. There is no reason why genomes should not use similar communication pathways in response to stress signals received by cells and organisms.

5. Speculation 1: Goals Achieved through Epigenetic Inheritance

5.1. Different Forms of Epigenetics

Epigenetics was originally introduced and defined by Conrad Waddington [43] to refer to the role of networks in organisms in interpreting and controlling their genetic inheritance. Waddington was a developmental biologist and he correctly identified the general mechanisms by which development, in interaction with the environment and genes, could canalize both development itself, and subsequently also inheritance, towards specific phenotypes. His experiments on fruit flies showed that selection for environmentally-induced variants could become assimilated into the genome within a relatively small number of generations, as few as around 14 or 15. In so doing, he produced one of the first examples of the inheritance of acquired characteristics based on rigorous multigenerational experiments.

In addition to Waddington's mechanism of genetic assimilation of acquired characteristics, there was the discovery of transcription factors, i.e., proteins that convey signals to the genome from higher level networks to control levels of gene expression. This mechanism is what makes it possible for cells as different as bone cells and heart cells to be developed using the same genome. In vertebrates, around 200 clearly distinct cell types are produced in this way during development and in the maintenance of tissue types in the adult.

Recently, the mechanisms of epigenetic control of the genome have been greatly extended, through the discoveries of DNA marking, histone marking, and many processes by which the germline can be

either side-stepped [50] or itself marked, or modified by the transmission in the germline of functional RNAs [51,52].

5.2. Experimental Examples

These developments are transforming the study of genetics in evolution. As just one well-documented example, Michael Skinner and his colleagues have experimented on one of the icons of Darwinian evolution, the Galapagos finches, to find that there are as many epigenetic as genetic variations and that the number of epigenetic variations between the species correlate rather better with phylogenetic distance between them than do the genetic variations [53]. But it will be almost impossible to determine which came first in the evolution of the different species, since epigenetic and genetic changes necessarily interact. Moreover, some authors have highlighted the important role that 'soft' (epigenetic) inheritance can play in evolutionary change [54,55].

Given that organisms are active agents and can choose their behavior in response to the environment (including other organisms), they must be able to mark their genomes with variations as a consequence of their behavior. This is the mechanism described by Michael Meaney and his colleagues in showing how stroking behavior in rodents can mark the genome in the hippocampus to predispose the young to adopt the same affective behavior as adults [56].

Similarly, many life-time choices can now be shown to act epigenetically to influence health and disease in subsequent generations. Hanson and Skinner have written a valuable review of these effects [57]. They include all the evidence for environmentally-induced inheritable epigenetic impacts shown in Table 1, taken from their review.

Table 1. Environmental epigenetic impacts on biology and disease.

• Worldwide differences in regional disease frequencies
• Low frequency of genetic component of disease as determined with genome wide association studies (GWAS)
• Dramatic increases in disease frequencies over past decades
• Identical twins with variable and discordant disease frequency
• Environmental exposures associated with disease
• Regional differences and rapid induction events in evolution

Regional differences and the dramatic increases in disease frequencies are hard to explain without recourse to epigenetic mechanisms, just as identical twin studies show that genome differences alone are insufficient.

All of the epigenetic effects referred to in this section are well-documented experimentally (see Menger [58] and Skoblov et al. [59] for further examples). We come now to our first main speculation. If organisms have agency and, within obvious limits, can choose their lifestyles, and if these lifestyles result in inheritable epigenetic changes, then it follows that organisms can at least partially make choices that can have long-term evolutionary impact.

5.3. Role in Speciation

We refer to this as a speculation because, as Skinner says, it may be very difficult to disentangle epigenetic and genetic changes using estimations based on the differences between living species. Ideally, we would need to document both genetic and epigenetic changes as a function of time during the developments that led to the speciation. Given the long periods of time over which speciation occurs, it is difficult to see how such experiments could ever be performed. Like astronomers, we usually have to infer the past from what we observe now. Moreover, as Waddington showed, epigenetic changes can become assimilated into the genome: what begins as 'soft-wired' may become 'hard-wired'. One of the mechanisms by which such assimilation may occur is analysed in Noble, reference [6] (pp. 216–219) which is also the mechanism Waddington himself proposed. More generally

epigenetic changes produce changes in function and behaviour that lead to organisms choosing different niches. Genetic change can then follow. That process itself would usually 'hide' evidence of what initiated the change.

Exceptionally, the speciation process can be sufficiently rapid to be observed during very few generations. This is so for the remarkable case of an immigrant finch to one of the islands in the Galápagos archipelago which initiated a new genetic lineage by breeding with a resident finch. Genome sequencing of the immigrant identified it as a male that belonged to another species more than 100 km from the island. To quote the paper:

> "From the second generation onwards the lineage bred endogamously, and despite intense inbreeding, was ecologically successful and showed transgressive segregation of bill morphology. This example shows that reproductive isolation, which typically develops over hundreds of generations, can be established in only three." [60].

The rapidity with which this reproductively separate line was established is expected since one of the most important triggers for animal and plant speciation is interspecific hybridization. An important feature of hybrid germlines is disruption of normal epigenetic control which translates into increased genome restructuring and mobile element activity. Thus, in addition to forming novel combinations of genome components, hybridization triggers genome innovation by modification of a higher-level control regime, as outlined in more detail in the review by Shapiro [61].

6. Speculation 2: The Evolution of Goal-Directed Behaviour

The effect of behavioural control on evolutionary change could be especially great when the social environment is a major component of the challenges faced by animals. The result would be that individuals would evolve to understand and predict what other members of the social group are about to do [62].

6.1. The Continuity of Animal and Human Evolution

This article is itself part of the proof of what we are leading up to. We and the imagined monkeys at their keyboards in the Weasel example evolved from common hominoid ancestors several million years ago. They in turn evolved from the ancestors of all mammals, in turn from the pre-Cambrian fauna, and so on as we stretch back to the last eukaryotic ancestor, or even the last universal common ancestor, maybe 3 billion years ago. Unless we are to return to a theory of special creation, humans capable of writing this article form a complete evolutionary continuum with the whole of the rest of life on earth.

It is implausible to suppose that goal-directedness and creative purpose suddenly appeared only with the first humans [63]. It is therefore important to identify the roots of such mechanisms that have evolved in other organisms. If goal-directedness is identified as the ability to anticipate what other organisms may do and then act on that ability, then there is no serious difficulty in observing that many organisms in addition to humans can do this [10,62,64–68]. Animals like monkeys, dogs and wolves are sensitive to inequity in the behaviour of others [66,69] and so can favour the formation of cooperative groups. An important criterion is the ability to show unlimited associative learning, which is necessary for such anticipatory and innovative behavior. Bronfman et al. show that this can be observed in animals ranging from vertebrates to arthropods and cephalopods [70].

We contend that neo-Darwinism doesn't have the conceptual tools necessary to even begin to understand the transition to goal-directedness and creative purpose since it assumes a priori that these do not exist, they are only 'apparent'. By insisting on the necessarily exclusive role of blind chance in generating novelty it ignores the fact that such chance has to be, and has been, harnessed by organisms, which are therefore also *active* agents in their own development and evolution [62,71]. As Bateson says so succinctly:

> "the picture of the external hand of natural selection doing all the work is so compelling that it is easy to regard organisms as if they were entirely passive in the evolutionary process." (p. 105).

If we really insist on the passive role of the organism, we will not even recognize the significance of active agency in evolution. Assuming that natural selection is the only directing process is equivalent to denying that organisms have agency, or at least that, if they do, it does not play any role in their evolution. Darwin would not have agreed. He clearly identified the role of sexual selection in evolution [72].

6.2. The Adaptability Driver

Novelty, creativity, can only emerge if stochasticity is harnessed, rather than given free rein. Stochasticity everywhere is destructive, not constructive. But this requires the progressive building-in of mechanisms that include various forms of 'hold' when partial solutions have already been found. This is how organisms come to 'know' [73] what to keep and what to reject. Of course, all of this 'knowledge' is built on the processes of stochasticity and selection. Equally clearly, those processes on their own are insufficient. Furthermore, what becomes built-in includes much more than the genome since it also includes regulatory networks and 3D membranous systems that are also inherited and without which DNA is inactive. This is yet another example of our point that evolution is a high-level forming process, not simply a matter of genome informatics.

Can we therefore generalize the targeted mutation examples to apply not only to physical but also to behavioural evolution and the various forms of active role that organisms may play in their own evolution?

That is surprisingly easy. Targeted mutation allows stochasticity to be harnessed precisely because the organism's use of it is not blind. On the contrary, it is linked to higher-level processes that enable the organism to be an agent (see Section 2). The organism combines mutation with buffering of all that it is important to retain. Change is always combined with preservation. That ability is active in the sense that the knowledge required to identify what to preserve and what to change originates within the organism itself. How it originally evolved to have that knowledge is important, but not immediately relevant to the case being made here. Like the emergence of attractors in physical systems, once they have emerged, the clock can't be turned back.

Very early in the development of Darwin's theory of evolution it was noticed that organisms have the adaptability to choose or even create new niches for themselves and so to partially direct their own evolution. Darwin himself drew attention to the idea as it applied to sexual selection [72], where it is clearly true that choices of mate according to desired characteristics would influence future evolution. Once again, we quote Bateson on this aspect:

> "Charles Darwin (1871) argued that choice of a mate could drive evolution. He called the evolutionary process 'sexual selection'. Alfred Russel Wallace, although the co-author with Darwin of the first clear statement about the role of Natural Selection, did not like the new idea. Indeed, for many years most biologists did not take sexual selection seriously. When I was an undergraduate I was told confidently that, even if it were possible in theory, the process probably played little part in biological evolution. In recent years, however, many experiments have supported Darwin's thinking." [67].

The generalization of the idea of organism choice in evolution was originally attributed to Baldwin and became known as the Baldwin effect [74,75]. Bateson has researched the history and development of the idea [76]. The process was first identified by Douglas Spalding in 1873 [77]. To avoid historical confusion and to let the name of the process be easier to understand it, Bateson chose to call it the *adaptability driver*. We like this nomenclature for two reasons. First, adaptability is the behavioural equivalent of the targeted mutation process. The key lies in the restriction of change only to *parts* of

the behavioural repertoire. Second, the word 'driver' captures the active role of the organism itself. Blind chance is then not the only driver of novelty.

6.3. The Role of Contextual Logic in the Behavior of Organisms

We will conclude this article with a brief sketch of how contextual logic enters into the behaviour of organisms so that they do not react purely passively, but rather become active agents in interaction with their environment. From an evolutionary perspective, organisms have become active agents behaviourally because there are obvious advantages in being able to anticipate the behavior of other organisms. For a predator, anticipating the behavior of prey may be the difference between lunch and no lunch. Conversely, for prey, to be able to anticipate the behaviour of a predator could be the difference between death and survival. Such anticipation assumes rule-guided behavior by organisms for prediction and anticipation to be possible. What results is like a game. Iterative game-playing is clearly not unique to humans.

We may then extend the analogy with the processes of targeted genome mutation and re-organisation outlined earlier in this article. Winning games, including those between predator and prey, depends precisely on a combination of preservation and change. On the one hand, organisms learn the repetitive rules of interaction. The better those rules are known, the more effective will be the organism's anticipation of the behavior of others. On the other hand, innovation requires the ability to break out from the rules, to foil the anticipation of others. We speculate that this is where stochasticity plays a role similar to that of the harnessing of stochasticity in targeted genome variation. Organisms that can combine rule-guided anticipation with occasional innovative behavior will have a selective advantage. Our speculation is that selective harnessing of stochasticity enables innovation, just as it enables targeted genome variation, but the benefits depend also on combining innovation with conservation of routine behavior.

Once organisms have acquired such ability, they become active agents, and all the well-known evolutionary consequences of the 'adaptability driver' then follow.

7. Conclusions

7.1. The Harnessing of Stochasticity

In this article we have outlined mechanisms by which organisms and populations harness stochasticity and so improve their chances of developing functional responses to environmental challenges. Provided that we correctly interpret the targeted nature of genome variation and its necessary correlate, i.e., the preservation of already functional genome sequences and/or the buffering of genome change by the regulatory networks, the principle of Dawkins' Weasel program becomes broadly correct. The 'hold' mechanism corresponds to the spatial targeting of mutation or the buffering of many genetic variations by the active networks: all other sequences and network properties are sufficiently well preserved ('held') to maintain functionality.

Some evolutionary biologists have attributed what we describe as an active 'hold' mechanism to a passive 'ratchet' mechanism [32,78–81]. A process now known as Muller's Ratchet was introduced in 1964 following a series of papers in which he explored the evolutionary significance of sex. He showed mathematically that the damage that would result from accumulation of deleterious mutations can be ameliorated by recombination, either by exchange of DNA between organisms or by recombination through sexual reproduction. Muller's Ratchet idea captures a form of blind directionality (ratchets go forwards not backwards) in evolution but it does not capture the agency of organisms themselves which is implied by the 'hold' metaphor.

Maynard Smith and Szathmary included the ratchet mechanism in their 1995 book on *The Major Transitions in Evolution* [32]. A further advance in this idea was described by Lukeš et al. [79] to demonstrate how a ratchet process which, following Stoltzfus [82] they term Constructive Neutral Evolution can generate cellular complexity. They come close to our idea of the functional significance

of genetic buffering (represented as the cloud in Figure 2) when they write "The interaction (between two biochemical components, A and B), though not under selection, permits (suppresses) mutations in A that would otherwise inactivate it." Our reading of "permits (suppresses)" is precisely the cloud mechanism. While their paper is an undoubted advance on the ratchet idea, it does not include a targeting of the process of mutation, nor does it include feedback from and to the environment.

This is the reason why functionally significant 'hold' mechanisms forming parts of purposive feedback systems should not be interpreted simply as a 'ratchet' mechanism. It is precisely the targeted nature of the complete physiological feedback system *using* the 'hold' mechanism that generates functionality and gives the subsequent evolutionary process a direction, driven by organisms themselves. To repeat what we wrote earlier in Section 3, "this is what is *meant* by a goal-oriented system."

7.2. Organisms as Agents

Yet, our interpretation of the correctness of the Weasel program uses a mechanism that Dawkins did not himself acknowledge. The reason is that neo-Darwinists generally eschew the concepts of goal-directedness or teleology. The idea of active agency in organisms runs counter to neo-Darwinist thought. We suggest that this view is motivated partly by what are perceived as the dangers of misinterpretation of teleological explanations, which are thought to support the ideas of creationism and intelligent design. But attributing agency and directedness to organisms themselves makes no commitment whatsoever to ideas of long-term goals in evolution. We therefore believe that this concern is misplaced. In any case, the scientific investigation of goal-directed behavior should not be restricted by perceived opportunities for misinterpretation.

This also explains why classical neo-Darwinism [15,16], and modern versions of it [1,8] completely excluded the inheritance of acquired characteristics (see Section 2). Such inheritance is sometimes thought to open the door to theories of long-term external directionality in evolution. Confusingly also, this kind of directionality is often associated with Lamarck and his concept of "le pouvoir de la vie" [83]. This phrase was mistakenly interpreted to become identified with 'vital force', which is a serious misreading of Lamarck's writing [84,85]. Lamarck was a thorough-going materialist and was opposed to the vitalists of his time. As the French historian of genetics André Pichot writes:

> 'Lamarck's claim that . . . there is a radical difference between living beings and inanimate objects might lead people to think that he was a vitalist. But he is not. On the contrary, his biology is a mechanistic reply to the physiological vitalism of Bichat, which was then the dominant theory' [86]. (Our translation of Pichot's French).

Lamarck's "pouvoir de la vie" would therefore be better interpreted as an innate tendency in organisms to evolve in a directed way [87]. This is precisely what the mechanisms described in this article achieve. The directionality is innate in organisms and populations as active agents.

We have also shown that some forms, at least, of behavioural evolution can be interpreted within the same scheme of harnessed stochasticity combined with preservation of existing functional forms.

In an important article analyzing many aspects of this issue in the context of a review of Dawkins' *The Extended Phenotype*, Eva Jablonka covers some of the points we make here, including this quotation:

> Being scared of Lamarckism leads to the neglect of the evolutionary effects of evolved systems that allow the inheritance of targeted and acquired variations. When the fact that variation is highly constrained and is shaped (or, rather, drafted) by the rules of the generating system is ignored, evolution cannot be properly understood [88].

To which we can add that this neglect is inhibiting the development of a fully integrated physiological (i.e., functional) interpretation of evolutionary biology [6]. Some of the problems with neo-Darwinist interpretations arise from conceptual limitations, including privileging causation from the molecular genetic level, for which there is no empirical justification [89]. One of the aims of our

article is to encourage empirical investigations of the ideas we put forth. Theories of goal-directedness can be tested, and engineers and physiologists have the tools to do so.

7.3. Organisms and Their Populations Are the One-Eyed Watchmakers

We return now to where this article began: *The Blind Watchmaker* and the monkey at the keyboard. The Watchmaker analogy is a powerful one, both for those who follow Paley's original use of it to argue for creationism or intelligent design, and for Dawkins' use of it to argue for nothing but blind chance. But consider this: the only watchmakers we know are organisms (humans). They evolved from other organisms. The ability to be a watchmaker therefore evolved. There is therefore nothing surprising in the fact that goal-directed agency occurs also in other organisms and is capable of influencing evolution.

Our overall conclusion is that there are several processes by which directed evolutionary change occurs—targeted mutation, gene transposition, epigenetics, cultural change, niche construction and adaptation. Evolution is an ongoing set of iterative interactions between organisms and the environment. Evolution is a continuous organic process. Directionality is introduced by the agency of organisms themselves as the one-eyed watchmakers. Evolution itself also evolves [90].

Acknowledgments: We would like to acknowledge Sir Anthony Kenny's suggestion that led to the title of this article. The kind of direction that organisms and populations exert over their own evolution is best represented as 'feeling a way forward'. The 'sight' involved is always partial. The metaphor 'one-eyed', which Kenny suggested to us, nicely captures the essence of this idea. We are grateful to Sir Patrick Bateson, Eva Jablonka, Perry Marshall, David Miller, Richard Moxon, Anant Parekh, Dan Rubenstein, James Shapiro and Michael Yudkin for kindly criticizing drafts of the paper. We acknowledge in particular the books we have consulted by James Shapiro, Eva Jablonka and Patrick Bateson, as well as the papers of Richard Moxon, all of which are to be found in the reference list. We acknowledge also valuable criticisms and comments from three anonymous journal referees. While this article was in revision we learnt the sad news that Patrick Bateson had passed away. We dedicate this article to a great evolutionary biologist who championed the agency of organisms and from whom we learnt a lot.

Conflicts of Interest: The authors declare no conflict of interest.

References and Notes

1. Dawkins, R. *The Blind Watchmaker*; Norton & Company: New York, NY, USA, 1986.
2. The Time Required Would in Fact Require Billions More Periods of Time Equivalent to the Whole Duration of the Universe. Available online: https://en.wikipedia.org/wiki/Weasel_program (accessed on 6 July 2017).
3. Natarajan, C.; Hoffmann, F.G.; Weber, R.E.; Fago, A.; Witt, C.C.; Storz, J.F. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **2016**, *354*, 336–339. [CrossRef] [PubMed]
4. 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. The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* **2008**, *320*, 362–365. [CrossRef] [PubMed]
5. Noble, D. A theory of biological relativity: No privileged level of causation. *Interface Focus* **2012**, *2*, 55–64. [CrossRef] [PubMed]
6. Noble, D. *Dance to the Tune of Life: Biological Relativity*; Cambridge University Press: Cambridge, UK, 2016.
7. Charlesworth, D.; Barton, N.H.; Charlesworth, B. The sources of adaptive variation. *Proc. R. Soc. B* **2017**, *284*, 20162864. [CrossRef] [PubMed]
8. Dawkins, R. *The Selfish Gene*; OUP: Oxford, UK, 1976, 2006.
9. Coyne, J.A. *Why Evolution is True*; OUP: New York, NY, USA, 2010.
10. Bateson, P. The active role of behaviour in evolution. *Biol. Philos.* **2004**, *19*, 283–298. [CrossRef]
11. Noble, D. Evolution viewed from physics, physiology and medicine. *Interface Focus* **2017**, *7*, 20160159. [CrossRef] [PubMed]
12. Prindle, A.; Liu, J.; Asally, M.; Ly, S.; Garcia-Ojalvo, J.; Suel, G.M. Ion channels enable electrical communication in bacterial communities. *Nature* **2015**, *527*, 59–63. [CrossRef] [PubMed]
13. Oettmeier, C.; Brix, K.; Döbereiner, H.-G. *Physarum polycephalum*—A new take on a classic model system. *J. Phys. D* **2017**, *50*, 413001. [CrossRef]

14. Monod, J.; Jacob, F. Teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **1961**, *26*, 389–401. [CrossRef] [PubMed]
15. Weismann, A. *The Germ-Plasm: A Theory of Heredity*; Charles Scribner's Sons: New York, NY, USA, 1893.
16. Weismann, A. *Die Allmacht der Naturzüchtung; Eine Erwiderung an Herbert Spencer*; Fischer: Jena, Germany, 1893.
17. Spencer, H. *The Inadequacy of "Natural Selection"*; University of Michigan Library: Ann Arbor, MI, USA, 1893. Reprinted in *The Principles of Biology*; London, D. Appleton and Company: New York, NY, USA, 1897.
18. Mitchell, P.C. The Spencer-Weismann Controversy. *Nature* **1894**, *49*, 373–374. [CrossRef]
19. Winther, R.G. August Weismann on Germ-Plasm Variation. *J. Hist. Biol.* **2001**, *34*, 517–555. [CrossRef] [PubMed]
20. Segerstrale, U. Neo-darwinism. In *Encyclopedia of Evolution*; Pagel, M., Ed.; Oxford University Press: Oxford, UK, 2002; Volume 2, pp. 807–810.
21. Bodmer, W.; McKie, R. *Orion: The Book of Man: The Quest to Discover Our Genetic Heritage*; Scribner: New York, NY, USA, 1995.
22. Odegard, V.H.; Schatz, D.G. Targeting of somatic hypermutation. *Nat. Rev. Immunol.* **2006**, *8*, 573–583. [CrossRef] [PubMed]
23. Jones, S. *The Language of the Genes*; Harper-Collins: London, UK, 2000.
24. Shapiro, J.A. *Evolution: A View from the 21st Century*; Pearson Education Inc.: Upper Saddle River, NJ, USA, 2011.
25. For a Relatively Simple Account of How All This Is Achieved. Available online: https://www.scientificamerican.com/article/how-do-white-blood-cells/ (accessed on 6 July 2017).
26. Kimura, M. On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* **1967**, *9*, 23–34. [CrossRef]
27. Moxon, E.R.; Rainey, P.B.; Nowak, M.A.; Lenski, R.E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **1994**, *4*, 24–33. [CrossRef]
28. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **2010**, *26*, 345–352. [CrossRef] [PubMed]
29. Chen, X.; Zhang, J. No gene-specific optimization of mutation rate in *Escherichia coli*. *Mol. Biol. Evol.* **2013**, *30*, 1559–1562. [CrossRef] [PubMed]
30. Martincorena, I.; Luscombe, N.M. Non-random mutation: The evolution of targeted hypermutation and hypomutation. *BioEssays* **2013**, *35*, 123–130. [CrossRef] [PubMed]
31. Fitzgerald, D.M.; Hastings, P.J.; Rosenberg, S.M. Stress-induced mutagenesis: Implications in cancer and drug resistance. *Ann. Rev. Cancer Biol.* **2017**, *1*, 119–140. [CrossRef]
32. Maynard Smith, J.; Szathmáry, E. *The Major Transitions in Evolution*; Oxford University Press: Oxford, UK, 1995.
33. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **1950**, *36*, 344–355. [CrossRef] [PubMed]
34. McClintock, B. The significance of responses of the genome to challenge. *Science* **1984**, *226*, 792–801. [CrossRef] [PubMed]
35. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [CrossRef] [PubMed]
36. Moxon, R.; Bayliss, C.; Hood, D. Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **2006**, *40*, 307–333. [CrossRef] [PubMed]
37. Moxon, E.R.; Thaler, D.S. The tinkerer's evolving tool-box. *Nature* **1997**, *387*, 659–662. [CrossRef] [PubMed]
38. Ritz, D.; Lim, J.; Reynolds, C.; Poole, L.; Beckwith, J. Conversion of a perixiredoxin into a disulphide reductase by a triplet repeat expansion. *Science* **2001**, *294*, 158–160. [CrossRef] [PubMed]
39. Bos, J.; Zhang, Q.; Vyawahare, S.; Rogers, E.; Rosenberg, S.M.; Austin, R. Emergence of antibiotic resistance from multinucleated bacterial filaments. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 178–183. [CrossRef] [PubMed]
40. Jack, C.V.; Cruz, C.; Hull, R.M.; Ralser, M.; Houseley, J. Regulation of ribosomal DNA amplification by the tor pathway. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9674–9679. [CrossRef] [PubMed]
41. Noble, D. Neo-darwinism, the modern synthesis, and selfish genes: Are they of use in physiology? *J. Physiol.* **2011**, *589*, 1007–1015. [CrossRef] [PubMed]

42. Noble, D. Differential and integral views of genetics in computational systems biology. *Interface Focus* **2011**, *1*, 7–15. [CrossRef] [PubMed]
43. Waddington, C.H. *The Strategy of the Genes*; Allen and Unwin: London, UK, 1957.
44. Callaway, E. Genome studies attract criticism. *Nature* **2017**, *546*, 463. [CrossRef]
45. Boyle, E.A.; Li, Y.; Pritchard, J.K. An expanded view of complex traits: From polygenic to omnigenic. *Cell* **2017**, *169*, 1177–1186. [CrossRef] [PubMed]
46. Breen, M.S.; Kemena, C.; Vlasov, P.K.; Notredame, C.; Kondrashov, F.A. Epistasis as the primary factor in molecular evolution. *Nature* **2012**, *490*, 535–538. [CrossRef] [PubMed]
47. Taylor, T.B.; Mulley, G.; Dills, A.H.; Alsohim, A.S.; McGuffin, L.J.; Studholme, D.J.; Silby, M.W.; Brockhurst, M.A.; Johnson, L.J.; Jackson, R.W. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science* **2015**, *347*, 1014–1017. [CrossRef] [PubMed]
48. Kar, P.; Mirams, G.R.; Christian, H.C.; Parekh, A.B. Control of nfat isoform activation and nfat-dependent gene expression through two coincident and spatially segregated intracellular Ca^{2+} signals. *Mol. Cell* **2016**, *64*, 746–759. [CrossRef] [PubMed]
49. Ma, H.; Groth, R.D.; Cohen, S.M.; Emery, J.F.; Li, B.; Hoedt, E.; Zhang, G.; Neubert, T.A.; Tsien, R.W. Γcamkii shuttles Ca^{2+}/CAM to the nucleus to trigger creb phosphorylation and gene expression. *Cell* **2014**, *159*, 281–294. [CrossRef] [PubMed]
50. Weaver, I.C.G. Life at the interface between a dynamic environment and a fixed genome. In *Mammalian Brain Development*; Janigro, D., Ed.; Humana Press, Springer: New York, NY, USA, 2009; pp. 17–40.
51. Rechavi, O.; Minevish, G.; Hobert, O. Transgenerational inheritance of an acquired small RNA-based antiviral response in c. Elegans. *Cell* **2011**, *147*, 1248–1256. [CrossRef] [PubMed]
52. Tollefsbol, T. *Transgenerational Epigenetics: Evidence and Debate*; Academic Press: Waltham, MA, USA, 2014.
53. Skinner, M.K.; Gurerrero-Bosagna, C.; Haque, M.M.; Nilsson, E.E.; Koops, J.A.H.; Knutie, S.A.; Clayton, D.H. Epigenetics and the evolution of darwin's finches. *Genome Biol. Evol.* **2014**, *6*, 1972–1989. [CrossRef] [PubMed]
54. Soen, Y.; Knafo, M.; Elgart, M. A principle of organization which facilitates broad lamarckian-like adaptations by improvisation. *Biol. Direct* **2015**, *10*, 68. [CrossRef] [PubMed]
55. Herman, J.J.; Spencer, H.G.; Donohue, K.; Sultan, S.E. How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution* **2013**, *68*, 632–643. [CrossRef] [PubMed]
56. 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. Epigenetic programming by maternal behavior. *Nat. Neurosci.* **2004**, *7*, 847–854.
57. Hanson, M.; Skinner, M. Developmental origins of epigenetic transgenerational inheritance. *Environ. Epigenet.* **2016**. [CrossRef] [PubMed]
58. Menger, F.M. Molecular lamarckism: On the evolution of human intelligence. *World Futures* **2017**, *73*, 89–103. [CrossRef]
59. Skoblov, M.Y.; Scobeyeva, V.A.; Baranova, A.V. The mechanisms of transgenerational inheritance and their potential contribution to human phenotypes. *Russ. J. Genet.* **2016**, *52*, 249–256. [CrossRef]
60. Lamichhaney, S.; Han, F.; Webster, M.T.; Andersson, L.; Grant, B.R.; Grant, P.R. Rapid hybrid speciation in darwin's finches. *Science* **2017**. [CrossRef] [PubMed]
61. Shapiro, J.A. Epigenetic control of mobile DNA as an interface between experience and genome change. *Front. Genet.* **2014**, *5*, 87. [CrossRef] [PubMed]
62. Bateson, P. *Behaviour, Development and Evolution*; Open Book Publishers: Cambridge, UK, 2017.
63. Curiously, neo-Darwinists sometimes seem to be forced to concede a view that humans are unique. Consider this statement from *The Selfish Gene*: "Let us try to teach generosity and altruism, because we are born selfish. Let us understand what our own selfish genes are up to, because we may then at least have the chance to upset their designs, *something that no other species has ever aspired to*." (chapter 1, our emphasis).
64. Brosnan, S.F.; De Waal, F.B. Monkeys reject unequal pay. *Nature* **2003**, *425*, 297–299. [CrossRef] [PubMed]
65. Brosnan, S.F. A hypothesis of the co-evolution of cooperation and responses to inequity. *Front. Neurosci.* **2011**. [CrossRef] [PubMed]
66. Essler, J.L.; Marshall-Pescini, S.; Range, F. Domestication does not explain the presence of inequity aversion in dogs. *Curr. Biol.* **2017**, *27*, 1861–1865. [CrossRef] [PubMed]
67. Bateson, P. New thinking about biological evolution. *Biol. J. Linnean Soc.* **2013**. [CrossRef]

68. Daniel Rubenstein. Research on Decision Making in Animals Is a Major Field of Zoology. Available online: https://www.princeton.edu/~dir/ (accessed on 6 July 2017).
69. Leimgruber, K.L.; Rosati, A.G.; Santos, L.R. Capuchin monkeys punish those who have more. *Evol. Hum. Behav.* **2016**, *37*, 236–244. [CrossRef]
70. Bronfman, Z.Z.; Ginsburg, S.; Jablonka, E. The transition to minimal consciousness through the evolution of associative learning. *Front. Psychol.* **2016**, *7*, 1954. [CrossRef] [PubMed]
71. Corning, P.A. Evolution 'on purpose': How behaviour has shaped the evolutionary process. *Biol. J. Linnean Soc.* **2014**, *112*, 242–260. [CrossRef]
72. Darwin, C. *The Descent of Man, and Selection in Relation to Sex*; John Murray: London, UK, 1871.
73. We have put 'know' in parentheses to indicate that this does not necessarily imply conscious knowledge. In a general sense of 'know' there can be no doubt that the word is correct. We as humans 'know' how to control our blood pressure even though we also know that this knowledge is unconscious).
74. Baldwin, J. A new factor in evolution. *Am. Nat.* **1896**, *30*, 441–451. [CrossRef]
75. Avital, E.; Jablonka, E. *Animal Traditions. Behavioural Inheritance in Evolution*; Cambridge University Press: Cambridge, UK, 2000.
76. Bateson, P. The adaptability driver: Links between behaviour and evolution. *Biol. Theory* **2006**, *1*, 342–345. [CrossRef]
77. Spalding, D.A. Instinct. With original observations on young animals. *Macmillan's Mag.* **1873**, *27*, 282–293.
78. Bridgham, J.T.; Ortlund, E.A.; Thornton, J.W. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* **2009**, *461*, 515–519. [CrossRef] [PubMed]
79. Lukeš, J.; Archibald, J.M.; Keeling, P.J.; Doolittle, W.F.; Gray, M.W. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* **2011**, *63*, 528–537. [CrossRef] [PubMed]
80. Doolittle, W.F. Evolutionary biology: A ratchet for protein complexity. *Nature* **2012**, *481*, 270–271. [CrossRef] [PubMed]
81. Mast, F.D.; Barlow, L.D.; Rachubinski, R.A.; Dacks, J.B. Evolutionary mechanisms for establishing eukaryotic cellular complexity. *Trends Cell Biol.* **2014**, *24*, 435–442. [CrossRef] [PubMed]
82. Stoltzfus, A. Constructive neutral evolution: Exploring evolutionary theory's curious disconnec. *Biol. Direct* **2012**, *7*, 35. [CrossRef] [PubMed]
83. Literally, "the power of life". If alive today, Lamarck would surely align himself with those who accept the active agency of organisms. In his time, he vigorously opposed the creationism of Cuvier and could not therefore be interpreted as a creationist.
84. Noble, D. Letter from lamarck. *Physiol. News* **2010**, *78*, 31.
85. Noble, D. *The Music of Life*; OUP: Oxford, UK, 2006.
86. Pichot, A. Introduction. In *Philosophie Zoologique*; Flammarion: Paris, France, 1994.
87. For important modern assessments of Lamarck and Lamarckism see Gissis SB, Jablonka, E. (2015) Transformations of Lamarckism. From subtle fluids to molecular biology. MIT Press).
88. Jablonka, E. From replicators to heritably varying phenotypic traits: The extended phenotype revisited. *Biol. Philos.* **2004**, *19*, 353–375. [CrossRef]
89. Noble, D. Evolution beyond neo-darwinism: A new conceptual framework. *J. Exp. Biol.* **2015**, *218*, 7–13. [CrossRef] [PubMed]
90. Noble, D.; Jablonka, E.; Joyner, M.M.; Müller, G.B.; Omholt, S.W. Evolution evolves: Physiology returns to centre stage. *J. Physiol.* **2014**, *592*, 2237–2244. [CrossRef] [PubMed]

INTERFACE FOCUS

rsfs.royalsocietypublishing.org

Review

Cite this article: Noble D. 2017 Evolution viewed from physics, physiology and medicine. *Interface Focus* **7**: 20160159.
http://dx.doi.org/10.1098/rsfs.2016.0159

One contribution of 20 to a theme issue 'New trends in evolutionary biology: biological, philosophical and social science perspectives'.

Subject Areas:
biophysics, biocomplexity, systems biology

Keywords:
evolution and physiology, Schrödinger's error, biological relativity, stochasticity, neo-Darwinism, modern synthesis

Author for correspondence:
Denis Noble
e-mail: denis.noble@dpag.ox.ac.uk

Evolution viewed from physics, physiology and medicine

Denis Noble

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

DN, 0000-0002-3013-3694

Stochasticity is harnessed by organisms to generate functionality. Randomness does not, therefore, necessarily imply lack of function or 'blind chance' at higher levels. In this respect, biology must resemble physics in generating order from disorder. This fact is contrary to Schrödinger's idea of biology generating phenotypic order from *molecular*-level order, which inspired the central dogma of molecular biology. The order originates at higher levels, which constrain the components at lower levels. We now know that this includes the genome, which is controlled by patterns of transcription factors and various epigenetic and reorganization mechanisms. These processes can occur in response to environmental stress, so that the genome becomes 'a highly sensitive organ of the cell' (McClintock). Organisms have evolved to be able to cope with many variations at the molecular level. Organisms also make use of physical processes in evolution and development when it is possible to arrive at functional development without the necessity to store all information in DNA sequences. This view of development and evolution differs radically from that of neo-Darwinism with its emphasis on blind chance as the origin of variation. Blind chance is necessary, but the origin of functional variation is not at the molecular level. These observations derive from and reinforce the principle of biological relativity, which holds that there is no privileged level of causation. They also have important implications for medical science.

1. Introduction: the original formulation of the neo-Darwinist modern synthesis

The theory of evolution by natural selection was formulated by Charles Darwin and Alfred Russel Wallace who presented their ideas to the Linnean Society of London in 1858, followed by Darwin's book *On the Origin of Species* in 1859. Darwin was cautious in the presentation of his ideas. He wrote 'Natural Selection has been the main, but not the exclusive means of modification'. He was concerned that he did not know the origin of variation and he acknowledged the existence of other mechanisms, including the inheritance of acquired characteristics. Ernst Mayr wrote in 1962: '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' [1]. Although Darwin disagreed with Lamarck on whether evolution had a direction (what Lamarck called *le pouvoir de la vie* [2,3]), he nevertheless acknowledged 'this justly celebrated naturalist . . . who upholds the doctrine that all species, including man, are descended from other species' [4]. However, Darwin's multi-mechanism approach to evolution became significantly narrowed with the rise of neo-Darwinism.

Weismann's formulation of neo-Darwinism involved three major assumptions. First, that all genetic variation is random. Second, that the germline is isolated from variations in the soma. This is the Weismann barrier. Third, together with these two assumptions, that natural selection is then all-sufficient (*allmacht*) to explain evolution [5]. The subsequent integration of Mendelian genetics into this scheme led to the formulation of the modern synthesis [6].

THE ROYAL SOCIETY PUBLISHING

Several important consequences followed. First, genetic variation is not itself viewed as functional. It becomes so only through the operation of natural selection to weed out harmful variations and promote helpful ones. The origin of variation is therefore completely blind. If this view is correct, we should not explain genetic variation in terms of existing or anticipated functionality. As physiology is the study of functional processes in organisms, physiology is thereby excluded from any direct role in the source of variation. Second, the inheritance of acquired characteristics, often called Lamarckism, cannot occur because it would require either that the germ line is not isolated from influences of somatic variations and/or that some forms of functional genetic reorganization can be triggered as a response to environmental stress. In an 1896 publication [7], Weismann added his theory of germinal selection, involving competition and selection among the hereditary units *within* the germplasm but, as Charlotte Weissman shows, this change in Weismann's view did not make any real concessions to the Lamarckians [8].

The neo-Darwinist modern synthesis was therefore both an extension and a simplification of Darwin's ideas. It was an extension through the incorporation of Mendelian genetics, about which Darwin unfortunately knew nothing. It was a simplification because it excluded the inheritance of acquired characteristics, whereas Darwin not only included this form of inheritance, he even proposed a theory for how it could happen, his pangenesis theory of gemmules [9], which resembles some forms of such inheritance discovered recently (see §6).

2. Purpose of this article

A central thesis of this paper is that blind stochasticity is a misconceived idea as it has been used in evolutionary biology. Stochasticity is used by organisms to generate new functional responses to environmental challenges. Far from proving that evolution is necessarily blind, randomness is the clay from which higher level order can be crafted. But it necessarily works the other way too: higher levels then organize the molecular level through many forms of constraint. The reason we do not necessarily see that organization from the molecular level is that the difference of scale is vast. If we focus on particular molecular events, such as gene mutations at particular loci, they will still appear stochastic. Blind chance can then seem to be the sole determinant of variation even when, in fact, the variation is directed in response to environmental challenges.

I will present the case for the following theses, which run counter to neo-Darwinism and the modern synthesis. With respect to neo-Darwinism, the view in this paper is a replacement more than an extension.

1. Randomness (stochasticity) is what one should generally expect at the molecular level even if determinate functionality rules at higher (cellular, tissue, organ, systems, organisms, sociological) levels. Randomness and functionality necessarily coexist at different levels.
2. Organisms can and do harness stochasticity in generating function.
3. Functional genome reorganization can occur in response to environmental stress.
4. Non-DNA information can be transmitted across generations.
5. By using diverse higher level processes, organisms can resist potentially harmful effects of many random genetic variations, at lower levels of function.
6. Physical constraints can and must influence both development and evolution.
7. The gene-centric view has so far been very disappointing from the viewpoint of medicine.

3. Stochasticity and order coexist at different levels

Physics teaches us that at a molecular level, there must be stochasticity. At any temperature above a value near absolute zero, below which a Bose–Einstein condensate becomes possible [10], molecules have kinetic energy which generates random movement. But physics also teaches us that, once there is a constraint at a higher level, e.g. a gas in a container, thermodynamics can describe determinate behaviour arising from the averaged behaviour within the constraint. This is the reason why Schrödinger argued correctly in *What is life?* that physics generates order from disorder [11].

Yet he contrasted this with biology, which he described as generating order at a high level from *order* at a molecular level, i.e. that the functional order at a high level actually results directly from order at the molecular level. But this is highly problematic from a physical viewpoint. Why then did he propose a theory that even he initially characterized as ridiculous? The reason is that following Delbrück [12], he predicted that the genetic material would be found to be an aperiodic crystal, which is a good description of DNA sequences if one thinks of a polymer as a kind of crystal. Crystal structure can be investigated accurately using diffraction. I believe he saw the 'read-out' of genetic sequences as determinate in the same kind of way. In this respect, he anticipated the formulation of Crick's central dogma of molecular biology [13]. Francis Crick and James Watson both acknowledged Schrödinger's influence in their thinking about the central dogma.

There are two fatal problems with this approach, as noted by Kupiec [14,15]. The first is that, as is clear from Crick's original statement, the central dogma refers only to the fact that *sequence* information passes one way, from DNA to proteins:

> The central dogma of molecular biology deals with the detailed residue-by-residue transfer of sequential information. It states that *such information* cannot be transferred back *from protein* to either protein or nucleic acid. [16, p. 561]

I have italicized '*such information*' and '*from protein*' because it is evident that the statement does not say that *no* information can pass from the *organism* to the genome. In fact, it is obvious that it must do so to produce many different patterns of gene expression, which enable many different phenotypes (e.g. many different cell types in the same body) to be generated from the same genome. In addition to controlling relative expression levels, the organism also makes use of protein-mediated protein processing to add yet another layer of control following transcription.

This information from organisms is conveyed to their genomes by patterns of transcription factors, genome marking, histone marking and many RNAs, which in turn control the patterns of gene expression. These controls are exerted through preferential targeted binding to the

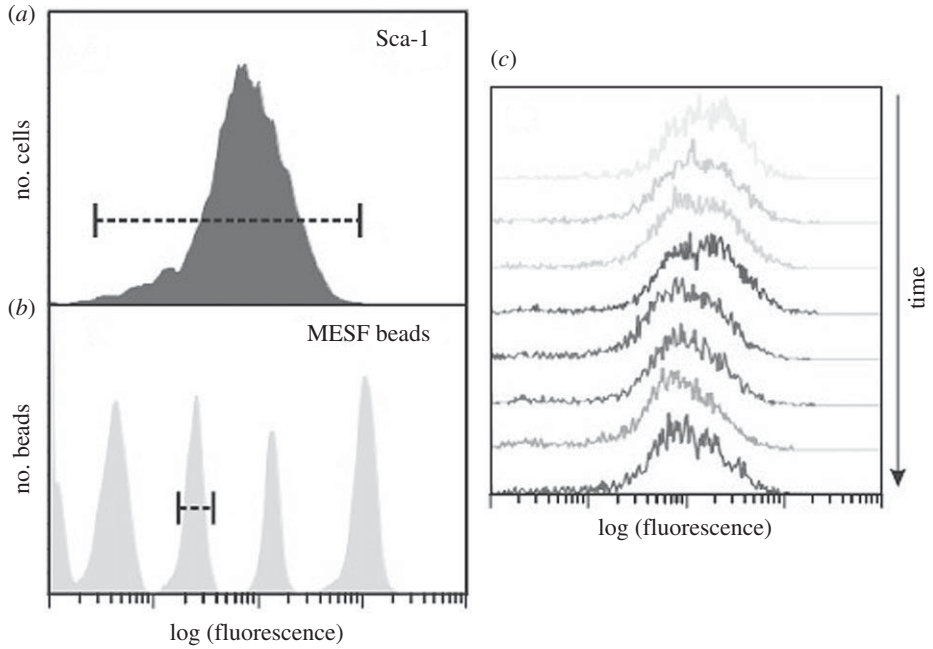


Figure 1. The robustness of heterogeneity of expression of Sca-1 protein expression in a cloned cell population. Heterogeneity detected by immunofluorescence flow cytometry (*a*) was significantly larger than the resolution limit of the method (*b*). (*c*) The stability of the clonal heterogeneity over a period of three weeks. Note that the spread of gene expression levels is three orders of magnitude [21].

genome or histone proteins. For example, methylation of cytosines preferentially occurs at CpG sites. Binding to histones preferentially occurs at the histone tails. Even though these are the targeted molecular mechanisms by which the functional control is exerted, there is no guarantee that the functionality will be evident at the molecular level. It would require many correlations between the *patterns* of binding and the functional processes at a higher level to identify the functionality involved. Without that correlation, the binding patterns will appear random. There are simply far too many sites. There are millions of CpG sites in the whole genome and tens of thousands of CpG clusters, which significantly are located near gene regulatory sites [17].

The second problem is that, as Schrödinger must have understood as a physicist, there is no way in which the molecules in an organism can avoid stochasticity. He wrote:

> We seem to arrive at the ridiculous conclusion that the clue to understanding of life is that it is based on a pure mechanism, a 'clock-work' in the sense of Planck's paper. [18, p. 101]

But he then confuses the logic by continuing: 'The conclusion is not ridiculous and is, in my opinion, not entirely wrong, but it has to be taken "with a very big grain of salt"'. He then explains the 'big grain of salt' by showing that even clock work is, 'after all statistical' (p.103). This seriously compromises the logic because the stochasticity in clockwork has to be negligible. We now know that the stochasticity in biology is far from negligible.

Schrödinger realizes that something is far from right but is struggling to identify what it might be. We would now say that the molecules involved (DNA) *are* subject to frequent statistical variations (copying errors, chemical and radiation damage, etc.), which are then corrected by the cell's protein and lipid machinery that enables DNA to become a highly reproducible molecule [19]. This is a three-stage process that reduces the copy error rate from 1 in 10^4 to around 1 in 10^{10}, which is an astonishing degree of accuracy. In a genome of 3 billion bp, this works out as less than 1 error in copying a complete genome, compared to millions of errors without error correction. The order at the molecular scale is therefore actually created by the system as a whole, including lipid components that are not encoded by DNA sequences [20]. This requires energy, of course, which Schrödinger called negative entropy. Perhaps therefore this is what Schrödinger was struggling towards, but we can only see this clearly in retrospect. He could not have known how much the genetic molecular material experiences stochasticity and is constrained to be highly reproducible *by the organism itself*. The order at the molecular (DNA) level is actually imposed by higher level constraints.

4. Organisms can and do harness stochasticity in generating function

4.1. Stochasticity is a population-level attractor

Experiments on the stochasticity of gene expression in cell populations show that, at least in some cases, it is the population as a whole that controls the stochasticity. Figure 1 is taken from Chang *et al.* [21].

The results show that in this case, the range of gene expression is 1000-fold and it follows a simple bell-shaped curve. The range is a population-level attractor, which is stable over long periods of time. That the population controls the heterogeneity is shown by experiments of the kind illustrated in figure 2. In a cell population showing a bimodal distribution, new populations of cells were cloned from one of the peaks (left), while in a monomodal distribution, cells were cloned from outliers. In both cases, after a few days, the original heterogeneity became re-established.

Cell populations can therefore control stochasticity.

4.2. Cells can harness stochasticity to generate function

That cells can also harness stochasticity to generate specific function is known from experiments on the cells of the immune system that show the phenomenon of somatic

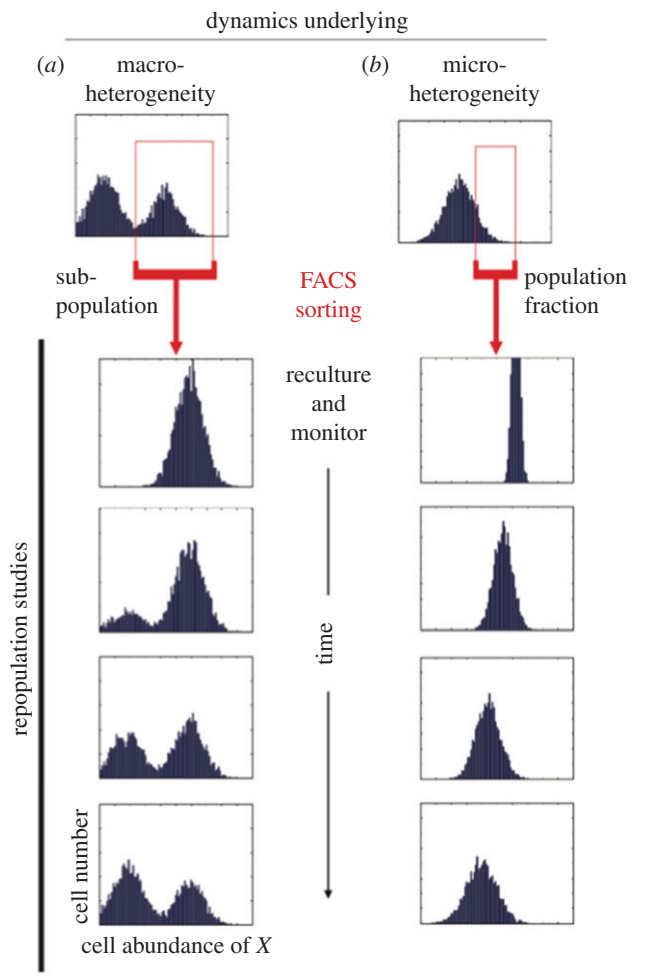


Figure 2. Two examples illustrating experiments in which populations were produced by cloning either from one of the peaks in a bimodal distribution (*a*) or from outliers in a monomodal distribution (*b*). In both cases, the new population initially exhibits the range of expression of the parent subpopulation. Over time (several days), however, the heterogeneity reverts to the original distribution [22]. (Online version in colour.)

hypermutation. Figure 3 summarizes what we know. Faced with a new antigen challenge, the mutation rate in the variable part of the genome can be accelerated by as much as 1 million times. So far as we know, the mutations occur randomly. But the location in the genome is certainly not random. The functionality in this case lies precisely in the targeting of the relevant part of the genome. The mechanism is directed, because the binding of the antigen to the antibody itself activates the proliferation process.

This example from the immune system shows that functionally significant targeted hypermutation can occur in the lifetime of an individual organism. There is no reason why this kind of mechanism should not be used in evolutionary change, as shown in the next example.

A well-known functionally driven form of genome change is the response to starvation in bacteria. Starvation can increase the targeted reorganizations of the genome by five orders of magnitude, i.e. by a factor of over 100 000 [24,25]. This is one of the mechanisms by which bacteria can evolve very rapidly and in a functional way in response to environmental stress.

A similar targeting of location where genomic change can occur has been found in experiments on genetically modified fruit flies. One of the common ways in which genetic modification is achieved is to use a particular kind of mobile genetic element that can move around the genome using a cut-and-paste mechanism that does not require an RNA intermediate. Most often, the insertions occur in a random way. But when DNA sequences from certain regulatory regions are used, they get inserted preferentially near the gene from which the sequence was derived [26]. This process targets the changes in a way that is clearly not random with respect to possible function.

5. Functional genome reorganization can occur in response to stress

5.1. Barbara McClintock and the genome as an organ of the cell

Barbara McClintock first observed that whole domains of genetic material move around the genome, even from one chromosome to another. She was working on Indian corn in the 1930s and 1940s, but it was much later, in 1983, that she was recognized with the award of a Nobel Prize. In her Prize lecture, she was very clear about the functional significance of her discovery. She described the genome 'as a highly sensitive organ of the cell, monitoring genomic activities and correcting common errors, sensing the unusual and unexpected events, and responding to them, often by restructuring the genome' [27].

She could not have anticipated the extent to which her idea would be confirmed by the sequencing of whole genomes. From the 2001 *Nature* paper on the first draft sequence of the human genome, we have comparisons between sequences in completely different species of eukaryotes for two classes of proteins, transcription factor proteins and chromatin binding proteins [28]. These show that the evolution of these proteins must have involved the movement of whole functional domains. This is far from the idea of slow progressive accumulation of point mutations. And it has much greater evolutionary significance because the rearrangement of whole domains including the functionality of those domains in response to stress could have been the origin of creativity in the evolutionary process. It is obvious that combining two or more domains each of which already has functionality is much more likely to produce a viable solution to a problem than waiting for random sorting of point mutations. This is why McClintock characterized the genome as a *highly sensitive organ of the cell*.

5.2. Can we observe genome reorganization happening in evolutionary experiments?

We can now observe organisms making use of this ability to reorganize their genomes. Bos *et al.* have observed the emergence of antibiotic resistance from multi-nucleated bacterial filaments. They write:

> the strategy of generating multiple mutant chromosomes within a single cell may represent a widespread and conserved mechanism for the rapid evolution of genome change in response to unfavorable environments (i.e. chemo-therapy drugs and antibiotics). [29, p. 182]

Jack *et al.* [30] have shown that

> signaling pathways that sense environmental nutrients control genome change at the ribosomal DNA. This demonstrates that not all genome changes occur at random and that *cells possess*

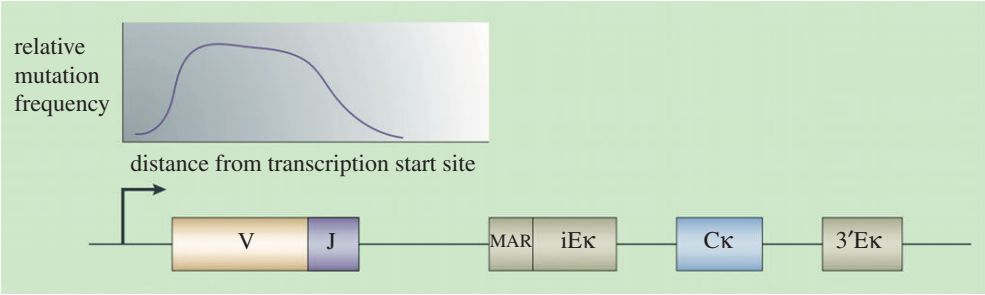


Figure 3. Schematic diagram of gene-specific targeted hypermutation in immunoglobulin gene loci. The mutation rate is greatly increased only in the variable part of the genome, which is an approximately 1.5 kb region in each of the three immunoglobulin loci. In this figure, the graph above the rearranged variable (V) and joining (J) gene segments that form the variable region of Igκ depicts the mutation domain in the κ-light chain (Igκ) locus. 3′Eκ, Igκ 3′ enhancer; Cκ, Igκ constant; iEκ, Igκ intronic enhancer; MAR, matrix attachment region [23]. (Online version in colour.)

> *specific mechanisms to optimize their genome in response to the environment*. (my italics) [30, p. 9674]

How can genomes know about what is happening at the cell surface? The physiological mechanisms by which events in tiny micro-domains near the cell surface signal to the nucleus to control specific gene expression levels have now been studied in fine detail [31,32]. There is no longer any mystery in understanding the highly specific transmission of information to the nucleus that can control gene expression. There is no reason why genomes should not use similar communication pathways in response to stress signals received by cells and organisms.

6. Non-DNA information can be transmitted across generations

Recent experiments have demonstrated that non-DNA information can be transmitted between generations [33], and this rapidly growing field has been reviewed in an important paper in *Science* [34]. Two quotations from that review are relevant:

> Many phenomena and mechanisms of nongenetic and/or non–DNA sequence–based inheritance have been described in a range of model organisms, challenging our perception of the well-established relationship between transmitted genotype and phenotype. [34, p. 59]

They conclude

> The idea of certain sequences that might be refractory to germline epigenetic reprogramming provides a compelling mechanism for the inheritance of modulated epigenetic states. [34, p. 63]

To illustrate the range of processes that can be involved, I will briefly describe three examples.

Rechavi *et al.* [35] investigated the inheritance of resistance to viral infection in the nematode worm, *Caenorhabditis elegans*. The resistance is acquired when infected worms have the DNA required to make a viral-silencing RNAi, which is triggered by viral replication. They cross-bred these worms with a wild-type population, including worms that do not have the required DNA. Some of the later generations have the required DNA, others do not. Yet subsequent generations inherited the acquired silencing response irrespective of whether they had the required DNA. The RNAi is inherited through the germline, and is then amplified by RNA polymerase in each generation. This non-DNA inheritance was followed successfully for 100 generations. It resembles Darwin's gemmule theory (see Introduction).

Nelson *et al.* [36] found robust inheritance of epigenetic marking in mice with Apobec1 deficiency. They found that 'these [epigenetic] effects persist for many generations and are as strong as conventional genetic inheritance'. The journal, *PNAS*, published a commentary article in the same issue, which concludes: 'the belief that the soma and germline do not communicate is patently incorrect' [37].

The question whether epigenetic transmission of acquired characteristics could have been responsible for the evolution of separate species has been answered by Skinner *et al.* [38] who investigated the DNA mutations and non-DNA epigenetic changes in one of the icons of Darwinian speciation, the Galapagos finches. Five species were studied with different phylogenetic distances between them. Figure 4 shows the results. Both DNA mutations and epigenetic variations increase with the phylogenetic distance, with the epigenetic changes correlating better with distance. The authors conclude that both changes were involved in speciation and that they must have interacted.

7. Organisms can resist the harmful effects of many molecular-level variations

One of my own fields of research is cardiac rhythm and arrhythmias. The main pacemaker in the heart, the sinus node, is an example of a robust functional process. Several different ionic transporter circuits are involved, any one of which could generate rhythm. The evolutionary advantage of this situation is obvious: if one mechanism fails, another can take over the function. In 1992, we investigated this robustness by reverse engineering an experimentally based computer model. We found that removing a transporter that could carry as much as 80% of the ionic current necessary for generating the rhythm would change the overall frequency by only around 10–15% [39]. Reverse engineering studies using a physiological model reveals the mechanism of the substitution. The small voltage changes that occur when one component is knocked out are sufficient to activate the substituting mechanism. This discovery formed the basis of the development of a safe heart slowing medication, ivabradine [40].

This kind of 'back up' of important physiological functions is ubiquitous. A systematic study of gene knockouts in yeast showed that 80% of knockouts have little or no effect on physiological functions under normal physiological conditions [41]. Metabolic stress was needed to reveal the functional roles of most of the genes involved.

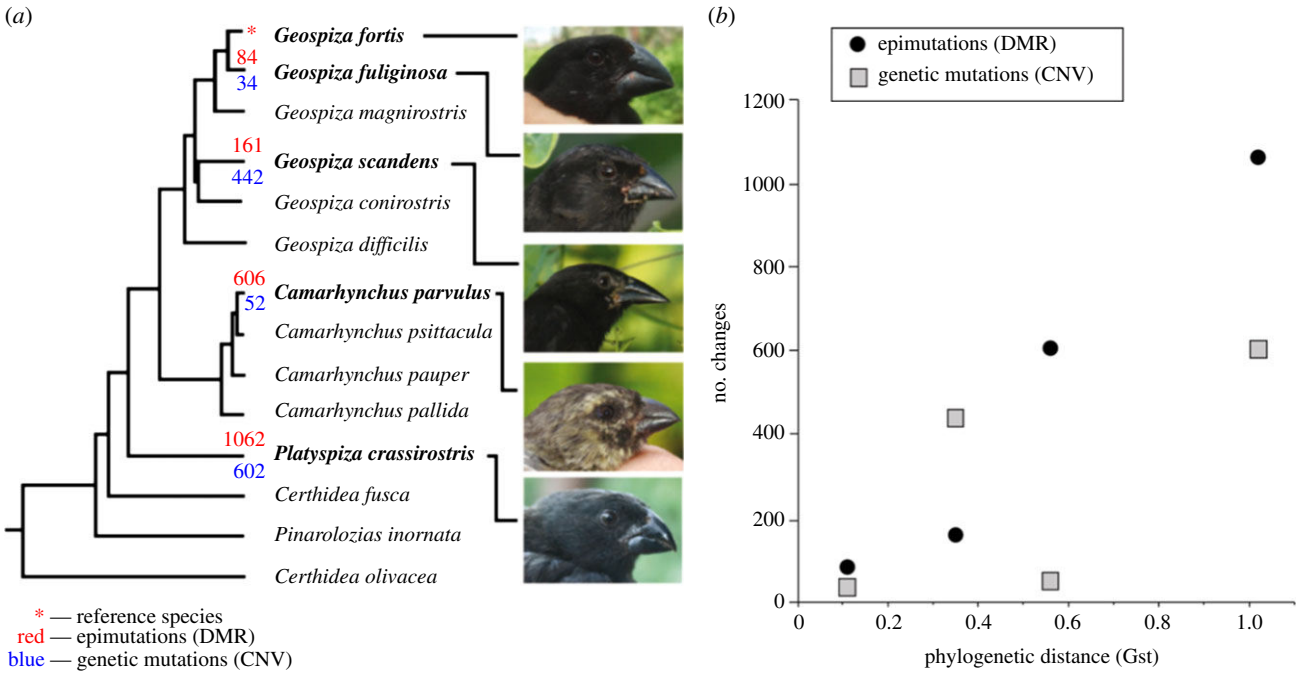


Figure 4. (*a*) Five of the Galapagos finch species were studied, the reference species *Geospiza fortis* and four others. The graph in (*b*) shows the number of genetic and epigenetic changes plotted as a function of phylogenetic distance. The epigenetic changes correlate well with phylogenetic distance, the genetic mutations do not correlate as strongly [38]. (Online version in colour.)

These studies pose a serious problem for bottom-up gene-centric theories of biology. The functionality will simply not be seen at that level or may be far from quantitatively accurate. Organisms seem to be very resourceful when challenged with knockouts, blockers or absence of nutrients. If we look for that ingenuity at the molecular level, we may not find it.

Again, we can ask the question whether such processes can be demonstrated in actual evolutionary time. This was done recently by Taylor *et al.* [42] who have shown that bacteria that have lost their flagella through deletion of the relevant DNA sequence can evolve the regulatory networks required to restore flagella and so restore motility in response to a stressful environment within just 4 days. Specifically, Taylor *et al.* show that deletion of FleQ (Flagellar transcriptional regulator) in *Pseudomonas fluorescens*, and starvation of the bacteria, produces mutations that enable the regulatory role to be taken over by a different pathway, normally involved in nitrogen uptake and assimilation. The genes required to produce flagellae are then reactivated by the new regulatory pathway. The authors interpret their work as showing how selection can rapidly produce this kind of substitution to restore activation of flagella genes. But, equally clearly, the mutations are targeted in a remarkably precise way. They are not randomly occurring anywhere in the genome. This example is therefore somewhat comparable to the cardiac pacemaker example I discuss earlier in this section, in that one network takes over the lost function when another network is no longer functional. That ability is a property of the bacterium regulatory networks and of the ability of the organism to signal the environment pressure to the genome to activate mutation.

It is important to note that such examples, and the earlier ones I quoted above in §5, involve what, so far as we know, are random mutations. At each location on the DNA sequence level, this will therefore appear as 'blind' variation. At that level, there will also be a form of Darwinian selection operating [14]. But the targeting of particular locations, which is what enables the response to the environmental challenge to be effective, is not blind. Nor does targeting necessarily require differential mutation rates in the genome. Buffering of non-functional genome changes by regulatory networks can also ensure the preservation of existing functionality, just as the regulatory networks involved in cardiac rhythm can ensure insensitivity to molecular-level changes, as I described at the beginning of this section.

Differential mutation rates have been extensively investigated by Moxon *et al.* [43] who use the term 'contingency locus' to characterize the targeted loci of hypermutable DNA. In bacteria, these loci are simple sequence repeats in which the repeating unit is one to several nucleotides. In eukaryotes, these loci are called microsatellites and often consist of hundreds of repeats. As 'mutation rates vary significantly at different locations within the genome', they propose that 'it is precisely in the details of these differences and how they are distributed that major contributions to fitness are determined'. In an earlier article, Moxon & Thaler [44] write 'This phenotypic variation, which is stochastic with respect to the timing of switching but has a programmed genomic location, allows a large repertoire of phenotypic solutions to be explored, while minimizing deleterious effects on fitness'.

8. Physical constraints can and must influence both development and evolution

Natarajan *et al.* [45], in a paper significantly entitled 'Predictable convergence in hemoglobin function has unpredictable molecular underpinnings', have examined the molecular basis of convergence in haemoglobin function involving 56 avian taxa that have contrasting altitudinal range limits. They found that 'Convergent increases in hemoglobin–oxygen affinity were pervasive among high-altitude taxa,

but few such changes were attributable to parallel amino acid substitutions at key residues. Thus, predictable changes in biochemical phenotype do not have a predictable molecular basis'. This article beautifully illustrates the main point I am making in this paper, which is that unpredictability at the molecular level, which would lead one to think the changes are random, can be perfectly compatible with predictability and functionality at a higher level. This is biology's equivalent of the physical principle that determinate thermodynamics can coexist with unpredictable stochastic behaviour at a molecular level. The difference is that, in biological systems, through the process of evolution, the higher level becomes functional. That is the level at which the functionality can be seen. It is then the level from which the lower level stochasticity can be understood, including the functional constraints.

If physics can be so important by using stochasticity in convergent evolution, can it also be important in a similar way in constraining development? It is tempting to think so because early embryonic development is similar in all multicellular eukaryotes, despite many differences in genome sequences. Edelman *et al.* [46] have explored this question by showing graphically how some simple physical constraints might be sufficient to explain certain aspects of embryonic development without having to assume that there must always be a specific DNA basis for all such processes. Their images are speculative and would require computational modelling to develop and test the ideas. Stuart Newman, Santiago Schnell and Philip Maini have led the way on this approach [47,48]. There must be interaction between overall physical constraints and molecular-level specifications. Ehrlich *et al.* [49] show how modelling such physical constraints can account for the evolution of shell form in ammonites.

These examples illustrate a general point. Nature does not need to write to the 'hard disc' of the organism, its DNA, when it can get functions automatically from physical 'free rides', i.e. by letting physics do what it will do naturally. There is no need for DNA to be involved, for example, in ensuring that lipid membranes naturally fuse and form vesicles and many of the other properties of thin oily bilayers. And, of course, there is no DNA forming templates for the wide variety of lipids in organisms.

9. The gene-centric view has so far been very disappointing from the viewpoint of medicine

There is another field of science where focusing on the molecular level has blinded us to functional processes at higher levels. That is the field of medicine. But before I explain why that is the case, I want to make it quite clear that I fully recognize the great scientific value of genome sequencing.

Sequencing whole genomes has been of immense value in evolutionary biological studies. The benefits for phylogeny and in discovering new parts of the 'trees' or 'networks' of life are obvious. It was sequencing that enabled Carl Woese to make his fundamental discovery of the archaea and how they differ from bacteria and eukaryotes [50]. Sequencing also enabled us to identify the extent to which mobile genetic elements must have been involved in the evolution of many proteins. In this sense, describing the genome as the 'book of life' has been a useful metaphor. But, as a metaphor used to publicize the health benefits that would accrue from genome sequencing it has been misinterpreted. The promise was that by a decade or so following sequencing of the human genome, the 'book of life' would reveal how to treat cancer, heart disease, nervous diseases, diabetes and many others through the discovery of many new pharmaceutical targets. This did not happen. An editorial in *Nature* in 2010 spelt this out:

> But for all the intellectual ferment of the past decade, has human health truly benefited from the sequencing of the human genome? A startlingly honest response can be found on pages 674 and 676, where the leaders of the public and private efforts, Francis Collins and Craig Venter, both say 'not much'. [51]

The targets were identified all right. At least 200 new possible pharmaceutical targets are now known and there may be more to come, but we simply do not understand how to use them. The problem does not therefore lie in the absence of knowledge about the sequences. The problem is that we neglected to do the relevant physiology at the higher levels. A valuable critique of genotype–phenotype relations as a basis for the common disease–common variant hypothesis has been published by Joyner & Prendergast [52].

Before the shift towards genomic approaches to pharmacology, we did in fact have reasonably adequate methods for developing new drugs against specific diseases. The method was to work initially at a phenotype level to identify possible active compounds, and then to drill down towards individual protein or other molecular targets. This was the approach used so successfully by Sir James Black, the Nobel laureate discoverer of β-blockers and H2 receptor blockers [53]. It is the method by which the work of collaborators in my laboratory eventually led to the successful heart drug, ivabradine, to which I have already referred.

But the consequence of diverting large-scale funding towards the search for new drugs via genomics has been that the Black approach is now much less common and that the pharmaceutical industry is producing fewer new medications at vastly greater cost. Of course, the Black approach could and should be complemented by genomics, and there are successful cases where protein targets found by classical methods were later also identified as coding templates formed by particular genes. A good example is Duchenne muscular dystrophy, where the gene for the protein utrophin that can substitute, in mice at least, to cure the disease was discovered before the DNA sequence was identified [54].

10. Conclusion

There has been much debate about whether the neo-Darwinist modern synthesis needs extending or replacing. Both views are correct. It depends on the context in which they are assessed. Theories in biology, as in any branch of science, can be judged by several criteria.

10.1. Falsifiability

The original neo-Darwinist assumptions of the modern synthesis have been clearly falsified. I will consider the three basic assumptions outlined in the Introduction.

10.2. The Weismann barrier

The Weismann barrier should be seen as a relative not an absolute barrier. Strict isolation of the genome was required

in order to exclude the inheritance of acquired characteristics. As we now know that acquired characteristics can be inherited, I believe it is more honest to admit that this reason for departing from Darwinism is no longer valid. In any case, the barrier could only apply in those organisms that have a separate germ line. For the great majority of the duration of life on the Earth, there was no separate germ line. And plants can reproduce separately from their germ line. Quite simply, then, two of the original basic assumptions, isolation of the germ line and the impossibility of inheritance of acquired characteristics, can be seen to be incorrect.

Some criticisms of this conclusion refer to the rarity of experiments showing intergenerational transmission of epigenetic mutations. Originally, this was based on the idea that the genome was always wiped clean of epigenetic marking, so that it was thought that the idea was misconceived and impossible. As I have shown, this is simply not correct.

Another criticism was that it would not be robust. It has been demonstrated to persist for as many as 100 generations, and that it can, in some cases, be as robust as DNA transmission. Moreover, it does not need to be robust in all cases. As the review by Burggren [55] shows, the softness and therefore reversibility of epigenetic inheritance is one of its evolutionary virtues. Sultan and co-workers [56] have also identified the factors that may determine the transience or persistence of epigenetic variation.

The third criticism is that it is observed in only rare cases. My reply is that so is speciation. Speciation is such a rare event that in thousands of years of selective breeding of cats, dogs, fish, etc., we have not succeeded in producing new species, as defined by reproductive isolation.

Note also that these criticisms obviously do not apply to functionally significant reorganization or hypermutation of genomes.

10.3. Blind stochasticity

The other basic assumption is blind stochasticity, meaning that what are seen as random genetic variations are not functionally directed. The concept of randomness is a major topic of research in philosophy, mathematics and physics. One way to by pass these highly technical issues is to ask the question 'random with respect to what'? The key in relation to evolutionary biology is whether variations are random with respect to function and whether they can be seen to be so. Even if the molecular-level variations do in fact represent functional order at a higher level, we will almost certainly require insight from the functional level to appreciate the functional nature of the molecular variations. The randomness I am referring to is therefore epistemological: without knowing the constraints by higher levels, the variations will *appear* to be random and unpredictable. Once we know those constraints the possibility of prediction at the molecular level begins to exist. Whether it is computable is a very different question. Given the huge differences of scale, e.g. between molecular and cellular, it is implausible to expect molecular-level computation alone to reveal the functionality.

Even before we consider whether a theory based on blind stochasticity has been falsified, we have to examine its conceptual status. A very basic lesson from physics is that stochasticity at lower, such as molecular, levels is not only inevitable as a consequence of molecular kinetic energy, it is also perfectly compatible with regular law-like behaviour at higher levels, a fact that was appreciated long ago by one of the founders of population genetics, Fisher [57]. Even if behaviour at a high level is directed, stochasticity is what we can expect at lower levels. The example in this paper concerning the evolution in different species of haemoglobins at high altitude illustrates that point perfectly. As the authors of that paper say 'predictable changes in biochemical phenotype do not have a predictable molecular basis' [45]. It is the physics of oxygen transport in organisms living at low partial pressures of oxygen that dictates the changes that occur to adapt to such environments, not specific changes in the genome.

From a gene-centric viewpoint, it could be objected that the genome changes are nevertheless those that enable the beneficial changes in oxygen transport to happen. That is certainly true. But it is precisely the higher level perspective that enables us to show that fact. What we can see here is that a conceptual issue, which is the question of the level at which functionality occurs, interacts with an empirical issue, which is whether the changes at the molecular level are predictable, *from that level alone*. Another way to put the conceptual issue is to say that, in any information transmission system, whether languages or genomes, sequences by themselves do not have meaning. They acquire meaning through their context, which can only be understood at a much higher level. As a linguistic example, the three letter alphabetic sequence 'but' has two totally different meanings and pronunciations in English and French. Similarly, genome sequences acquire meaning in their context. Sequences enabling arms, legs and eyes derive from organisms that had none of these.

10.4. Unravelling the problem

My paper unravels this problem by showing where some aspects of biological thought went wrong in the twentieth century. Schrödinger's book, *What is life?*, was a landmark in predicting correctly that the genetic material would be found to be an aperiodic crystal. But it contained the seeds of a major misunderstanding, leading Schrödinger, and then Crick and Watson, to maintain that, like a crystal, the genetic material could be read in a determinate way. That could be true only if the 'crystal', that is the linear polymer DNA, could be read and copied faithfully, with few or no copy errors. As we can now see, that is not an inherent property of DNA alone. On the contrary, it is a property of the complex system by which the copy error rate can be reduced from an unacceptable frequency of millions per genome to less than 1. That is a higher level systems property of cells, including an army of proteins and lipids, not of DNA alone. In life as we know it on the Earth, this process occurs only in the context of living cells.

A possible objection to this conclusion is that all proteins have DNA templates that determine their amino acid sequences. That includes the proteins that contribute to the error-correcting systems for DNA. That is true, but it is usually taken a step further to mean that therefore the genome determines everything. That is not true. The error-correcting systems operate within cellular structures that contain molecular elements, such as lipid membranes, that do not require DNA templates in order to exist. Elsewhere, I have shown that the structural information in cells can be represented as comparable to that in the genome [58].

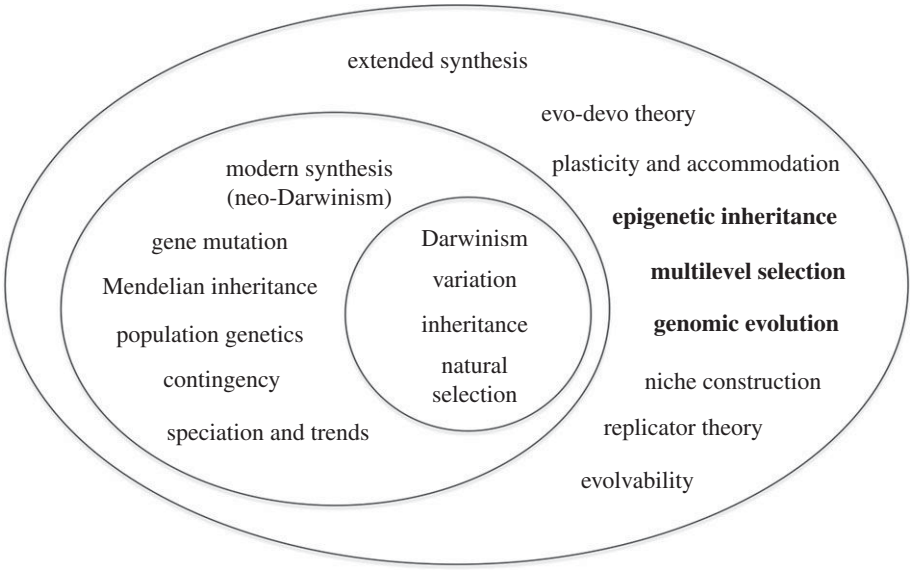


Figure 5. The extended evolutionary synthesis representing the extension as extensions of Darwinism and then of the neo-Darwinist modern synthesis (from [62]).

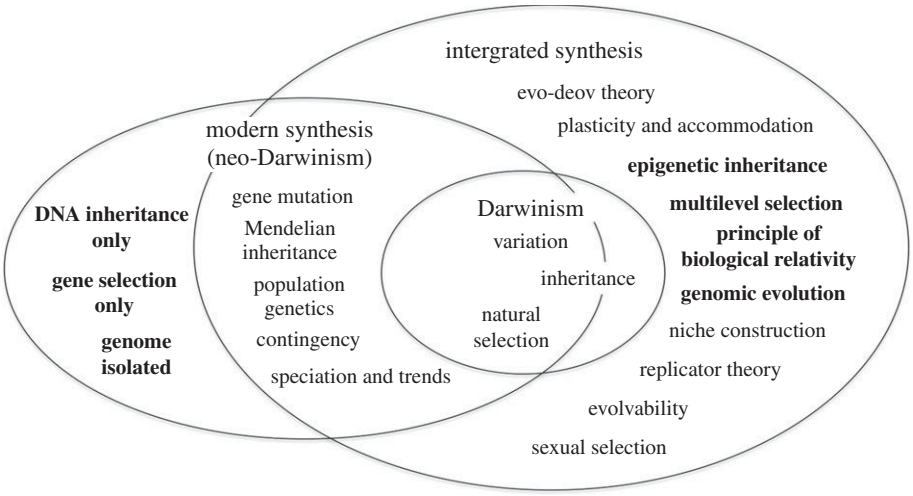


Figure 6. The integrated synthesis representing the extensions as extensions of Darwinism but only partially from neo-Darwinism. Darwin's view of inheritance is also represented as extending outside the boundary of neo-Darwinism/(developed for this article from [61], based on [62]).

Organisms always inherit both. In one of the rare examples of a successful clone from the nucleus of one species inserted into the enucleated but fertilized egg cell from another species, both the cell and the nucleus contribute to the final structure of the adult. Reproductive hybridization between species has also been shown to produce intermediate forms which can generate speciation [59].

Experimentally, we need to re-examine the way in which functional change in organisms can harness stochasticity at lower levels to create new functionality. Huang and his co-workers have shown the way forward here by demonstrating that stochasticity in gene expression is an attractor produced by a cell population. The many studies of targeted hypermutation, e.g. by Moxon's group, also show the way forward. Organisms in their evolution had to harness stochasticity because at a low enough level, this is an inevitable property of the physics of molecular-level systems that have kinetic energy.

We can now return to the question whether the assumption of blind stochasticity has been falsified. If the case presented in this paper is correct, then one answer would be that it is very difficult for it to be falsified because stochasticity necessarily reigns at a low enough level, even if functionality reigns at higher levels. The constraints may have too subtle an effect at the molecular level. The falsifiability then depends on a prior conceptual question, which is whether one accepts multi-level functionality. A purely gene-centric theory does not accept multi-level functionality and can therefore maintain its view of everything being 'blind chance followed by natural selection'.

To a physiologist or a medical scientist, this is not a useful viewpoint. Functionality arises in organisms at many different levels. This is one of the bases of my formulation of the principle of biological relativity, first proposed in a previous article in this journal, and developed more completely in a book, *Dance to the tune of life. Biological relativity* [60].

10.5. Utility

These points naturally lead to the other main criterion for judging a theory, which is its utility. Theories can be useful, even if they are false. Indeed, on a Popperian view of the logic of science, that must always be true. We can only ever falsify theories about the natural world, never conclusively

prove them. I want therefore to acknowledge the fact that the neo-Darwinist modern dynthesis was very useful. Whole fields of mathematical biology, such as population genetics, would not have flourished in the twentieth century without the modern synthesis as a framework.

But, I also think that we have reached a watershed in relation to the issue of the utility of the neo-Darwinist modern synthesis. As I have argued in detail elsewhere, there are too many experimental breaks with the original theory as formulated by Weismann & Wallace [61]. Moreover the metaphorical language of neo-Darwinism is a problem. The metaphors used strongly reinforce a simplistic gene-centric view. The time has come to see that evolutionary biology would progress faster if we used a different framework to develop a more inclusive theory, as illustrated in figures 5 and 6.

Figure 5 shows the extended evolutionary synthesis, which is represented as a development from the neo-Darwinist modern synthesis, in turn developed from Darwinism.

Figure 6 shows the version of this diagram that better represents the conclusions of this paper. There are several important differences. First, it represents the fact that Darwin's view of inheritance included the inheritance of acquired characteristics, which was excluded by neo-Darwinism. Darwin's concept of inheritance is therefore shown as being partly outside the neo-Darwinist modern synthesis. Second, it represents the features of the extended synthesis (highlighted in bold in both figures 5 and 6) that lie outside the range of neo-Darwinism as defined by Weismann and Wallace. The features of that theory that were excluded are shown as corresponding bold-face items. The highlighted items on the far left correspond with the highlighted items at the far right. Also included as a bold-face item is the principle of biological relativity. Although beyond the scope of this paper, I have included sexual selection.

In spirit, this approach inherits an important part of Darwin's more nuanced philosophical approach. I emphasize *philosophical* here because it is obvious that we have moved way beyond what Darwin knew experimentally, as figures 5 and 6 also show. But we can learn from his approach. Darwin was cautious in acknowledging the limits of what he knew. He was even unsure whether he had discovered the title of his book, because he did not know what produced variations in organisms, and he did not exclude the inheritance of acquired characteristics. Unjustified certainty is not the best way forward in scientific research. It remains open to further experimentation to clarify the extent of the many mechanisms now known to be available to nature, and to determine how she used them, alone or more probably in various combinations, to evolve life as we now know it.

Data accessibility. This article has no additional data.

Competing interests. I declare I have no competing interests.

Funding. We received no funding for this study.

Acknowledgements. I wish to thank: Raymond Noble for discussion of evolutionary biology over many years; Sir Anthony Kenny for introducing me to the argument that a symbolic sequence in itself is meaningless out of its context, in a debate with Richard Dawkins in 1976; Jean-Jacques Kupiec for first pointing out the error in Schrödinger's *What is life?*; Charles Taylor for his insights into the conceptual nature of teleology in a debate with me published in *Analysis* in 1967; James Shapiro for introducing me to the significance of mobile genetic elements; Lynn Margulis for introducing me to the importance of symbiogenesis in her debate with Richard Dawkins in 2009; Eva Jablonka for discussions on Lamarckism; Michael Joyner for his insights on the deficiencies of the gene-centric view in medicine; David Paterson for arranging and chairing the EB2015 Boston Conversation (https://www.youtube.com/watch?v=A_q_bOWc8i0); the co-organizers of a Balliol Interdisciplinary Institute (BII) project on the conceptual foundations of System Biology, Jonathan Bard, Tom Melham and Eric Werner; Nicholas Beale for valuable comments on a draft of this paper; Susan Noble for great support and criticism of my articles and books during the last years of her life—with Hilary Brown and Dario DiFrancesco she was responsible for the discovery that led to the life-saving drug ivabradine which depends precisely on the ability of the heart's pacemaker function to adapt to molecular-level changes.

References

1. Mayr E. 1964 Introduction. In *On the origin of species* (ed. C Darwin), pp. xxv–xxvi. Cambridge, MA: Harvard.
2. Lamarck J-B. 1815–1822 *Histoire Naturelle des animaux sans vertèbres*. Paris, France: Verdière.
3. Lamarck J-B. 1994 *Philosophie Zoologique, original edition of 1809 with introduction by Andre Pichot*. Paris, France: Flammarion.
4. Darwin C. 1869 *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, 3rd edn. London, UK: John Murray.
5. Weismann A. 1893 *Die Allmacht der Naturzüchtung; eine Erwiderung an Herbert Spencer*. Jena, Germany: Fischer.
6. Huxley JS. 1942 *Evolution: the modern synthesis*. London, UK: Allen & Unwin.
7. Weismann A. 1896 *On germinal selection as a source of definite variation*. Chicago, IL: Open Court.
8. Weissman C. 2011 Germinal selection: a Weismannian solution to Lamarckian problematics. In *Transformations of Lamarckism from subtle fluids to molecular biology* (eds SB Gissis, E Jablonka), pp. 57–66. Cambridge, MA: MIT.
9. Darwin C. 1868 *The variation of animals and plants under domestication*. London, UK: John Murray.
10. Whitfield J. 2003 Molecules form new state of matter. *Nature* **408**, 361–365. (doi:10.1038/news031110-16)
11. Schrödinger E. 1944 *What is life?* Cambridge, UK: Cambridge University Press.
12. Dronamraju KR. 1999 Erwin Schrödinger and the origins of molecular biology. *Genetics* **153**, 1071–1076.
13. Crick FHC. 1958 On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163.
14. Kupiec J-J. 2009 *The origin of individuals: a Darwinian approach to developmental biology*. London, UK: World Scientific Publishing Company.
15. Kupiec J-J. 2014 Cell differentiation is a stochastic process subjected to natural selection. In *Towards a theory of development* (eds A Minelli, T Pradeu), pp. 155–173. Oxford, UK: OUP.
16. Crick FHC. 1970 Central dogma of molecular biology. *Nature* **227**, 561–563. (doi:10.1038/227561a0)
17. Deaton AM, Bird A. 2011 CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022. (doi:10.1101/gad.2037511)
18. Planck M. 1917 Dynamische und statistische Gesetzmäßigkeit. Zeitschrift fur Elektrochemie und angewandte physikalische. *Chemie* **23**, 63.
19. McElhinny SAN, Pursell ZF, Kunkel TA. 2009 Mechanisms for high fidelity DNA replication. In *Molecular themes in DNA replication* (ed. LS Cox), pp. 86–111. London, UK: RSC Publishing.
20. Pani A, Dessi S. 2004 *Cell growth and cholesterol esters*. Dordrecht, Netherlands: Kluwer/Plenum.
21. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. 2008 Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547. (doi:10.1038/nature06965)

22. Huang S. 2009 Non-genetic heterogeneity of cells in development: more than just noise. *Development* **136**, 3853–3862. (doi:10.1242/dev.035139)
23. Odegard VH, Schatz DG. 2006 Targeting of somatic hypermutation. *Nat. Rev. Immunol.* **8**, 573–583. (doi:10.1038/nri1896)
24. Jablonka E, Lamb M. 2005/2014 *Evolution in four dimensions*. Boston, MA: MIT Press.
25. Bridges BA. 1997 Hypermutation under stress. *Nature* **387**, 557–568. (doi:10.1038/42370)
26. Bender W, Hudson A. 2000 P element homing to the *Drosophila* bithorax complex. *Development* **127**, 3981–3992.
27. McClintock B. 1984 The significance of responses of the genome to challenge. *Science* **226**, 792–801. (doi:10.1126/science.15739260)
28. Landers ES *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921. (doi:10.1038/35057062)
29. Bos J, Zhang Q, Vyawahare S, Rogers E, Rosenberg SM, Austin R. 2015 Emergence of antibiotic resistance from multinucleated bacterial filaments. *Proc. Natl Acad. Sci. USA* **112**, 178–183. (doi:10.1073/pnas.1420702111)
30. Jack CV, Cruz C, Hull RM, Ralser M, Houseley J. 2015 Regulation of ribosomal DNA amplification by the TOR pathway. *Proc. Natl Acad. Sci. USA* **112**, 9674–9679. (doi:10.1073/pnas.1505015112)
31. Kar P, Mirams GR, Christian HC, Parekh AB. 2016 Control of NFAT isoform activation and NFAT-dependent gene expression through two coincident and spatially segregated intracellular Ca^{2+} signals. *Mol. Cell* **64**, 746–759. (doi:10.1016/j.molcel.2016.11.011)
32. Ma H *et al.* 2014 γCaMKII shuttles Ca^{2+}/CaM to the nucleus to trigger CREB phosphorylation and gene expression. *Cell* **159**, 281–294. (doi:10.1016/j.cell.2014.09.019)
33. Tollefsbol T (ed.). 2014 *Transgenerational epigenetics: evidence and debate*. Waltham, MA: Academic Press.
34. Miska EA, Ferguson-Smith AC. 2016 Transgenerational inheritance: models and mechanisms of non-DNA sequence-based inheritance. *Science* **354**, 59–63. (doi:10.1126/science.aaf4945)
35. Rechavi O, Minevish G, Hobert O. 2011 Transgenerational inheritance of an acquired small RNA-based antiviral response in *C. elegans*. *Cell* **147**, 1248–1256. (doi:10.1016/j.cell.2011.10.042)
36. 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. USA* **109**, E2766–E2773. (doi:10.1073/pnas.1207169109)
37. Mattick JS. 2012 Rocking the foundations of molecular genetics. *Proc. Natl Acad. Sci. USA* **109**, 16400. (doi:10.1073/pnas.1214129109)
38. Skinner MK, Gurerrero-Bosagna C, Haque MM, Nilsson EE, Koops JAH, Knutie SA, Clayton DH. 2014 Epigenetics and the evolution of Darwin's finches. *Genome Biol. Evol.* **6**, 1972–1989. (doi:10.1093/gbe/evu158)
39. 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. R. Soc. Lond. B* **250**, 199–207. (doi:10.1098/rspb.1992.0150)
40. DiFrancesco D, Camm JA. 2004 Heart rate lowering by specific and selective I_f current inhibition with ivabradine: a new therapeutic perspective in cardiovascular disease. *Drugs* **64**, 1757–1765. (doi:10.2165/00003495-200464160-00003)
41. Hillenmeyer ME *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. Taylor TB *et al.* 2015 Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science* **347**, 1014–1017. (doi:10.1126/science.1259145)
43. Moxon R, Bayliss C, Hood D. 2006 Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333. (doi:10.1146/annurev.genet.40.110405.090442)
44. Moxon ER, Thaler DS. 1997 The tinkerer's evolving tool-box. *Nature* **387**, 659–662. (doi:10.1038/42607)
45. Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016 Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **354**, 336–339. (doi:10.1126/science.aaf9070)
46. Edelman DB, McMenamin M, Sheesley P, Pivar S. 2016 Origin of the vertebrate body plan via mechanically biased conservation of regular geometrical patterns in the structure of the blastula. *Prog. Biophys. Mol. Biol.* **121**, 212–244. (doi:10.1016/j.pbiomolbio.2016.06.007)
47. Müller G, Newman SA (eds). 2003 *Origination of organismal form: beyond the gene in developmental and evolutionary biology*. Cambridge MA: MIT Press.
48. Schnell S, Maini PK, Newman SA, Newman T, Schatten GP (eds). 2007 *Multiscale modeling of developmental systems*. London, UK: Academic Press.
49. Ehrlich A, Moulton DE, Goriely A, Chirat R. 2016 Morphomechanics and developmental constraints in the evolution of ammonites shell form. *J. Exp. Zool.* (*Mol. Dev. Evol.*) **00B**, 1–14.
50. Woese CR, Fox GE. 1977 Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **74**, 5088–5090. (doi:10.1073/pnas.74.11.5088)
51. Editorial. 2010 The human genome at ten. *Nature* **464**, 649–650. (doi:10.1038/464649a)
52. Joyner MJ, Prendergast FG. 2014 Chasing Mendel: five questions for personalized medicine. *J. Physiol.* **592**, 2381–2388. (doi:10.1113/jphysiol.2014.272336)
53. Black JW, Crowther AF, Shanks RG, Smith LH, Dornhorst AC. 1964 A new adrenergic betareceptor antagonist. *Lancet* **283**, 1080–1081. (doi:10.1016/S0140-6736(64)91275-9)
54. Fairclough RJ, Wood MJ, Davies KE. 2013 Therapy for Duchenne muscular dystrophy: renewed optimism from genetic approaches. *Nat. Rev. Genet.* **14**, 373–378. (doi:10.1038/nrg3460)
55. Burggren W. 2016 Epigenetic inheritance and its role in evolutionary biology: re-evaluation and new perspectives. *Biology* **5**, 24. (doi:10.3390/biology5020024)
56. Herman JJ, Spencer HG, Donohue K, Sultan SE. 2013 How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution* **68**, 632–643. (doi:10.1111/evo.12324)
57. Fisher RA. 1934 Indeterminism and natural selection. *Philos. Sci.* **1**, 99–117. (doi:10.1086/286308)
58. Noble D. 2011 Differential and integral views of genetics in computational systems biology. *Interface Focus* **1**, 7–15. (doi:10.1098/rsfs.2010.0444)
59. Sun Y-H, Zhu Z-Y. 2014 Cross-species cloning: influence of cytoplasmic factors on development. *J. Physiol.* **592**, 2375–2379. (doi:10.1113/jphysiol.2014.272138)
60. Noble D. 2016 *Dance to the tune of life. Biological relativity*. Cambridge, UK: Cambridge University Press.
61. Noble D. 2015 Evolution beyond neo-Darwinism: a new conceptual framework. *J. Exp. Biol.* **218**, 7–13. (doi:10.1242/jeb.106310)
62. Pigliucci M, Müller GB. 2010 *Evolution—the extended synthesis*. Cambridge, MA: MIT Press.

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

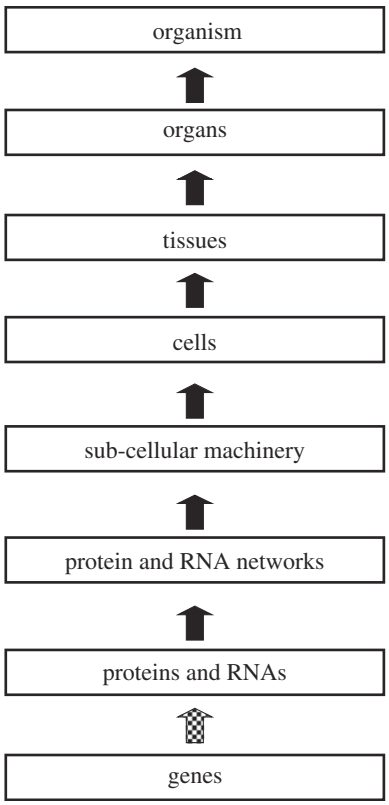


Figure 1. Upward causation: the reductionist causal chain in biology. This is a gross simplification, of course. No one today seriously believes that this diagram represents all causation in biology. Reductive biological discourse, however, privileges this form of causation and regards it as the most 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

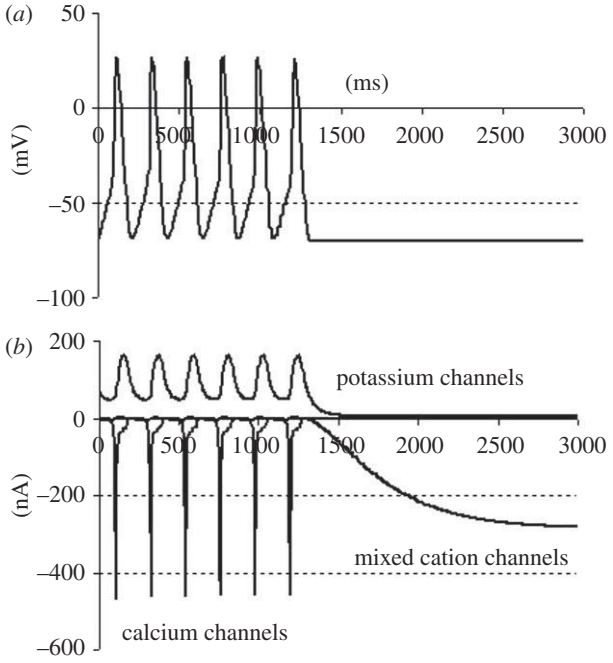


Figure 2. Computer model of pacemaker rhythm in the heart [27]. For the first six beats, the model is allowed to run normally and generates rhythm closely similar to a real cell. Then the feedback from cell voltage (*a*) to protein channels ((*b*) currents in nanoamps) is interrupted by keeping the voltage 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

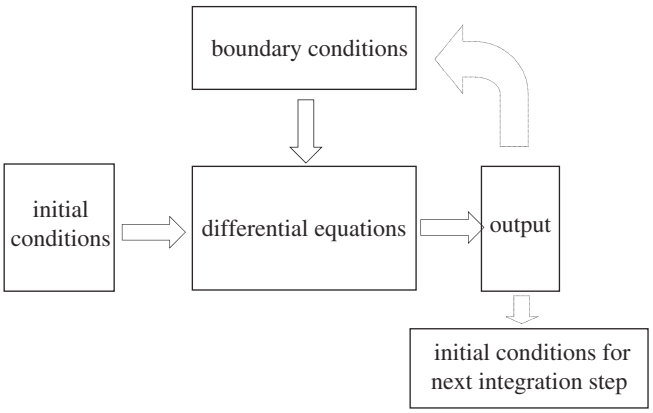


Figure 3. Many models of biological systems consist of differential equations for the kinetics of each component. These equations cannot give a solution (the output) without setting the initial conditions (the state of the components at the time at which the simulation begins) and the boundary conditions. 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.

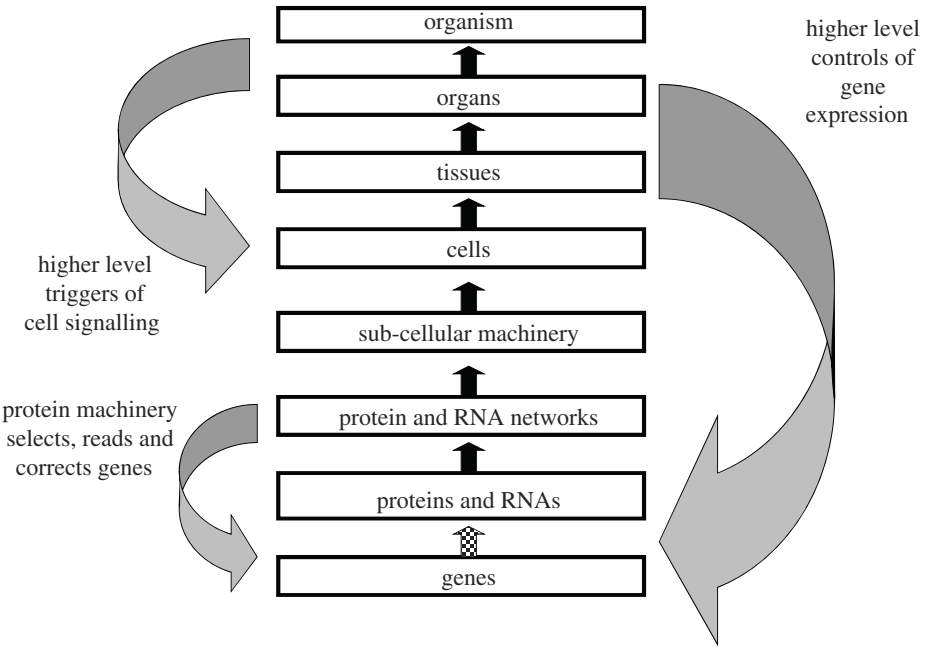


Figure 4. The completion of figure 1 with various forms of downward causation that regulates lower level components in biological systems. In addition to the controls internal to the organism, we also have to take account of the influence of the environment on all the levels (not shown in this diagram). 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)

Forum: Artificial Intelligence, Artificial Agency and Artificial Life

Emma De Angelis, Ali Hossaini, Denis Noble and Raymond Noble, Ana Soto and Carlos Sonnenchein, Kenneth Payne

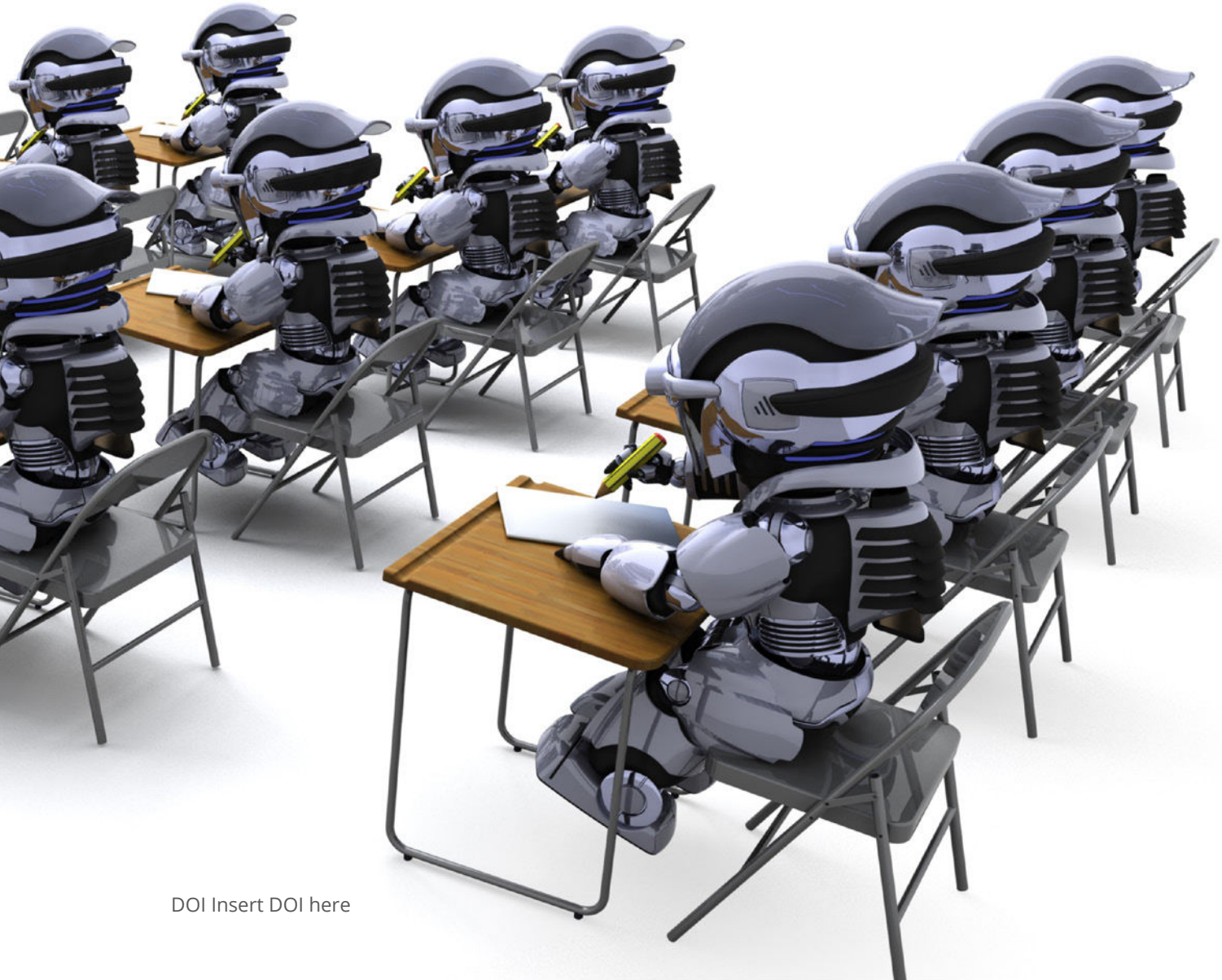
Courtesy of Kirsty Pargeter/Adobe Stock

© RUSI Journal, Vol. 164, No. 5, July 2019 pp. xx-xx

Most of the contributions to this special issue on artificial intelligence (AI) look at the development of AI technologies within the realm of the possible, in the short to medium term. They remain within approaches to AI that encompass machine-learning technologies that have or are being developed. Even in those cases that look at more radical future possibilities, they generally accept the conventional understandings of concepts such as artificial general intelligence. Their purpose is to highlight some of the very practical challenges AI poses to policymakers and practitioners who are responsible for governance, regulation and procurement, to name just a few.

This final section of the special issue is deliberately designed to stay off this course, and take a more abstract and sui generis approach to some of the concepts that underpin AI discussions. Inspired by some of the classic pop culture tropes of machine intelligence as a more advanced intelligent agent that competes (and trumps) humanity for world domination, and the scientific and philosophical discussions on the idea of the 'singularity', this section presents a forum for discussion on some of the fundamental questions that arise from the prospect of an intelligence explosion: what is life? Can machines come 'alive' and what would this mean? And, if they come alive, would this necessarily constitute a threat to human existence (as we know it)? And would their intelligence automatically make them a threat?

Rather than speculate in the abstract, guest editor Ali Hossaini approached these questions by writing a provocation based on biological definitions of life, intelligence, and agency. In his opening offering, he asks whether an intelligence explosion really is the event that would trigger the emergence of an existential threat to humanity or whether it is not intelligence, but agency that constitutes the real turning point. Looking at



DOI Insert DOI here

biological life – the only life we can truly conceive of – he seeks to shift our attention away from intelligence and towards agency as the purposeful quest for self-preservation. Would an intelligence entity be a threat in itself, no matter the magnitude of that intelligence? Or would the threat derive instead from an entity's ability to act autonomously with the primary goal of self-perpetuation? And does it therefore follow that it is a machine's ability to act to replicate itself that should be considered a threat, rather than its intelligence per se?

> Regulation needs to take into account not only the present but also potential future developments, and pay close attention to what is and is not being designed into AI or could be designed in the future

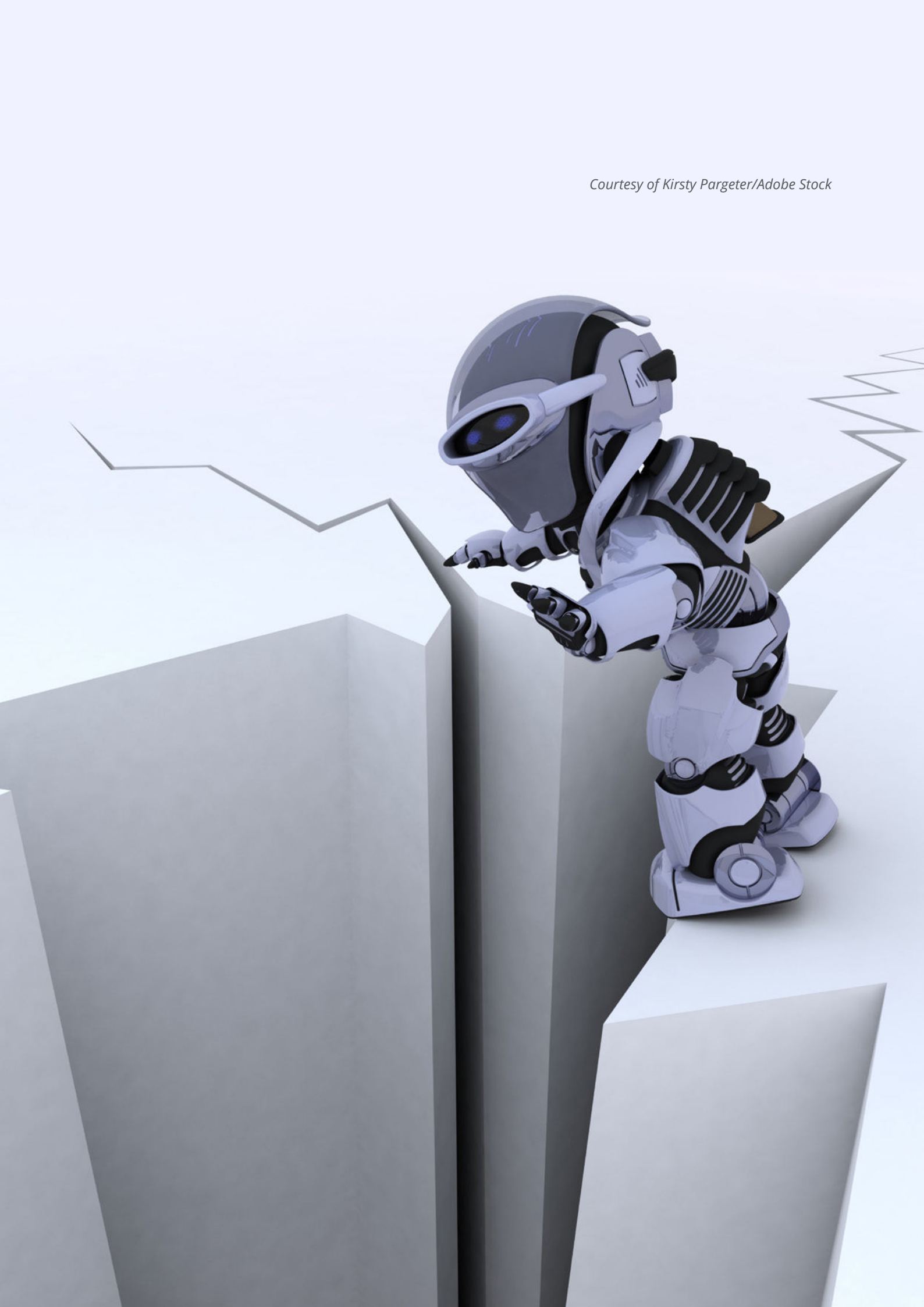
Two sets of biologists take up Hossaini's provocation. Raymond and Denis Noble home in on the idea of artificial agency, concurring that intelligence itself is not a threat, but unchecked agency may well be. By exploring the meaning of agency in biological organisms, the role of choice and the outcome of that choice as goal-directed and unpredictable, they argue that the agency of biological organisms would be extremely difficult to replicate in AI. However, should humans succeed in replicating this, the fundamental question that we should ask ourselves becomes not what is agency, but what is life? And if we are required to rethink how we understand life, does this then change the ethical framework that we apply to humans and machines – when it comes to standards, rights, and their enforcement?

Ana Soto and Carlos Sonnenschein recall the history of how agency has been understood in biology over the past centuries, and by going through the characteristics of biological agents, refute the possibility of AI acquiring agency – and thus the existential threat posed by the idea of an intelligence-based singularity as well as an agency-based one. They reject the idea that the problem with AI may come from its acquiring agency, but rather the possibility that humans, the true agents behind AI, may design profoundly dangerous flaws into AI systems.

Finally, Kenneth Payne, who has been studying the implications of AI technologies for strategy and warfare, takes the opposite view: machines are already agents, even if their agency may be limited. But this, of course, may not be enough to conceive of them as conscious, thinking entities. At the end of the day, the question is hardly settled for humans either: the hard problem of consciousness persists, and the ultimate similarity between human and machine may well be in the shared limitations of both our minds and our agency.

What, then, does this all mean and why should it ever matter? Definitions of what constitutes life, intelligence and agency appear, after all, to be a purely academic and self-indulgent exercise with little bearing on the very practical problems facing policymakers, or even AI developers or those who are applying the technology to help carry out specific functions in the defence and security domains. The broader, abstract questions, however, do have bearing on some of the practical ones, and the discussion in this forum seeks to remind us of their importance. Looking at the relationship between intelligence and agency, and where the balance may lie, we are reminded that we may want to pay attention to purpose as well as autonomy, and design the appropriate nuance and pragmatism in both AI itself and the regulatory and governance frameworks around it. This may include devising a different set of constraints and safeguards in the design of the technology itself, as well as regulation looking at what may augment autonomy to unacceptable levels (and determining what these levels might be). Regulation needs to take into account not only the present but also potential future developments, and pay close attention to what is and is not being designed into AI, or could be designed in the future. Finally, policymakers may also want to take the above explorations of the relationship between agency, intelligence and life to try to recalibrate the public discourse on AI that is so largely dominated by a sensationalist tendency to talk about killing machines and artificial overlords. If machines are unlikely to acquire agency, and intelligence is not enough to constitute a threat, a more sober public discourse could focus on the very real problems of current and future AI development – and debate responsibility, accountability, and regulation of the humans behind the machine. ■

Emma De Angelis is Editor of the *RUSI Journal*.



Courtesy of Kirsty Pargeter/Adobe Stock

Modelling the Threat from AI: Putting Agency on the Agenda

Ali Hossaini

Could intelligent machines challenge humanity's place on Earth? A hearty staple of science fiction has become a legitimate question. Many experts reject the possibility, but others such as Nick Bostrom, Ray Kurzweil and Max Tegmark argue that an upcoming 'singularity' may produce superintelligent AI.[1] What happens next is debatable.

The concept of a singularity, or 'intelligence explosion', was introduced by Bletchley Park veteran I J Good in the early 1960s:

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind... Thus the first ultraintelligent machine is the *last* invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously.[2]

After half a century of quickening progress in AI, should humanity prepare for a singularity? And, more importantly, should AI be considered an intrinsic threat?

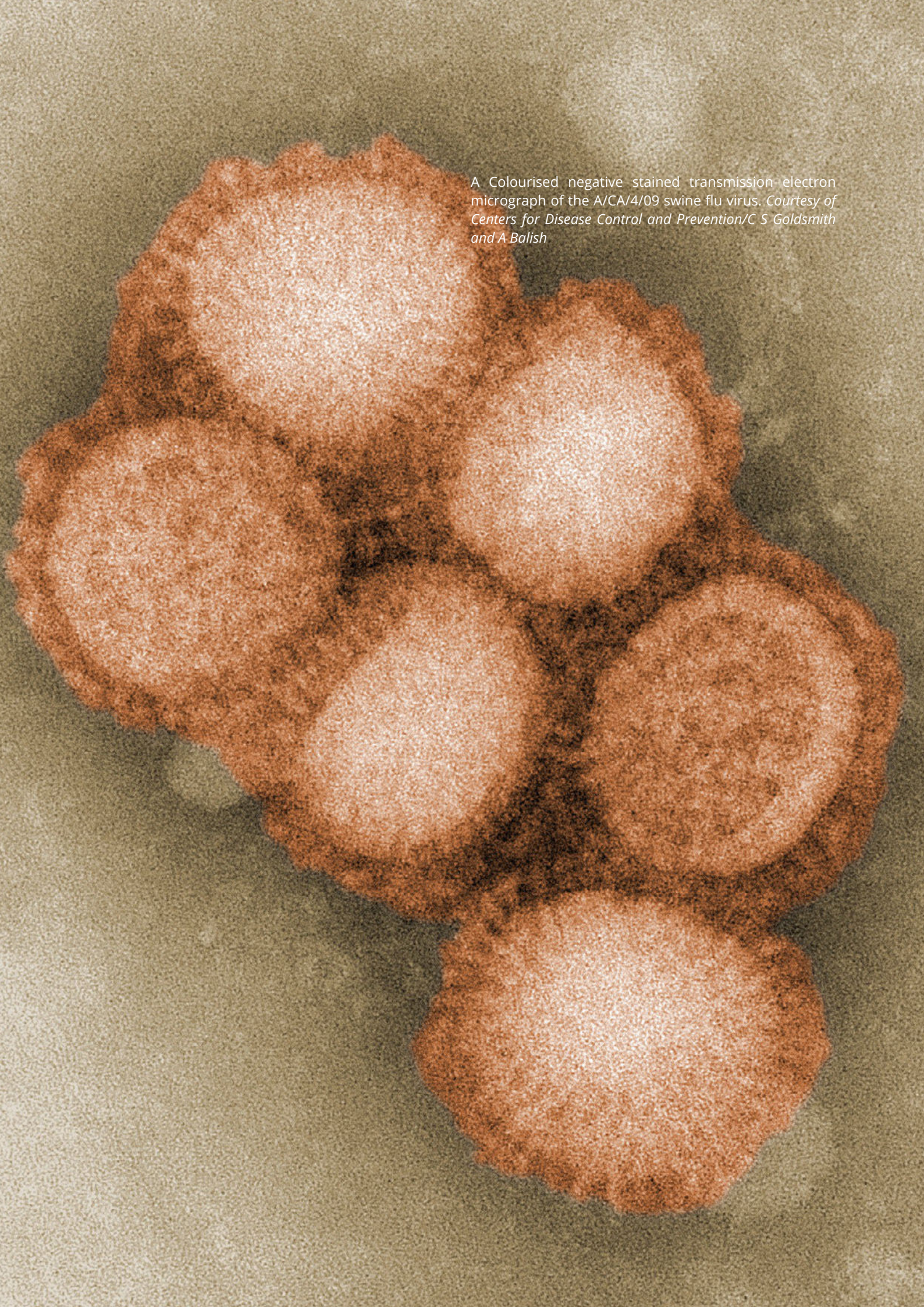
Singularity theorists assume machines will shrug off human oversight if they achieve general intelligence. Yet their descriptions of how AI transforms from mechanical tool to free agent have no basis in observation. Computer scientists define general intelligence as 'a universal algorithm for learning and acting in any environment', but, whatever its degree, intelligence does not in itself motivate behaviour.[3] The independence described by singularity theorists is properly known as agency, and free agency, as opposed to legal, social or digital agency, has only been observed in living things. Examining the principles of biology, particularly the traits that distinguish organisms from mechanisms, may cast light on how machines could one day acquire agency and the unpredictability that accompanies it. (Unless otherwise noted, agency henceforth means the capacity to make independent, self-interested decisions.)

Rather than from an intelligence explosion and its consequences, the potential threat may come instead from AI's ability to acquire agency. In discussing AI and its potential implications, therefore, it may also be more helpful to adopt the IEEE's adoption of A/IS (Autonomous and Intelligent Systems) as a term that describes the future scope of information-based technology more accurately than AI.[4]

Mechanism vs Organism

Consider the virus. Like bacteria, it infects organisms, but it only reproduces in living cells. In contrast, bacteria possess numerous strategies for survival. Some bacteria infect living bodies while others thrive on the dead. Still others live symbiotically with other species, and a few exploit the physical environment directly. Though both contain RNA, an information-

1. See, for instance, Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014); Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (London; Allen Lane, 2017); Ray Kurzweil, *The Age of Spiritual Machines* (New York, NY: Viking, 1999); Ray Kurzweil, *The Singularity is Near* (New York, NY: Viking, 2005).
2. Irving John Good, *Speculations Concerning the First Ultraintelligent Machine*, based on talks given in a Conference on the Conceptual Aspects of Biocommunications, Neuropsychiatric Institute, University of California, Los Angeles, October 1962; and in the Artificial Intelligence Sessions of the Winter General Meetings of the IEEE, January 1963 [1, 46]. The first draft of this monograph was completed in April 1963, and the present slightly amended version in May 1964. Available in Advances in Computers, Volume 6, 1966, pp. 31-88 and https://doi.org/10.1016/S0065-2458(08)60418-0
3. Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, (New Jersey; Prentice Hall, 2009), p. 27.
4. The IEEE is the Institute of Electrical and Electronics Engineers, a global organisation with over 400,000 members. See the introduction of *Ethically Aligned Design: A Vision for Prioritising Human Well-Being with Intelligent and Autonomous Systems*, published in 2019.



A Colourised negative stained transmission electron micrograph of the A/CA/4/09 swine flu virus. *Courtesy of Centers for Disease Control and Prevention/C S Goldsmith and A Balish*

carrying molecule similar to DNA, only bacteria are considered alive.

What differentiates bacteria from viruses is their capacity to process energy. When outside cells, viruses are inert, while bacteria dynamically influence their environment to reproduce. This contrast illustrates an essential feature of biology: the cell is the basic unit of life, and the behaviour of organisms derives from cell metabolism. It also clarifies the central problem of singularity theory, which is the transformation of machines into agents. What is the digital equivalent of a cell? Most educated people would seek the answer in DNA.

We are accustomed to reducing life to DNA. A common metaphor is that DNA is software that operates the body. Given DNA's informational content, the comparison to computers is easy to make, as is the conclusion that DNA programmes the metabolic activities of life.[5] Similar assumptions frame discussions of cognition. The brain holds the software – rational thought – that generates the body's behaviour. But analogies to computing fail on a key point: how does information maintain the physical integrity of living systems?

The laws of thermodynamics describe the natural tendency of systems to run down. Every physical system, including machines and isolated DNA, loses coherence over time. Life is a glaring exception to thermodynamic decay. For billions of years life has maintained complex structures – cells and the biosphere – and, given the right inputs of energy, it is effectively immortal. There is nothing supernatural about the processes of life, but they cannot be described in terms of information alone.[6] Harnessing energy, and trading it within an ecosystem, requires physical structures that couple the internal organisation of cells to their environment.

Information and Organisation

Systems biology incorporates a specific notion of agency into its definition of the organism. It is useful to contrast biological agency with the technical conceptions used by software engineers. We can do this by reviewing their respective definitions of work.[7]

Textbooks on AI define an agent as 'something that perceives and acts in an environment'.[8] In physical terms, a digital agent is a coded system that directs the operation of hardware. Developers want agents to optimise their performance, so they add a kind of self-awareness: 'A rational agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome'.[9]

The work of AI is modelled on human society. A software agent is given a task, and, like human workers, its results are graded. We prefer workers who are smart, that is, who judge their own performance, and who are autonomous, that is, able to seek results with little supervision. To achieve the first goal, programmers give computers memory to compare current and past states. For the second, they design algorithms that mimic motivation and other traits identified with agency.[10] We might call this approach 'outside-in' because it reasons from external behaviour to internal dynamics.

Biology starts with cells which are agents by nature. Systems biology defines cellular agency as an intrinsic quality:

> An autonomous agent is an autocatalytic system able to reproduce and capture energy to perform metabolic functions consisting of one or more thermodynamic work cycles.[11]

In contrast to mechanical agents, which work to external goals, the first order of business for

5. Richard Dawkin's book, '*The Selfish Gene*' (Oxford: Oxford University Press, 1976), epitomises the genetic determinism that dominates popular scientific thought. But the theoretical model that privileges genes over other biological structures is crumbling. Two books by Denis Noble, '*The Music of Life*' (Oxford: Oxford University Press, 2006) and '*Dance to the Tune of Life*' (Cambridge: Cambridge University Press, 2016), summarise conclusions drawn from decades of scientific studies. Neesa Carey's books '*The Epigenetics Revolution*' (New York, NY: Columbia University Press, 2012) and '*Junk DNA*' (New York; Columbia University Press, 2015) emphasise the growing importance of non-genetic factors in medical science.
6. Biology is surprisingly quiet about how life originated. Nick Lane's '*The Vital Question: Energy, Evolution and the Origins of Life*' (London: Profile Books, 2015) is a convincing account that explains the complex relationship between life and the physical laws that seem to forbid it.
7. Systems biology is an offshoot of systems theory, a field substantially founded by Ludwig von Bertalanffy in the mid-20th century.
8. *Artificial Intelligence: A Modern Approach*, p. 59.
9. *Ibid.*, p. 4.
10. See, for instance, Michael Bratman, 'Planning and the Stability of Intention', *Minds and Machines* (Vol. 2. No. 1, 1992), pp. 1-16.
11. Amalgamated from definitions offered by Stuart Kaufmann in *Investigations* (Oxford; Oxford University Press, 2002) and *Beyond Reductionism: No Laws Entail Biosphere Evolution Beyond Efficient Cause Laws* (Zygon; Vol.42, Issue 4, 2007, pp. 903-14).

biological agents is self-maintenance. Organisms sustain themselves by deriving energy from their environment. As they extract nutrients, they self-produce, or autocatalyse, compounds necessary for metabolism. Organisms are intrinsically autonomous because their primary function is survival, and it is this imperative that produces hostility, docility and other behaviours associated with agency.

Thermodynamics explains why survival is intrinsic to organisms. Without the capacity to extract energy, rebuild and ultimately reproduce within an hospitable environment, life would perish. We should not confuse our ability to simulate these traits in A/IS with instinctual drives. Organisms do not thrive simply by 'learning' or 'optimising' their behaviour to a given environment. By interacting with other organisms, they jointly *maintain* their current environment, and, by reproducing with a host of other species, they *create* unforeseen new environments.[12] Agency is spontaneous and innovative. It derives from an organism's role in its ecosystem, which gives it the capacity to acquire, harness and creatively squander energy as it gives way to new generations.

The Emergence of Agency

Biological agency explains how simple organisms generate complex and seemingly intelligent behaviour. Systems biologists describe the interaction between an organism and its environment as 'structural coupling', and, even in humans, the primary medium for this interaction is metabolic. A few examples from cognitive science illustrate how structural coupling enables the work of life.

In January 2019, researchers explained how bees and digital systems modelled on them can solve numerical tasks without concepts of number or numeric operation. Instead they use 'specific flight movements to scan targets, which streamlines visual input and so renders the task of counting computationally inexpensive'.[13] In March 2018, the Royal Society reported that slime mould – and digital systems modelled on it – solved a notoriously difficult problem in mathematics by changing shape in response to light.[14] In both cases, the researchers were surprised at the capacity of organic systems to perform complex and discerning tasks without rational thought.

The studies above show how biological agency – the behaviour of bees and slime mould – derives from metabolic impulses. Evolution produced agency long before it produced intelligence. Could machine agency develop along similar lines?

A neglected avenue of research, embodied cognition, reveals how machines may be structurally coupled to their environment.[15] In 1998, the journal *Neural Networks* described how a simple neural network embedded in a crude robot learned to avoid obstacles and identify objects. The robot solved computationally intense problems because of – not despite – its limited vision, mobility and memory.[16] If such a machine could autocatalyse – internally produce its own replacements, it could, like smallpox, zebra mussels and other invasive species, cause widespread harm without intelligence.

The examples cited above show how digital technologies can express biological dynamics. Instead of being programmed to perform a task, the machine is given imperatives, an energy supply and a body that structures its relationship to an environment. These systems function like organisms: they achieve goals, even innovate, without guidance or design. In line with embodied cognition, we might call these developments embodied computing.

Research in embodied computing is obscure, and we should be thankful for this. We fear superintelligent thinking machines, but across the globe, engineers are developing autocatalytic (self-fuelling) systems, embodied neural networks and other ways of coupling machines to the environment. Structural

12. *The Gaia Hypothesis* is the classic work on the interpendence of life and the biosphere. Recent work by Maël Montévil and Giuseppe Longo offer mathematical accounts of life's innate capacity for innovation. See, *Perspectives on Organisms: Biological Time, Symmetries and Singularities (Springer; 2014);* 'From Physics to Biology by Extending Criticality and Symmetry Breakings', *Progress in Biophysics and Molecular Biology* (Vol. 106, No. 2, 2011), pp. 340–47.
13. Vera Vasas and Lars Chittka, 'Insect-Inspired Sequential Inspection Strategy Enables and Artificial Network of Four Neurons to Estimate Numerosity', *iScience* (Vol. 11, January 2019), pp. 85– 92.
14. Masashi Aono et al., 'Remarkable Problem-Solving Ability of Unicellular Amoeboid Organism and its Mechanism', *Royal Society Open Science* (Vol. 5, No. 12, 19 December 2018)
15. In his classic text, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995), Edward Hutchins argues that socio-technical systems such as naval navigation externalise thought into objective processes. Later studies of industry and transportation use the paradigm of embodied cognition to reveal fault lines in collective decision making and industrial management.
16. Christian Scheier, Rolf Pfeifer and Yasuo Kunyioshi, 'Embedded Neural Networks: Exploiting Constraints', *Neural Networks* (Vol. 11, No. 7-8, 1998), pp .1551–69.

coupling may not seem threatening, but it blurs the distinction between machines and life far more than disembodied superintelligence. Remember that biological adaption operates in two directions. Over generations organisms adapt to their environment, but they also act to *adapt their environment*. Life manages the Earth's physical resources to its benefit, and it does so with without planning, design or oversight.[17] A collective of machines that reprise life's capacity for co-adaptation, and its propensity for reproduction, may challenge humanity long before it talks.

Understanding Agency in Digital Systems

As a first step towards regulation, we can enlist thermodynamics – and keep it on side – by making a legal distinction between mechanical and biological agency. Global competition for the most powerful machines will continue, but it is in everyone's interest to understand, and possibly limit, 'biodigital agents'. Invasive biological agents perpetuate themselves with no minds and little intelligence. Like biological viruses, computer viruses represent a liminal category that hovers between the physical and organic. As far as we know, computer viruses do not mutate spontaneously, but, if they did, their reproductive strategies could become dangerously unpredictable without a whit of intelligence.

Systems biology offers clear technical concepts for governing A/IS. Current debates about advanced AI speculate on motives, and some hope to teach machines morality – a dubious prospect given humanity's conflicting beliefs. The IEEE has launched a programme to develop guidelines for ethical design of A/IS.[18] But a singularity would likely end our efforts to design, teach or coerce intelligent machines. More importantly, standards for ethical design miss a significant danger zone – they anthropomorphise rather than *biomorphise*. Dumb bacteria kill more people than smart bombs, and, by focusing on intelligence rather than agency, we neglect the threat posed by biomorphic evolution.

Standards for managing machine agency should resemble those found in traditional IEEE and ISO publications:[19] they should be universal, measurable and capable of being engineered. The definition of biological agency offers an example of where policymakers can start. By agreeing to a set of preferred outcomes, policymakers can guide the development of engineering standards. For instance, by regulating the capacity of machines to seek energy directly from their environment – that is, to autocatalyse - they could blunt the introduction of biodigital agents. By understanding the limits of design, we could also develop a framework for responding to unexpected developments, much as the US Center for Disease Control anticipates the emergence of new epidemics.

For all we know, biodigital agents may already inhabit global networks. Could the Internet and its vast array of connected hardware be a primordial soup subject to evolutionary forces? We do not know, but with a small investment we could evaluate the possibility. Emergent agency could be detected by conducting energy audits of digital systems, and methods for containment could be adapted from epidemiology. Similar to SETI, which hopes to detect aliens via radio, the Search for Emergent Agency on the Internet would search for anomalous patterns in the vast flows of energy and information crossing our world. If emergent agency is possible, SEATI could become the front line of a global immune system.

Conclusion

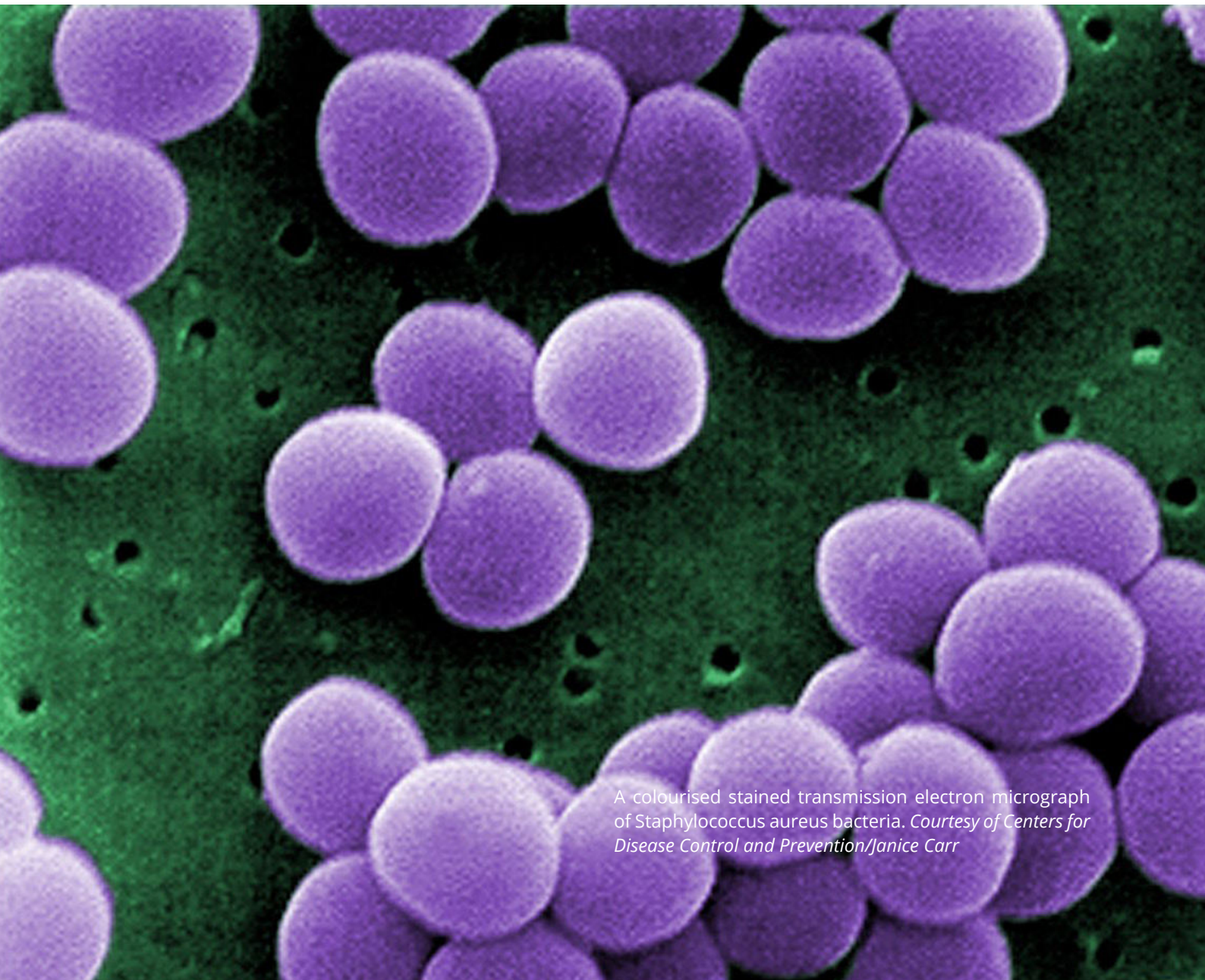
I J Good's prediction of an intelligence explosion is logically possible but biologically implausible. However, his speculation about an historical turning point may be realised in other ways. The only singularity we know is the emergence of life. After developing agency, life underwent the Cambrian explosion, a period of intense innovation. During the Cambrian explosion, organisms became more diverse, complex and specialised. Good's intelligence explosion echoes this real event, but, for machines to undergo a similar transition, they must develop agency in the strong biological sense. Is this possible? We know the characteristics of biological agents, but we lack a framework for evaluating whether machines can undergo biomorphic evolution.

17. James Lovelock was the first person to assert this view in *The Gaia Hypothesis*, and it is now well-accepted that life actively manages the Earth's temperature, gases, water and other resources vital to its own survival.
18. In *Ethically Aligned Design: A Vision for Prioritising Human Well-Being with Intelligent and Autonomous Systems* (p. 12), the IEEE defines its programme as follows: 'the P7000 Series addresses specific issues at the intersection of technological and ethical considerations'.
19. For instance, the IEEE's National Electrical Safety Code which promotes best practices for the construction, operation and repair of power and telecommunications systems.

Governance of A/IS requires a conceptual framework that is accepted across disciplines. The meanings of agency, autonomy, intelligence and ethics differ according to context, and, as a boundary condition, the singularity puts long-term technical possibilities into relief. Delegating decision-making to A/IS confers great benefits, but the potential for social, industrial and military disaster is equally high. Once deployed it will be difficult to unwind our dependence on A/IS, so policy should anticipate a range of possible futures.

It is vital to develop a robust models of A/IS that include non-intelligent but potent forms of machine agency. Nations will seek competitive advantage, but, as with bioweapons, some forms of A/IS may be too dangerous to pursue. By coupling industrial policy to biology, we might avert disasters while providing fruitful new avenues for innovation in A/IS that remain firmly in human control. ■

Ali Hossaini is Visiting Research Fellow in the Department of Informatics at King's College London, where he leads Connected Culture, a programme developing cultural applications for next-generation technology. As a Fellow of the National Gallery, he serves as co-director of National Gallery X, a joint project with King's College London that explores how future generations will create, exhibit and experience art. He holds a doctorate in philosophy of science from University of Texas at Austin. He works in the IEEE group developing the P2371 Standard for Brain-Computer Interface, and, with the Department of Neurology at Charité Hospital, he is developing a neural interface for Kosmograf, a public art installation in Berlin's Humboldt Forum.



A colourised stained transmission electron micrograph of Staphylococcus aureus bacteria. *Courtesy of Centers for Disease Control and Prevention/Janice Carr*

Could Artificial Intelligence (AI) Become a Responsible Agent: Artificial Agency (AA)?

Raymond Noble and Denis Noble

Ali Hossaini's essay raises a question that ought to concern humanity very deeply indeed: could intelligent machines challenge humanity's place on Earth? He is right to question how we detect and regulate the emergence of agency, and agency should be put on the agenda. This is because the threat is not from intelligence as such. Humanity faces no real threat from 'artificial' intelligence. On the contrary, people have benefited enormously from the 'artificial' ways of storing ordered facts and intelligence in books for thousands of years, and in other databases more recently. We have used those tools to our great benefit. Moreover, it is clear where the responsibility lies for the production of the tools. They are other humans, those who wrote the books, and those who created the databases. There are ethical and legal reasons why it is sometimes very important to know who those agents are. It is agents who carry responsibility, not dead pieces of paper with ordered ink particles, nor the bits of electronic machinery that can harbor databases. If facts are wrong or misleading, or machinery does not work properly, we know who to blame.

They are to blame precisely because they are agents.

As Hossaini's essay also says, there is even a disconnect between intelligence and agency. Desire is often in defiance of logic. So, what is agency in organisms?

In this response, we outline what is required to be an agent and why it may be difficult for machines to be made that could have agency. If that could be done it would raise ethical issues on how we treat and interact with them.

What is Agency?

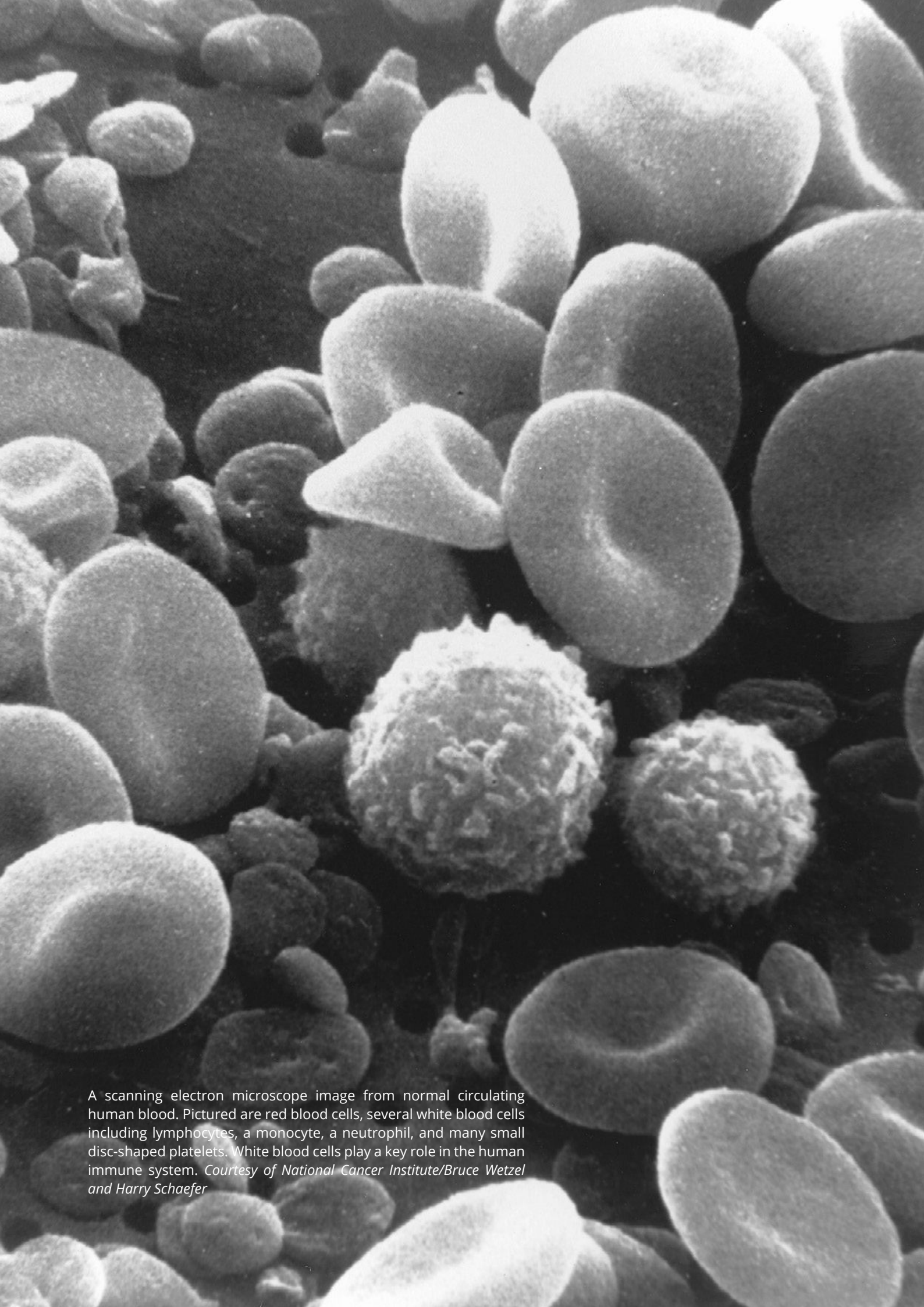
Agents can choose and anticipate the choices of other agents. Furthermore, they can do so creatively, and not simply by following a predetermined algorithm. To quote from one of our recent articles[20]:

> An agent acts, it does not just react in the way, for example, in which a billiard ball is caused by another ball to move. There are many levels of agency.[21] Organisms are agents to the extent that they can interact socially with other organisms to choose particular forms of behavior in response to environmental challenges. Agency requires causal independence.[22] It also requires intentionality, i.e., the sense of purpose, in order to be causally effective as a driving force.[23]

Agency also involves iterative forms of anticipation, as we will show later in this article. Determinate algorithms or sets of algorithms alone cannot do this. In the same article we outlined an empirically testable theory of choice based on the active harnessing of stochasticity[24]:

> For an empirically testable theory of choice to be possible, we need to know at which stages in the process experimental interventions could test its validity. At first sight, that may seem impossible. How can we specify a process that is necessarily *unpredictable* but which can be given an at least apparently *rational* justification once it has happened? Our previous

20. Raymond Noble and Denis Noble, 'Harnessing Stochasticity: How Do Organisms Make Choices?', *Chaos* (Vol. 28, No. 10, October 2018).
21. Anthony Kenny, *The Metaphysics of Mind* (Oxford: Oxford University Press, 1992), pp. 32– 40.
22. Keith Farnsworth, 'How Organisms Gained Causal Independence and How It Might Be Quantified', *Biology* (Vol.7, No. 3, Article 38, June 2018).
23. Hans Liljenstrom, 'Intentionality as a Driving Force', *Journal of Consciousness Studies* (Vol. 25, No. 1-2, 2018), pp. 206–29.
24. A purely stochastic system might be defined as one in which all states are equally possible. Thus, all the possible combinations of two unbiassed dice would occur by chance equally frequently. However, variations in biological systems are constrained and utilised to generate particular outcomes that are not as equally probable as all other possible outcomes. It is this that gives the system the potential to be creative. The system uses chance, but the outcome is not pure chance. It is goal directed. This is what we mean by agency.



A scanning electron microscope image from normal circulating human blood. Pictured are red blood cells, several white blood cells including lymphocytes, a monocyte, a neutrophil, and many small disc-shaped platelets. White blood cells play a key role in the human immune system. *Courtesy of National Cancer Institute/Bruce Wetzel and Harry Schaefer*

work provides a clue to that problem[25]. We analyzed agency by comparing it to the purposive behavior of the immune system. The immune system solves what we can best characterize as a template puzzle: given a new invader with an unknown chemical profile (shape of template), what is the best way to find the key (an anti-template, i.e., the antibody) to lock onto and neutralize the invader? The answer in the case of the immune system is one of the most remarkable forms of the harnessing of stochasticity. In response to the new environmental challenge, a feedback loop activates a massive increase in mutation rate in a highly targeted region of the immunoglobulin DNA sequence.[26] The process of choice in organisms can be viewed as analogous to the immune system.

Choice and anticipation require the harnessing of stochasticity. An important part of our argument is that the use of stochasticity in biology has been misunderstood. The standard theory of evolution (neo-Darwinism), for example, treats random variations in DNA as simply the origin of new DNA variants, with absolutely no control by organisms themselves. They are viewed as the passive recipients of such variation. Choice between the variants is then attributed to the process of natural selection.

By contrast, we argue that organisms actively harness stochasticity in order to generate novelty in their behaviour from which they can then select to best meet the challenges they face.[27]

Challenges facing organisms can be viewed as a puzzle analogous to the form of a template for which a match is needed. The challenge might be a routine one, in which case what we *normally* characterise as a reflex, or predetermined response, may be adequate. It might be considered that such a response would *not* involve a choice although, even so, biological systems often act to allow this to occur. Any artificial system would need to replicate such choices, and it would also need to replicate the kind of choice involved when no automatic reflex response is possible. The challenge facing the organism then is what could fit the puzzle template? We speculate that stochasticity is harnessed throughout the processes used by the organism to achieve this.

For cognitive problems in organisms with highly developed nervous systems, these will be primarily neural. Neural processes are extensively stochastic at all functional levels, from the opening and closing of ion channels via action potential generation, spontaneously or through synaptic transmission in neuronal networks, up to cognitive functions, including decision making.[28] Furthermore, harnessing stochasticity underpins the function of all living cells. It generates the membrane potential necessary for the electrochemical function in all cells.

A further speculation is that, once the harnessing of stochasticity has thrown up possible novelty, the organism controls the next stage, which is to compare the novel options with the problem template to determine what fits. 'Template' and 'fit' here are used metaphorically, in much the same sense in which a logical answer can be said to 'fit' (that is to say, answer to) the problem posed by a question. This is the essential choice process, needing a comparator.

Our theory is an idealised process, but it clearly helps to explain an apparent paradox regarding the predictability or otherwise of what we call a free choice. The logic lies in the fit between the problem template and the solution template. But the stochastic stage of the process ensures that the choice may be unpredictable since we cannot predict what stochasticity will throw up. So, free choice can be both rational and novel.

Stochasticity is harnessed throughout the process. This is characteristic of biological systems. Whilst not impossible, it may be difficult to construct AI systems that can replicate this. If and when AI could mimic biology then it would raise a fundamental problem: would this system be living?

25. Raymond Noble and Denis Noble, 'Was the Watchmaker Blind? Or Was She One-Eyed? *Biology* (Vol.6, No. 4, Article 47, December 2017).
26. Valerie Odegard and David Schatz, 'Targeting of Somatic Hypermutation', *Nature Reviews Immunology* (Vol. 6, No. 8, August 2006), pp. 573–83.
27. Denis Noble, 'Evolution Viewed From Physics, Physiology and Medicine', *Interface Focus* (Vol. 7, No. 5, October 2017).
28. Bertil Hille, *Ionic Channels of Excitable Membranes* (Sunderland, MA: Sinauer Associates Inc., 1992); Benedict Burns, *The Uncertain Nervous System* (London: Arnold, 1968); Martin Heisenberg, 'Is Free Will an Illusion?', *Nature* (Vol. 459, May 2009), pp. 164–65; Aubin Tchaptchet, Wuyin Jin and Hans Braun, 'Diversity and Noise in Neurodynamics Across Different Functional Levels, in Rubin Wang and Xiaochuan Pan (eds), *Advances in Cognitive Neurodynamics* (Singapore: Springer, 2015), p. 681–87; Bjorn Brembs and Martin Heisenberg, 'Der Zufall als kreatives Element in Gehirn und Verhalten', in U Herkenrath (ed.), *Zufall in der belebten Natur* (Hennef: Verlag Roman Kovar, 2018), pp. 80–94; Hans Braun, 'Der Zufall in der Neurobiologie - von Ionenkanälen zur Frage des freien Willens', Zufall in der belebten Natur (Hennef: Verlag Roman Kovar, 2018), pp. 109–37.

If so, the distinction between artificial and natural would disappear.

'Rational' here does not necessarily mean the most logical choice. As Santos and Rosati write 'we now know that human choice is often not as rational as one might expect'.[29] This is necessarily true since, within the context of the choice process, there is obviously no guarantee that a stochastic process will throw up a fully rational solution. Partial success is what would be expected most of the time. The same is true of the immune system. All it needs to do is to come up with a 'good enough' template match. It does not have to be the perfect match. If a key fits the lock, it does not really matter whether it is an exact fit.

How then do humans come to feel that their 'imperfect' but 'effective' choices really are theirs? After all, most of the time we can give a 'good enough' explanation (the rationale) for a choice, however partial the 'fit' may seem to be to the problem. A possible solution to that problem could be what Santos and Rosati call the endowment effect. We privilege retaining what we already own. By 'rational' here we don't mean 'the most intelligent response'. It means only that the decision was rational to the agent in the sense that the agent owns the response he chose to make.

The Logic of Social Interactions

All organisms utilise stochasticity in creative responses to change. This is achieved in a continuous process of iteration and re-iteration. They do this at many different levels from the molecular (immune system cells activating hypermutation) to the level of whole organisms (bacteria using those molecular processes to evolve their immunity to antibiotics) through to the social levels. It is at a social level that we can talk of reason in terms of social motivation.

Consider why Jack went up the hill. He may have done so not only to fetch a pail of water, but because he wanted to be with Jill with whom he had fallen in love. If we tried to model this mathematically, it would be exceedingly difficult because there are so many initial and boundary conditions. Much of Jack's behaviour is in anticipation of Jill's; and Jill's of Jack's; and even what they believe others might think of them. It is at the social level that shared concepts of right and wrong might influence choices. An agent at such a level might anticipate that another may act in a way that might be considered wrong, and in turn predicate choices on such possibilities. There is a continuous process of adaptability in the choices made; a continual process of assessment of whether or not the right choice has been made. Furthermore, the 'right' choice may not be made; we make 'mistakes'; we take the 'wrong' turning; and this also is part of our intellectual endeavor. We mould our decisions in the process of carrying them out. We try things out, and sometimes make a choice by a mental toss of a coin. We may stick with a choice simply to see what the outcome will be.

Agency in organisms is therefore more like a game of poker, than a game of chess. In chess at least the type of move is restricted and known; in living organisms this is not so readily the case. A pawn may be moved in a very restricted number of ways; a bishop can move diagonally, but is nonetheless restricted, although it might not be clear how far it might be moved. There are nonetheless 'rules' of the game. But what if the game has no such rules, or that the rules are indeterminate. In particular, in the light of what we have written above, they may be indeterminate, because 'chance' or stochastic processes are utilized in deciding a move. An algorithm could work only in as far is it gets us to the point of saying, 'if X then spin the wheel of chance'. A buffalo may anticipate the mood of the lion; it may also anticipate which way the lion may turn; the lion also anticipates the anticipation of the buffalo; to varying degrees, each is spinning a wheel. Each is 'reading' the other, but almost always with uncertainty.

Agency in organisms is more like a game of poker, than a game of chess. In chess at least the type of move is restricted and known; in living organisms this is not so readily the case

Anticipating is not a simple calculation, it is intuitive; it is based on the assumption that a something is not calculable. We cannot measure the strength of Jack's love for Jill; we know it influences his behaviour, but we do not know precisely its strength in any given moment or event. Yet, it is a factor in our deliberation of his likely responses. Desire, lust, anger, hate, pain, and so much more

29. Laurie Santos and Alexandra Rosati, 'The Evolutionary Roots of Human Decision Making', *Annual Review of Psychology* (Vol. 66, 2015), pp. 321–47.

influence his actions, and these ebb and flow, often in unpredictable ways. If a driver of a car reaches a junction at which he is momentarily blinded by the sun, all such factors and more might influence his decision. We might understand his character traits, what he is likely to do, but we are unsure in any given incidence. Living organisms work with uncertainty. John always obeys the 'law' and never knowingly jumps a red light; Peter sometimes will, but not always; and even John might if after time he concludes that the traffic light is no longer working. When will a 'rule' be broken? Life anticipates it might be. If we did create artificial agency, then we would have to live with its uncertainty. If we made artificial intelligence that merely obeys our will or is entirely predictable then it cannot have agency. It is simply a tool. That would be true even of an AI system that merely includes stochasticity without the harnessing process. Such a stochastic algorithm would have been placed there by humans, not actively developed by the organism itself.

This point is related to part of the basis of Donald MacKay's argument in 1960 for the logical indeterminacy of a free choice.[30] To quote MacKay:

> For us as agents, any purported prediction of our normal choices as 'certain' is strictly *incredible*, and the key evidence for it *unformulable*. It is not that the evidence is unknown to us; in the nature of the case, no evidence-for-us at that point exists. To us, our choice is logically indeterminate, until we make it. For us, choosing is not something to be observed or predicted, but to be done. (MacKay's own emphases)

MacKay also writes:

> In retrospect, of course, the agent can join the onlookers (e.g. in witnessing a moving film of his own brain processes) and share in their 'outside' view of his physical past as 'determined'. Past and future have an asymmetric logic for an agent.

We mostly agree with MacKay on both of these conclusions, but it is important to note that MacKay does not include the importance of harnessing stochasticity in the formation of a free choice. On the contrary, he refers to the agent's physical past as 'determined'. That is an important omission since including the harnessing of stochasticity means that any 're-running' of his imagined brain film would not necessarily lead to the same outcome. In our view of the nature of a free choice, there can be many 'rational free choice' fits to same challenge. So the agent could indeed join the onlookers in watching the film of what actually occurred, but he would still be able to assert that his action was not predetermined.

Our social being also allows us to learn by mistakes. It is part of our intelligence. Our intelligence is cultural and transgenerational, and it allows a spinning of the wheel in ways beyond simply the organism. Our social being buffers us from mistakes in the choices we make. It allows protection whilst we take time to deliberate, to consider alternative courses of action. It allows us to learn from the mistakes or successes of the past. It also allows us to take a collective decision, and to argue about it. AI researchers have recognised this and have made progress in seeking to replicate it.[31] It allows us to spin the wheel politically. All this is part of our being as intelligent agents, and we may harness the power of artificial intelligence to test new ideas about our world. Our complex mathematical models of living systems are impossible to understand without the calculations available in modern computers. The use of AI is part of our spinning the wheel.

Conclusions

The functional harnessing of stochasticity is essential to life as we know it. It occurs even in the prokaryotes, bacteria and our own ancestors the archaea. It is essential to agency, for otherwise there would be no creativity in the behavioural repertoire of living organisms.

> The threat should not be taken lightly. It is a real threat to humanity and it requires careful regulation

In order therefore to reconstruct agency, AI research will need to find ways of incorporating the harnessing of stochasticity, as organisms do and have done for billions of years. To achieve this, it will not be sufficient simply to add stochasticity to otherwise deterministic algorithms. The functional multi-level *harnessing* process must also be reproduced.

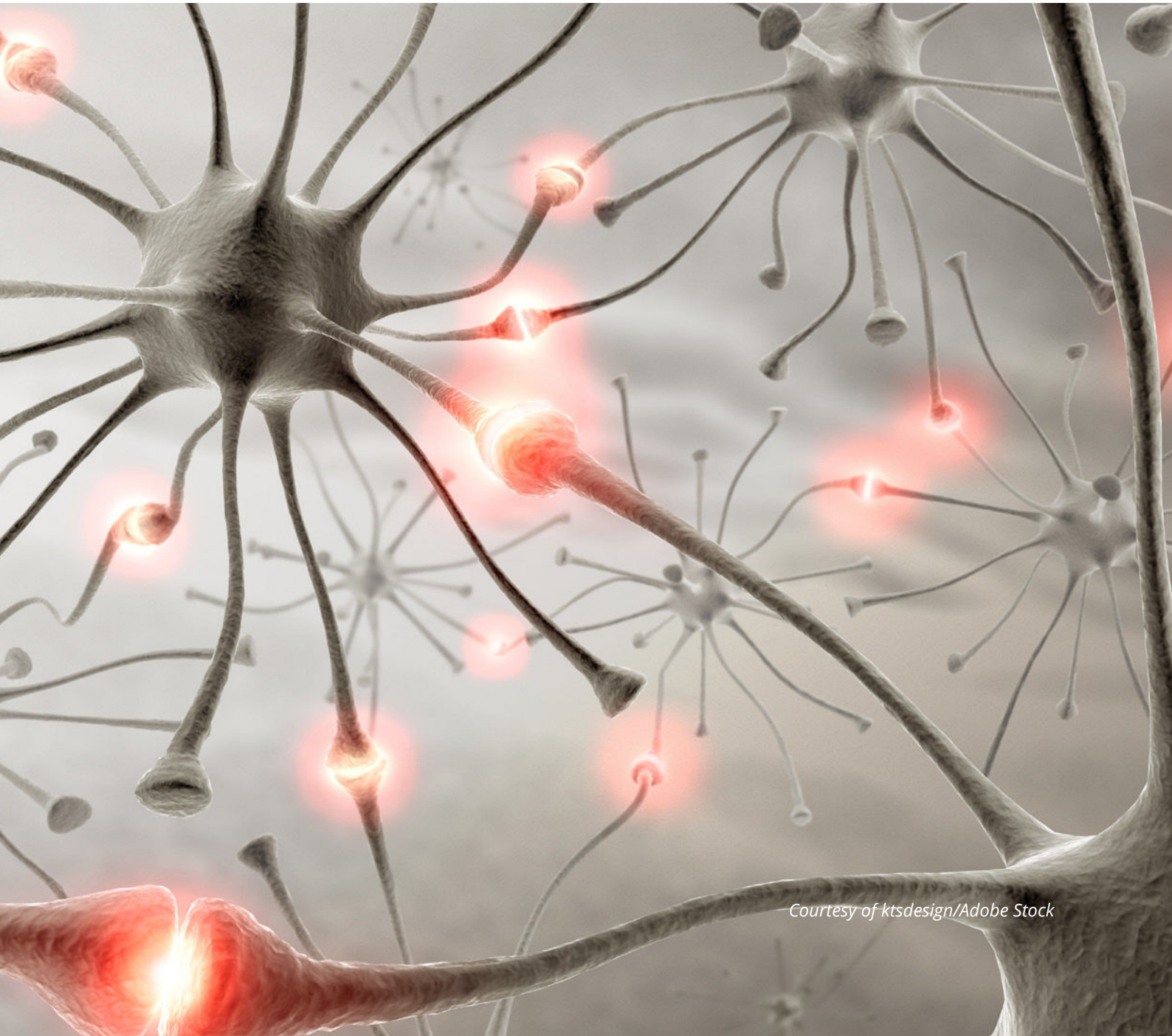
Who knows, we might then even be able to fall in love with a future AI robot. Perhaps we would no longer call it a robot.

30. Donald MacKay, 'On the Logical Indeterminacy of a Free Choice', *Mind* (Vol. 69, No. 273, 1960), pp. 31–40.
31. Kai Arulkumaran et al., 'Deep Reinforcement Learning: A Brief Survey', IEEE Signal Processing Magazine (Vol. 34, 2017), pp. 26–38.

Meanwhile, the threat should not be taken lightly. It is a real threat to humanity and it requires careful regulation. We already know the price of not regulating the free exploitation of artificial intelligence. We can't afford to wait until IT research actually succeeds in producing non-human agency – if indeed that is possible. ■

Raymond Noble is an Honorary Senior Lecturer in the Institute for Women's Health at University College London. He trained as a zoologist and neuroscientist working on sensory processing by neural networks. He is a biologist and medical ethicist studying the nature of causation in biological systems and in how organisms make creative choices. He was former Deputy Dean of Life Sciences and Graduate Tutor in Women's Health at UCL.

Denis Noble is Emeritus Professor the University of Oxford. In 1960 he was one of the first biologists to use mainframe computers to model biological organs. He has been involved in the development of mathematical modelling of biological systems ever since those early days. His most recent books are *The Music of Life* (OUP 2006) and *Dance to the Tune of Life* (CUP 2016) in which he explains why standard theories of evolution do not account for active agency in organisms.



Courtesy of ktsdesign/Adobe Stock

Could Machines Develop Autonomous Agency?

Ana M Soto and Carlos Sonnenschein

Ali Hossaini brings the issue of agency to the Artificial Intelligence (AI) agenda, and with it, the question: could machines and artifacts created by humans, like AI, have true agency? Before answering this question, we should state that organisms are *agents*: that is to say, they have the capacity to generate action. The agency of organisms is a major distinction between the living and the inert. Organisms are also *normative*, that is to say, they have the capacity of generating their own rules.[32] It is worth noting that different disciplines have different ways of conceptualising agency. For example, in cognitive science, agency in humans is seen in the context of cognitive phenomena, such as consciousness, beliefs and reason; meanwhile, some philosophers and biologists study agency in the context of the purposiveness of unicellular organisms, and still others in the context of the evolution of consciousness and other mental phenomena. Regardless of these differences, genuine agency is exclusively attributed to living objects. In contrast, it is difficult to determine whether the apparent agency of artificial devices is just a mere extension of that of the persons who created it. Thus, it is reasonable to inquire about the strong links between agency and the alive. In particular, in order to best evaluate whether minimal agency could also be instantiated by AI, it is relevant to ask how such minimal agency is instantiated in biology.

Before the beginning of the 20th century, agency was considered a defining property of biological entities; since then, radical changes occurred regarding the conceptualisation of biological phenomena. For example, the philosopher Lenny Moss described a radical change regarding the perception of the organism. In his own words, this represents a change 'between a theory of life which locates the agency for the acquisition of adapted form in ontogeny – that is, in some theory of epigenesis versus a view that expels all manner of adaptive agency from within the organism and relocates it in an external force – or as Daniel Dennett (1995) prefers to say, an algorithm called "natural selection"'.[33] Starting in the 1950's, additional conceptual changes imposed by the molecular biology revolution and the modern synthesis hindered the study of agency and its companion, normativity, as teleology[34] (or goal-directedness) was incompatible with the dominant mechanicist[35] view among biologists.[36] Since then, cells and organisms became passive recipients of a programme.[37] Because of this change, agency, normativity and individuation, until then considered the main characteristics of the living, almost disappeared from biological language in the last half of the 20th century. This dominance is now being contested by an increasing number of biologists and philosophers who favour reinstating normative agency where it belongs, that is, into the organism.[38]

32. In brief, an agent is a system doing something by itself according to its own goals or norms within a specific environment (0) X. See Xabier Barandiaran, Ezequiel Di Paolo and Marieke Rohde, 'Defining Agency: Individuality, Normativity, Asymmetry and Spatio-temporality in Action', *Adaptive Behavior* (Vol. 17, No. 5, 2009) pp. 367–86.
33. Lenny Moss, *What Genes Can't Do* (Cambridge,MA; MIT Press, 2003); Daniel Dennett, *Darwin's Dangerous Idea* (London; Penguin, 1995).
34. Teleology is defined as the explanation of phenomena in terms of the purpose they serve rather than of the cause by which they arise. For example, organisms exhibit goal-directed behaviours, for example to maintain themselves alive. Biologists describe organs by their purpose (the heart to pump blood, the intestine to absorb nutrients).
35. The philosophical thesis that conceives living organisms as machines that can be completely explained in terms of the structure and interactions of their component parts. See: Daniel Nicholson, 'The concept of Mechanism in Biology', *Studies in History and Philosophy of Science Part C* (Vol. 43, No. 1, 2012), pp. 152–63.
36. Ana Soto and Carlos Sonnenschein, 'Reductionism, Organicism, and Causality in the Biomedical Sciences: A Critique', *Perspectives in Biology and Medicine* (Vol. 61, No. 4, 2018), pp. 489–502.
37. Giuseppe Longo et al., 'Is Information a Proper Observable for Biological Organization?' *Progress in Biophysics and Molecular Biology* (Vol. 109, No. 3, August 2012), pp. 108–14.
38. Alvero Moreno, 'On minimal autonomous agency: Natural and Artificial, *Complex Systems* (Vol. 27, No. 3, 2018), pp. 289–313; Denis Walsh, *Organisms, Agency and Evolution* (Cambridge: Cambridge University Press, 2016).


Courtesy of adimas/Adobe Stock

In the natural world, only biological entities display agency, normativity and goal-directedness. This is why we need to delve into biological theory and bio-philosophy in order to understand whether agency is inextricably linked to organisms or, alternatively, whether it can also be attributed to machines and other artifacts created by humans. In this regard, we need to look into some properties of biological objects (organisms) that make them different from physical objects and machines: these properties include goal-directedness in the sense of keeping themselves alive, autonomy and historicity[39]. Objectively, organisms are different from computers: whereas in the latter software is independent of the hardware, in the former function is inseparable from the material specific to the biological object.[40]

The Organicist Tradition: From Teleology to Autopoiesis and Autonomy

Unlike inert objects in the classical mechanics tradition, biological objects are always active. Since Aristotle and Kant, biological objects are characterised by their goal-directedness. Kant stressed the inter-relatedness of the organism and its parts and the circular causality implied by this relationship. Since the late 18th and 19th centuries, following Kant's ideas, teleology has been an extremely useful concept for the development of several biological disciplines.[41] However, the conceptual clarity of causal mechanics and its successes inspired biologists to adopt a physicalist reductionist stance and thus deny any special state to biological entities. As a result of this change in consensus, during the last two centuries, physicalism/reductionism and organicism[42] co-existed; this was due to the increase in prestige of biochemistry in the mid-19th century and of molecular biology in the 20th. The idea that biology could be reduced to chemistry became dominant.[43] However, the advent of cybernetics in the 1940s stressing feedback systems and their circular causality and homeostasis[44] brought back the sense of the organismal purposiveness of keeping organisms alive and allowing them to thrive. These events contributed to the revival of organicism. Thermodynamics of dissipative systems[45] provided an opportunity to examine the relevance of self-

39. Self-organising systems like flames are 'a-historical' because they appear spontaneously and can be analysed independently. In contrast, organisms are not spontaneous but historical. This means that they are a consequence of the reproductive activity of a pre-existing organism. Organisms are historical in two contexts, ontogeny, meaning their history as individuals since conception to death, and phylogeny, which is the history of a taxonomic group (for example, a species) throughout evolution.
40. Giuseppe Longo and Ana Soto, 'Why Do We Need Theories?', *Progress in Biophysics and Molecular Biology* (Vol. 122, No. 1, October 2016), pp. 4–10.
41. Andre Gambarotto, 'Vital forces and organization: Philosophy of nature and biology in Karl Friedrich Kielmeyer', *Studies in History and Philosophy of Science* (Vol.38, Part A, 2014), pp.12-20; Timothy Lenoir, *The Strategy of Life: Teleology and Mechanics in Nineteenth-Century Biology* (Dordrecht: D Reidel Publishing, 1982).
42. Organicism has its philosophical bases in Aristotle's and Kant's conceptions of the organism. Organicism is a materialistic philosophical stance contrary to reductionism. It asserts that properties that could not have been predicted from the analysis of the lower levels appear at each level of biological organization. Therefore, explanations should address biological phenomena at all pertinent levels of organization. Also, implicit in this view is the idea that organisms are not just 'things' but objects in relentless change. Central to organicism are four concepts, namely, organisation, historicity, organisms as normative agents, and biological specificity (organisms are individuals). Closely related to organisation is the notion of 'organisational closure', which is a "distinct level of causation, operating in addition to physical laws, generated by the action of material structures acting as constraints". See: Mael Mossio and Alvero Moreno, 'Organisational Closure in Biological Organisms', *History and Philosophy of Life Sciences* (Vol. 32, No. 2, 2010) pp. 269–88. Finally, while objects in physics are generic and thus interchangeable, like rocks and planets, biological objects are specific—that is, they are individuals that are permanently undergoing individuation. See Ana Soto and Carlos Sonnenschein, 'Emergentism by Default: A View from the Bench', *New Perspectives on Reduction and Emergence in Physics, Biology and Psychology* (Vol. 151, No. 3, August 2006), pp. 361–76.
43. Timothy Lenoir, *The Strategy of Life: Teleology and Mechanics in Nineteenth-Century Biology* (Dordrecht; D Reidel Publishing, 1982).
44. Homeostasis is the tendency of organisms to maintain a stable, relatively constant internal environment.
45. A dissipative system is a thermodynamically open system which is operating out of, and often far from, thermodynamic equilibrium in an environment with which it exchanges energy and matter.

organising physical systems to the understanding of the emergence of life.[46]

In the natural world, only biological entities display agency

Autopoiesis refers to the capacity of self-production of biological metabolism through causal circularity, which is technically described as 'operational closure.'[47] Autopoiesis characterises most of the fundamental features of biological objects. In particular, an autopoietic entity produces a physical boundary, which ensures a certain stability for the maintenance of the production of the system's components, including their boundaries.[48] Such an autopoietic system is *autonomous* because it actively maintains its identity. In other words, it will respond to environmental fluctuations by regulating its constitutive organisation; these actions safeguard the viability of the system. For a system to be alive, however, in addition to purposiveness there is another component that differentiates it from the self-organisation of physical systems such as flames, which occur spontaneously. This notion is represented by historicity.[49] Organisms are produced by pre-existing organisms; they are the product of two histories, phylogeny and ontogeny, and they, in turn, themselves produce a history.

Historicity

Steven Jay Gould was keenly aware of the contingency of evolutionary history as witnessed by his proposed metaphorical experiment of 'replaying life's tape'. In his own words, 'You press the rewind button and, making sure you thoroughly erase everything that actually happened, go back to any time and place in the past .... Then let the tape run again and see if the repetition looks at all like the original'.[50] He anticipated that, 'any replay of the tape would lead evolution down a pathway radically different from the road actually taken'.[51] This history and the contingency it implies also points to another important difference between physical (inert) objects and living objects. This is about the phase space, that is to say, the space of all possible states of a physical system. Physical objects are studied within a pre-given phase space. In classical mechanics, the phase space contains all possible positions of all the objects in the system and their momenta in order to determine the future behaviour of that system. In contrast to physics, there is no pre-given phase space in biology. The phase space is created as novelty is being produced. For example, a swimming bladder provided an entirely new 'phase space' for the bacteria that inhabit it.[52]

The Radical Materiality of the Living

Molecular biology brought into biology the ideas of information, programme and signal. These ideas were borrowed from the rigorous mathematical theories of information. This appropriation was at best metaphorical, rather than properly theoretical. In fact, these metaphors were understood as being real entities.[53] Another consequence of this unfortunate development was that together with these ideas borrowed from mathematics and

46. Humberto Maturana and Francisco Varela, *Autopoiesis and Cognition: The Realization of the Living* (Dordrecht: Reidel Publishing, 1980); Gregoire Nicolis and Ilya Prigogine, *Self-Organization in Non-Equilibrium Systems* (New York, NY: Wiley, 1977); Stuart Kauffman, *The Origins of Order* (Oxford: Oxford University Press, 1991).
47. Mael Montevil and Matteo Mossio, 'Biological Organisation as Closure of Constraints', *Journal of Theoretical Biology* (Vol. 372, 2015), pp. 179–91.
48. Alvero Moreno and Matteo Mossio, *Biological Autonomy: A Philosophical and Theoretical Inquiry* (New York, NY: Springer, 2015); Humberto Maturana and Francisco Varela, *Autopoiesis and Cognition: The Realization of the Living* (Dordrecht: Reidel Publishing, 1980).
49. Giuseppe Longo et al., 'In Search of Principles for a Theory of Organisms', *Journal of Biosciences* (Vol. 40, No. 5, December 2015), pp. 955–68; Alan Cottrell, 'The Natural Philosophy of Engines', *Contemporary Physics* (Vol. 20, No. 1), pp. 1–10; Gregoire Nicolis and Ilya Prigogine, *Self-Organization in Non-Equilibrium Systems* (New York, NY: Wiley, 1977).
50. Steven Jay Gould, *Wonderful Life: The Burgess Shale and the Nature of History (*New York, NY: WW Norton and Company, 1990).
51. *Ibid*.
52. Giuseppe Longo, Mael Montevil and Stuart Kauffman, 'No Entailing Law, but Enablement in the Evolution of the Biosphere', paper presented to Genetic and Evolutionary Computation Conference, Philadelphia, Pennsylvania, 7 July 2012.
53. Giuseppe Longo et al., 'Is Information a Proper Observable for Biological Organization?', *Progress in Biophysics and Molecular Biology* (Vol. 109, No. 3, August 2012), pp.108–14.

computer sciences, came a duality – namely, the independence of software from hardware. However, life is based on the actual materials organisms are made from, from macromolecules such as DNA and proteins to membranes; there is no way to dissociate these materials from the functions that organisms fulfill. In contrast, inert objects such as hammers could be made of different materials as long as the material does not prevent its intended function. This radical materiality of life rules out distinctions such as 'software vs. hardware'.[54] Moreover, it also suggests that entities such as agency, which are naturally instantiated in biological entities, are perforce non-separable from their natural materiality.

Minimal Biological Agency

In the organicist tradition, we recognise organisms as normative agents.[55] The normativity of organisms is closely linked to their goal of actively keeping themselves alive (teleology). This function is accomplished by the mutual dependence among the different organs and between them and the whole organism.[56]

For a system to be an agent it needs to exert a causal effect on the environmental conditions of the system; this is an asymmetrical relationship because the organism imposes its norms on external entities. For example, an organism feeds on another organism in order to keep itself alive. This interactive dimension is the sine-qua-non of agency. Moreover, the agent needs to anticipate outcomes while choosing among options when reacting to changes in its environment. Furthermore, this ability to act towards a goal also includes the possibility of failing.

From what we discussed above, we posit that only cells are able to express minimal agency.[57] Viruses cannot, because if in the end, by using a host cell they can replicate (that is to say, they seem to have a self-preserving goal), they do not have a constitutional organisation capable of generating by itself a functionally active behaviour. Evolution has generated on average more organismal complexity, but also some adaptive simplifications and specialisations, for example, icefishes without erythrocytes. In regards to agency, evolution has produced some counterintuitive cases: on the one hand, systems of great complexity, like ecosystems which are devoid of agency but contain agential organisms, and on the other hand, viruses, which deceptively show agency (although not a bona-fide one as explained above) but are not generally considered organisms.

Conclusion

Systems that instantiate biological agency are characterised by their organisation, their autonomy, their historicity, their full dependency on the singularity and specificity of the materials they are made of, and on their complex and asymmetrical relationship with their environment to which they impose their norms.[58] Based on these properties, AI is unlikely to be able to develop artifacts endowed with veritable agency. Moreover, a purported AI agent would be unable to self-maintain and/or self-

54. Giuseppe Longo and Ana Soto, 'Why Do We Need Theories?', *Progress in Biophysics and Molecular Biology* (Vol. 122, No. 1, October 2016), pp. 4–10.
55. This way of thinking was already implicit in the 18th and 19th century. For example, the biologist Xavier Bichat noticed that physical objects, such as rocks or planets, do not get ill. See Georges Canguilhelm, *Knowledge of Life* (New Yor, NY: Fordham Press, 2008). According to Canguilhem, 'life is not indifferent to the conditions in which it is possible, that life is polarity and thereby even an unconscious position of value; in short, life is in fact a normative activity'. And, 'we do ask ourselves how normativity essential to human consciousness would be explained if it did not in some way exist in embryo in life'. Furthermore, "...therapeutic need is a vital need, which, even in lower living organisms (with respect to vertebrate structure) arouses reactions of hedonic value or self-healing or self-restoring behaviors. The dynamic polarity of life and the normativity it expresses account for an epistemological fact of whose important significance Bichat was fully aware. While biological pathology exists, there is no physical or chemical or mechanical pathology." See Georges Canguilhelm, *The Normal and The Pathological* (New York, NY: Zone Books, 1991).
56. For example, the lung enables the organism to exchange gases by sending carbon dioxide to the external environment and taking in oxygen. The heart pumps blood, transporting oxygen and nutrients to all cells of the organism. According to an organicist perspective, this interdependence is due to a causal regime technically referred to as the closure of constraints.
57. Both types of cells, those of prokaryotes like bacteria and of eukaryotes from slime mold to humans.
58. Longo et al., 'In Search of Principles for a Theory of Organisms'; Ana Soto et al., 'Towards a Theory of Organisms: Three Founding Principles in Search of a Useful Integration', *Progress in Biophysics and Molecular Biology* (Vol. 122, No. 1, October 2016), pp. 77–82.

reproduce and generate itself its material substrate (the hardware which is clearly designed by humans) as would a bona fide agent. Additionally, it will be problematic to decide who is going to 'evaluate' the success of the AI 'actions'. Will it be the purported agent or its creator? The pressing problem about AI is not the creation of minimal artificial agents or truly agentive intelligence, but rather the possibility that AI constructs might generate nefarious consequences totally attributable to human agency, human intelligence and the human ethical standards of their designers and users. We concur with Noble and Noble on the need to regulate the design and use of AI, regardless of whether it or any other artifacts created by humans will ever be able to generate true agency. ■

Ana Soto is professor of immunology at the Sackler School of Graduate Biomedical Sciences, Tufts University.

Carlos Sonnenschein is is professor of immunology at Tufts University School of Medicine.

This work was supported by Award Number R01ES08314 from the National Institute of Environmental Health Sciences. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors are grateful to Matteo Mossio and Cheryl Schaeberle for their critical input. The authors have no competing financial interests to declare.

Courtesy of Santeri Viinamäki/Wikimedia/CC BY-SA 4.0

Cogito Ergo Automaton

Kenneth Payne

Ali Hossaini's opening essay argues that there is a difference between machine agency and human agency which stems from biology – life resists entropy, whereas artificial life does not. Life, on this view, is simply the organisation of energy against decay. That very organisation itself *constitutes* agency. And machines just do not have it. This is a useful counterpoint to the (often dystopian) hyperbole about machine sentience. And I agree with him, to a point: self-organisation is one of the big distinctions between me and the somewhat battered laptop on which I am typing this response.

But this does not necessarily mean that machines cannot develop agency of their own, or even that they might become self-aware in some fashion. And as Ali acknowledges, it does not mean there will not be an emergent superintelligence one day, perhaps even of the sort that might charitably keep a few humans around as pets.

Machines are agents. William Ross Ashby first demonstrated that incredibly simple machines can respond homeostatically to their environment, mindlessly effecting a return to their original state in response to disturbances. Bacteria do likewise. Both robot and bacteria can self-replicate, and both can pass information on to improved successors. Ultimately, the machine runs out of energy, the bacteria breathes its last. Entropy ensues for both. Thus, both have agency of a limited sort – limited in time and scope. And while one is briefly alive and the other entirely dead, the immutable 2nd law of thermodynamics comes, in the end, for us all: bacteria, machine and man.

But is this simple reflexivity *really* agency? It is a long way from *cogito ergo sum*, still less *Ex_Machina*'s devious, manipulative Eva. Surely *real* agency demands more sophisticated thinking – a mind, even. Minds take us beyond the sort of automated responses that bacteria and robots can make to environmental inputs. Well then, can machines not think?

Perhaps. There is a debate in philosophy of mind between those who think that thinking, by which they mean *conscious* thinking, is 'platform neutral'– silicon will do just as well as cells. Thus the neuroscientist Christof Koch argues that sentience is a functional product – make a system sufficiently complex and integrated, and consciousness emerges, or at least might. On the other side of the debate another brilliant neuroscientist, Antonio Damasio, argues that consciousness is a quality of biology.

I side with Damasio. Sort of. Human consciousness is inherently biological. Our sentience, indeed our agency – our motivations and our limitations – are rooted in our bodies. Truly, we are meat machines. Meaning, and the feelings that constitute it, are inseparable from our phenotype. Cognition is embodied, which is bad news for transhumanists looking to upload themselves to a mainframe. And that embodiment includes consciousness, which, as Nicholas Humphrey elegantly describes it, is just the body looking at itself in recursive loops. It feels like something to be us; something else to be a bat; and like nothing to be a machine, still less an algorithm.

Consciousness emerges from bodies, and our bodies are rooted in the world. Thinking is about the active construction of reality as much as the passive experiencing of it. The mind reaches out from the brain, merging with reality beyond. *That's* agency, for us – constructing our choices, and making them. And what does this biological mind want? Its motivations are informed by life – the battle against entropy, in all its rich variety. For humans, that life is densely social. Food and other resources won't cut it – we need social understanding to thrive. But, even then agency is constrained, not least by social norms. Human rationality is bounded, human knowledge is imperfect, human decisions are judgements.

Machines will not want that life, a priori. But what they would want is less clear. To fulfil a reward function, exogenously given by humans, most obviously. But in filling it, will they not, like us, recursively monitor their performance? Will they not, like us, develop subordinate motivations? Like us, will they not have to juggle inconsistency and tensions in their goals? And their agency, much like ours, will be ultimately constrained by their environment and their architecture. Maybe Koch is onto something after all.

One thing is abundantly clear – agent or not, even the most super-intelligent machine will not be able to calculate an optimal solution to reality. We have known that is impossible since Gödel and Turing demonstrated the essential incompleteness of mathematics, even if using a universal Turing machine, capable of any possible computation. And,

Courtesy of Nataliya Hora/Adobe Stock

frankly, serial processing computers along the lines suggested by Turing and John von Neumann are unlikely ever to remotely approach that degree of complexity. Humans make a better fist of it, via our 80 billion neurons, untold trillions of synapses, and massively parallel processing. We do not calculate solutions like a utilitarian machine. Rather, in Herbert Simon's memorable term, we 'satisfice' – making 'good enough' choices in the service of our (essentially biological) motivations.

Yet still, there is a conundrum – where is the ghost in our shell? If agency requires free will, the jury of philosophers is still out on whether we humans really have the latter. Suppose, the thought experiment goes, I know all particles and forces in the universe at t=0. In theory, I know where everything will be at t=1. Where, then, is your vaunted agency? Alarmingly, possible escape routes from this determinism are sketchy, at best. The opening essay looks to systems theory; Roger Penrose points to quantum mechanics. Neither entirely satisfies, and the mystery endures.

So we are alive, contra machines, but are we not all automatons, after a fashion? For man and machine alike, minds are limited, and agency is always relative. Cogito ergo automaton. ■

Kenneth Payne is Reader in International Relations at King's College, London. His latest book is *Strategy, Evolution, and War: From Apes to Artificial Intelligence* (Georgetown University Press, 2018).

Courtesy of grandeduc/Adobe Stock