Richard Johns
draft of August 2016
not to be distributed

Physicalism and the Problem of Biological Form

1. Introduction

Much has been written over the past few decades about the explanatory adequacy of physicalism. Supporters of physicalism point to the success enjoyed by sciences like biochemistry, which now offer detailed physical descriptions of living processes such as DNA replication and protein synthesis. Critics, on the other hand, point to the remaining gaps in physical explanations, especially in understanding human consciousness, intentionality and free will. These criticisms have a long history, and continue to be pressed by philosophers such as Nagel (2012) and BonJour (2010).

In this paper I will explore another long-standing problem for physicalism, the *problem of biological form*, which is the apparent difficulty of explaining the origin of biological structures. These forms exist at many scales, from the gross anatomy of an animal down to the primary structures of proteins, and have two general features. First, they are functional – they contribute to biological activities such as sensation and motion. Second, these functions require them to be complex in very specific ways, just as machines like automobiles and photocopiers have to be complex in order to function well. Such functional structures seem, at least at first sight, to be difficult to explain in physical terms on account of the extreme simplicity of the physical laws that would govern their formation from (presumably) simple beginnings. As Goldenfeld and Kadanoff (1999) put it, "...why, if the laws are so simple, is the world so complicated?" noting that "biological systems [form] a limiting case of exceptional complexity".

Goldenfeld and Kadanoff's question might seem to be based on confusion, as if one had asked why present automobiles are so complicated, if the laws of physics are so simple. Automobiles are complicated, one is likely to say, because they *need* to be, in order to perform all of the functions that customers demand. The simplicity of the laws of physics is quite beside the point here, as the properties of the vehicle are determined by consumer preferences, the creativity of engineers, and so on, rather than the laws of physics. In a similar way, one might argue that living structures are a product of evolution, not the laws of physics. As with automobiles, living organisms are just as complex as they need to be in order to perform their functions. If an animal's way of life requires flying, for example, then it needs flight control and navigation systems, which are necessarily rather complex. And if we ask why some animals fly in the first place, the rough answer (which goes back to Darwin, and is maintained in the modern sythesis[1]) is that the ability to fly is advantageous in some contexts, and so slight changes toward flying ability that arise initially by chance will be selected for. Thus, while the whole universe is ultimately governed by physical laws, their simplicity (or otherwise) simply isn't relevant to the complexity of life.

This solution to the problem of biological complexity seems to regard physical matter, governed by the laws of physics, as akin to a collection of generic[2] Lego blocks, which has no inherently preferred configuration, but may be assembled by a person or other external influence into a bewildering variety of arrangements. In other words, the laws of physics on this view are *neutral* toward the existence of birds and automobiles (for example), just as Lego blocks have no endogenous disposition to assume the shape of a bird or a car rather than a formless blob. Once a machine or organism exists, its operation is of course fully governed by the laws of physics, but on this view the laws themselves have no propensity to arrange matter into functional machines or advanced[3] organisms in the first place. We will introduce the term *neutralism* to refer to this view of physical laws. According to neutralism, the bias toward functional structures displayed in evolutionary history is due to specific biological circumstances rather than the underlying laws of physics. These circumstances might include, for example, the existence of populations of self-reproducing organisms, with variation of heritable traits within each population, and some of these traits being correlated to reproduction rates.

Opposed to the neutralist view of physical laws is the *constructionist* view. This term 'constructionist', which was coined in a more general sense by Anderson (1972, p. 393), refers to the view that the laws of physics themselves have an inherent disposition to arrange matter into

[1] I shall also use the term 'selectionism' to refer to this view, that the natural selection is the primary source of functional complexity in the biological world. As Gould put it, "natural selection creates the fit".

[2] I am referring here to the general-purpose blocks, shaped as rectangular prisms, not the specific-purpose pieces that many sets now include.

[3] By *advanced* organisms I mean the more complex and sophisticated organisms, such as mammals and birds for example, rather than simpler organisms such as prokaryotes. Advanced organisms are the focus of the paper since they are apparently the hardest to explain in physical terms.

functional objects such as organisms and machines. We may say that the laws have a strong 'functional bias', according to constructionism, one that is sufficient to account for the observed existence of functional objects in the universe. Now evolutionary history displays an enormous variety of solutions to the basic problems of obtaining and digesting food, reproducing, moving around, sensing one's environment, etc., so that evolutionary forces must be similar to pragmatic engineers, who aim to solve a problem using any kind of device that will work, and are not wedded to particular materials or technologies. In light of this, constructionism should claim that there is a bias toward function generally, rather than toward specific ways to realise those functions.

A third view of physical laws is also possible, of course, according to which these laws are biased *against* the formation of functional objects. However, no one has advocated such a view, as far as I am aware, and apparently for good reason. This view would combine all the difficulties of the neutralist and constructionist views, without enjoying any of their advantages. In this paper, therefore, only the neutralist and constructionist views will be considered (together with intermediate views, according to which the laws of nature have a weak functional bias, but one that is not sufficient by itself to produce life).

In the following sections constructionism and neutralism will both be examined in detail, and I will argue that neither one (nor an intermediate view) is viable. I will conclude therefore that physicalism cannot solve the problem of biological form.

2. Constructionism

Constructionism, as defined above, is the view that the laws of physics have an inherent bias toward the formation of functional objects. Now we know that matter, as governed by physical laws, is disposed to form some structures more readily than others. For example, the universe is full of spherical stars, and contains not even a single cubic star, for reasons that are easily explained in terms of physical laws. Snowflakes are also common, and always have six points, again due to the laws of physics and the structure of a water molecule. We might say that the laws of physics 'prefer' or 'favour' spherical stars over cubic stars, and six-pointed snowflakes over seven-pointed ones. Constructionism claims that physical laws have a similar preference for functional objects, and that this functional preference accounts (at least to a great extent) for the existence of life as we find it.

According to constructionism, the functional bias in evolution arises from the laws of physics themselves, rather than from specific biological conditions, such as those appealed to in the modern synthesis. For an analogy, compare the forms of stars and snowflakes mentioned above with a sloping soccer field on which the ball has a tendency to move in an easterly direction. The eastward bias of the soccer ball, in this case, is not due to underlying laws of physics (since space is isotropic), but arises instead from the local geography. At another field, across town, a quite different bias may be present.

2.1 Constructionism Defined

In order to define constructionism precisely, we need to be very clear about what it means for the laws of physics to 'prefer' one possible state over another, or to have a greater readiness to produce one of those states than the other. In the general case of stochastic laws, I will show that the degree of dynamical favour for a state is best understood as the *stationary probability* of the state, to borrow a term from the formalism of Markov processes. To support this claim, I will first briefly summarise this formalism.

A Markov chain has a set of possible states $\{0, 1, 2, 3, \ldots\}$, and a matrix of transition probabilities p_{ij} that specifies the probabilities of direct transitions from state i to state j. Time consists of a series of discrete moments. Note that the transition probability p_{ij} depends only on i and j, and so is independent of the history of the system prior to being in state i. Physical systems are all assumed to have this 'memoryless' or 'Markov' property, as long as the state space is sufficiently fine grained. We will also make other generally-accepted assumptions about physical systems, as described below.

In a physical system, these transition probabilities are determined by the laws of physics together with the physical characteristics of the system. For any physical state j there will be what one might call 'incoming' and 'outgoing' transition probabilities. For example p_{ij}, the probability that state i evolves directly to state j, is an incoming transition probability for j and an outgoing one for i. The sum of the outgoing transition probabilities for each state must be 1, but the sum of the incoming transition probabilities could be as low as 0, or as high as *N*, for a system with *N* possible states. A system with states {1, 2, 3, 4} for example might be such that every state evolves to state 1 with probability 1 (including state 1 itself), as shown in Fig. 1. In that case, the total incoming probability for state 1 is 4, and for the three remaining states it is 0.

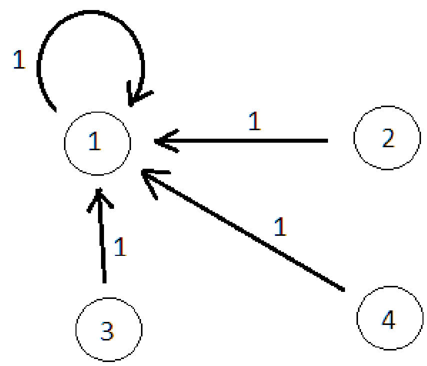

Figure 1

In the Markov chain of Fig. 1, the laws (represented by transition probabilities) obviously have a strong preference for state 1 over the others, as all the probability is flowing into state 1, but how is this preference to be represented numerically? In this example the sum of the incoming transition probabilities may look like a suitable measure, as it captures the fact that the dynamics is steering the system toward state 1 and away from the other states. This approach does not work in general, however, as can be seen from considering the system shown in Fig. 2 below.

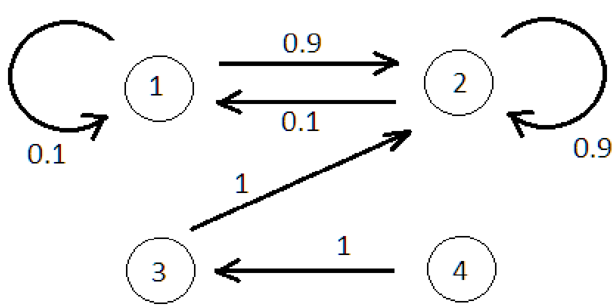

Figure 2

To what extent is each state in Fig. 2 favoured by the dynamics? If we simply add up the incoming transition probability for each state then state 2 is the highest, with a total of 2.8, while state 4 is the lowest, with 0. Also state 3 has a total of 1, which beats state 1's total of 0.2. But do the transition probabilities really favour state 3 over state 1? Perhaps not, since state 1 has (we might say) "more useful connections" than state 3. Some of state 1's incoming probability is from state 2, which is highly favoured, whereas all of state 3's incoming probability is from state 4, which is completely shunned by the dynamics. If we start the system in state 4, then what will happen? The state immediately evolves to state 3, then state 2. Thereafter it is mostly in state 2, with occasional brief excursions to state 1, but it never returns to state 3. The same final pattern occurs (even more rapidly) for the other possible initial states, so state 3 is apparently just as repugnant to the dynamics as state 4, despite its higher incoming probability.

It becomes clear that dynamical preference is a rather holistic property. In order to be favoured, a state needs incoming probability from *favoured* states, which in turn need incoming probability from other favoured states, and so on. Fortunately there is a well-studied probability measure that exactly captures what we need here: the *stationary probability*. I will now define this probability and explain what it represents.

Suppose we have somehow defined a probability function Π that measures the degree to which each state is favoured by the dynamics. In that case, since the transition probability p_{ij} confers dynamical favour on state j only to the extent that state i is itself favoured, the product $p_{ij}.\Pi(i)$ measures the degree of dynamical favour that a state j actually gains from the incoming probability p_{ij}. Summing up these terms, for all states i, we obtain a term that equals $\Pi(j)$, the dynamical favour of state j. In other words,

$$(1) \qquad \Pi(j) = \sum_{i=1}^{N} \Pi(i) p_{ij}$$

Now this equation (1) doesn't seem to provide a definition of Π, since Π also appears on the right-hand side – it would be a circular definition. However, with certain constraints on the transition probability matrix, such a probability Π provably exists, and is unique. Thus when these constraints are satisfied, equation (1) does define the probability Π, and it is called the stationary probability.

What constraints are needed, for the stationary probability to exist? For a finite state space, the Markov chain simply has to be *irreducible*, which means it is possible to eventually get from any state to any other state, with positive probability (i.e. all pairs of states are mutually accessible). If a system is reducible (not irreducible) then its state space consists of equivalence classes of mutually-accessible states and in that case it doesn't seem meaningful to compare the dynamical favourability of states in one such class with those of another. (At best, then, one could define a separate favourability measure for each class.) However, actual physical systems are thought to be irreducible, so we shall not consider this issue further.

The stationary probability is not normally introduced in the manner above, as a measure of dynamical favour. It is usually defined as the probability vector (if any) that maps to itself under multiplication by the transition matrix, so that it is invariant, or 'stationary', under this operation. In other words, if T is the matrix of transition probabilities, then a stationary probability vector Π is such that $\Pi.T = \Pi$. $\Pi(j)$ also represents the proportion of time that the system spends in state j, in the limit as time tends to infinity. (This limiting proportion is independent of the initial state.)

With the stationary probability function thus defined, we can now define very simply what it means for the laws of physics to favour one state over another. If S and S' are two possible physical states, and Π is the stationary probability, then the physical laws favour S over S' just in case $\Pi(S) > \Pi(S')$. The laws are neutral with respect to S and S', on the other hand, just in case they have equal stationary probability, i.e. $\Pi(S) = \Pi(S')$.

To apply the notion of dynamical favour to particular structures, rather than states of the whole system, is only slightly more complicated, since the stationary probability of a certain type of object (such as a certain type of star) can be defined as the stationary probability of the set of states in which such an object exists. To describe effectively the bias (or otherwise) of the physical laws toward functional objects, however, it is best to compare functional and non-functional arrangements of the same components. For example, functional proteins and non-functional proteins have exactly the same components (amino acids) but are arranged in different sequences. At a larger scale, the molecular components of a living organism can also be arranged in non-functional ways. To compare the preferences of the physical laws toward these alternative arrangements, one can use conditional probabilities. Let the proposition F define a particular functional arrangement of the components stated to exist by proposition C, and let G define a specific non-functional arrangement of the same components. Then the relevant comparison is between $\Pi(F \mid C)$ and $\Pi(G \mid C)$. If these are equal, then the laws of physics are neutral between arrangements F and G.

It is somewhat trickier to define the bias of the laws toward functionality in general, rather than toward particular functional structures. There are two key issues here. First, the laws could be biased in favour of some functional structures, while being biased against others. In that case, how could the overall bias be defined? A second issue is that the functional arrangements of a given set of components are generally very sparse in the set of all possible arrangements. So, for functional structures to be probable overall, the average (stationary) probability of the specific functional arrangements must be far greater than the average non-functional arrangement of those components. In light of these issues, the simplest way to define constructionism is to compare the stationary probability of the *class* of functional arrangements with the class of non-functional arrangements of a given set of components. Constructionism is the view that these two (conditional) probabilities are comparable, i.e. equal to within a few orders of magnitude, so that the average stationary probability for a functional arrangement is far greater than the average non-functional arrangement.

Constructionism, as stated in the introduction, is the view that the laws of physics are (like a pragmatic engineer) biased toward the formation of functional objects generally, rather than toward some specific technology such as camera eyes. However, constructionism entails that many (not necessarily all) specific functional objects will have much higher stationary probability than their non-functional counterparts (i.e. non-functional arrangements of the same components). If the overall stationary probability of functional arrangements is high, despite their extreme rarity, then the stationary probabilities of many specific functional arrangements must be high as well.

2.2 Stable States and Newton's Error

The Markov chain in Fig. 1 above might give the impression that a stable state (one that is likely to persist, according to the transition matrix, after the system enters that state) will also have high dynamical favour. After all, state 1 in Fig. 1 is completely stable, as well as highly favoured. However, in principle it is possible for a state S to be highly stable while also having very low stationary probability. Consider for example the Markov chain in Fig. 3 below, where the most stable state is #1, in the sense that it has the highest probability (0.99) of persistence from one instant to the next. However, state 1 has a rather low stationary probability of about 0.0014. State 6, by contrast, has a stationary probability of almost 0.89, despite its lower stability.

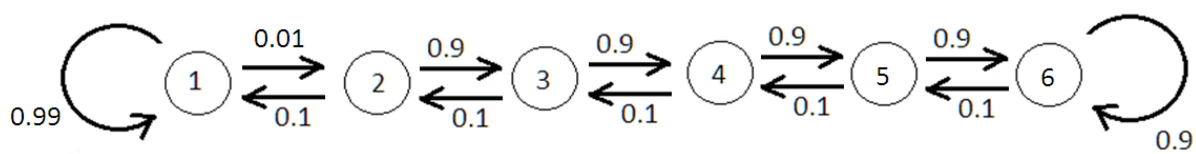

Figure 3

In general, a state (or group of states) can be very stable if transition probabilities are high between members of the group. These states may yet have low stationary probability if there is very little probability flow to the state or group itself, from the rest of the state space. That is the situation in Fig. 3, where the general probability flow is from state 2 up to state 6.

In a similar way, it would be a mistake to infer, merely from the stability of life, that life is dynamically favoured. Physical states where life exists are indeed stable, since life has persisted on earth for billions of years, and advanced life for around 500 million years. It does not follow however that life is also favoured by the dynamics. Despite the fact that living states are likely to persist once they are 'found' by the dynamics, it may be that these states are very hard to 'find' in the first place.

A second mistake concerning stability and favour is subtle, and more easily made. This is the idea that, if S is but one of a huge set of equally stable alternative states, then state S is not favoured. This is an error, since it may be that there is strong probability flow through the state space to S, while not to the alternatives, even though the alternatives are equally stable if they should come to exist. (See Fig. 3 for example.) Surprisingly, this second error was apparently committed by Isaac Newton. Newton writes in the *Opticks* (1704, Qu. 31, p. 402):

> … it's unphilosophical to seek for any other Origin of the World, or to pretend that it might arise out of a Chaos by the mere Laws of Nature; though being once form'd, it may continue by those Laws for many Ages. For while Comets move in very excentrick Orbs in all manner of Positions, blind Fate could never make all the Planets move one and the same way in Orbs concentrick, some inconsiderable Irregularities excepted, which may have risen from the mutual Actions of Comets and Planets upon one another, and which will be apt to increase, till this System wants a Reformation. Such a wonderful Uniformity in the Planetary System must be allowed the Effect of Choice. And so must the Uniformity in the Bodies of Animals…

Here Newton refers to the fact that the orbits of all the planets in our solar system are coplanar and unidirectional. For some reason, Newton concludes that such a structure could not "arise out of a Chaos by the mere Laws of Nature"; in other words, he believes that the laws are generally neutral, or at least not favourable, toward unidirectional orbits. The evidence for this conclusion is not stated in the passage, but presumably his reason is that other configurations are equally stable. If the planets (somehow) came to orbit the sun in different directions, then Newton's laws would entail that they continue to do so, at least for a very long time.

In the 18th century physicists such as Kant and Laplace showed that a nebula of gas and dust, collapsing under its own weight, would spontaneously develop a unidirectional rotation. Thus work on the 'nebular hypothesis' shows that unidirectional solar systems are favoured by the laws of physics after all, since rather arbitrary initial states evolve to unidirectionality under Newton's laws.

At the end of the quoted passage, Newton states his view that the bodies of animals are also structures that aren't favoured by the laws of physics – and concludes that they are the effect of "Choice", i.e. rational design. If Newton's argument for this is also based on the stability of alternative arrangements, then the argument fails. Interestingly, a similar argument is made much later by Polanyi (1968), who concludes that "the morphology of living things transcends the laws of physics and chemistry". Polanyi's main argument for this concerns the sequence of base pairs in a DNA molecule. Since the chain is held together by the sugar-phosphate ‘backbone’ rather than by bonds between adjacent base pairs, one cannot derive the base-pair sequence from chemical stability considerations. In fact, non-functional DNA sequences just as chemically stable as the ones used by life. As pointed out by Giere (1968) and Causey (1969), however, this tells us nothing about the propensity for these sequences to form in a physical system.

2.3 Objections to Constructionism

Now that it has been defined precisely, I will state four arguments against constructionism. These arguments arise from some general features that the laws of physics are understood to have, namely locality, invariance (across space and time, and under rotations) and the Markov (‘memoryless’) property. For brevity, we will refer to such laws as Local Invariant Markov (LIM) laws. To understand how LIM laws work, it is helpful to consider a simple system that is governed by such a law, such as a 2D cellular automaton. This is a square grid of cells, each cell representing a small region of space. Each cell has a finite number of possible states, and the state of the whole grid at one time consists simply of the states of all the individual cells at that time. The time evolution of the grid is governed by transition probabilities, of course, but these are in turn determined by a local updating rule which uses what we might call *local* transition probabilities. For each possible future state of a given cell there is a probability of a transition to that state, which depends only the present state of the cell, together with the states of the surrounding eight cells. For example, if the state of each cell is either 0 or 1, then each ‘local block’ of 3 by 3 cells has $2^9 = 512$ different possible states, and the local updating rule specifies the probabilities that the central cell will become 0 or 1 for each of these 512 possible states.

The invariance of the LIM law means that many of these 512 probabilities will be equal – for example, the local transition probabilities will be invariant under rotation at least, and may have other symmetries as well. Moreover, the local transition probabilities are the same for all cells, and do not vary with time.

The transition probabilities for the whole grid (which we might call ‘global’ transition probabilities) are determined from the local transition probabilities, using the fact that the local transitions are entirely independent of each other, after conditionalising on the present state of the grid. In other words, the probability of a transition from one state of the entire grid to another is simply the *product* of the local transition probabilities, as defined above, each depending only on its own local block.

A LIM law can favour particular structures only through a tendency to produce those structures by *self-organization*. (A structure forms through self-organization when it is the result of simple, local interactions between the components of the system, rather than through the control of external instructions, such as an architect's blueprint or a genome.) This is not to say that structures cannot be formed through external organization in a system governed by LIM laws. The point is that, in such cases, the bias toward a particular structure is due to the external instructions rather than the laws of physics. Since, under the operation of LIM laws, all interactions are local, any structure favoured by the laws themselves must arise from these local interactions. As Camazine et al (2001) put it,

> Self-organization is a process in which pattern at the global level of a system emerges solely from numerous interactions among the lower-level components of the system. Moreover, the rules specifying interactions among the system's components are executed using only local information, without reference to the global pattern.

With the nature of LIM laws defined, we will consider four arguments that LIM laws cannot favour functional arrangements.

2.3.1 The concept of function is foreign to physics

The first objection to constructionism is that the concept of function seems entirely foreign to physics, and belongs exclusively to "higher levels" such as biology and engineering. Why, and how, would the basic physical laws favour the emergence of *functional* structures? A functional structure is *for* something: it exists in order to solve some problem, such as how to metabolise available food, move more efficiently, or see in the dark. Such problems do not even exist until organisms do, so it hardly seems possible for the solutions to be implicit in the laws of physics before life even began.

2.3.2 Absence of empirical evidence.

The second objection to constructionism is based on the total absence of direct empirical evidence for it, either from real physical systems or computer models. This is an argument from ignorance, and hence cannot be conclusive, but it is powerful nevertheless.

Perhaps the best way to survey the empirical evidence for constructionism is to look at some of the evidence for it raised by its proponents, such as Stephen Wolfram, Nigel Goldenfeld and Leo Kadanoff. Wolfram (2002) investigates the fertility of LIM laws by studying one-dimensional binary cellular automata. There are only 256 possible such laws that are deterministic, and Wolfram examined the behaviour of every one, starting from a simple initial state – a single cell containing '1', and the rest zeros. Wolfram found most of them to produce very regular, repetitive, self-similar patterns. Moreover, the rules that have left-right symmetry (i.e. the output for a three-cell neighbourhood in the state 'abc' is the same as for the state 'cba') all result in regular patterns. For example, Figure 4 shows the output for Rule 22, for the first 64 time steps (time increases from top to bottom):

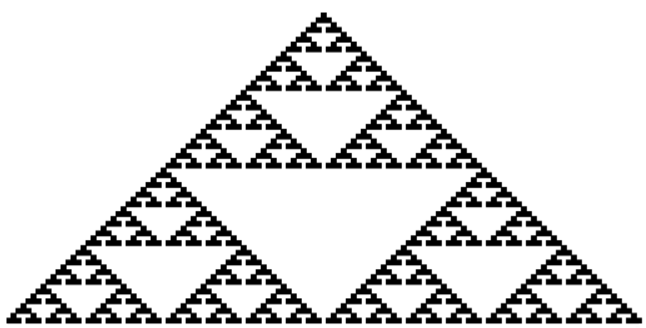

Fig. 4 Rule 22 output from a simple initial state

On the other hand, Wolfram discovered that rules without such left-right symmetry can exhibit more interesting, irregular behaviour. Rule 30, for example, generates the aperiodic pattern shown in Figure 5:

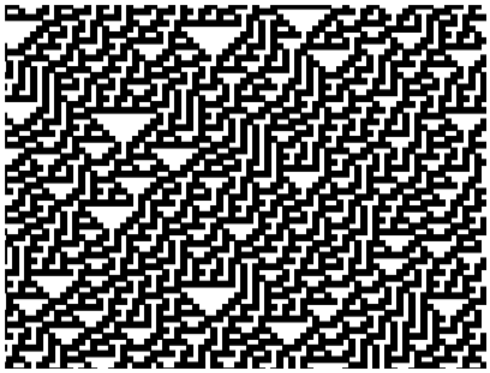

Fig. 5 Part of Rule 30 output from a simple initial state

Wolfram apparently uses the case of Rule 30 to argue for constructionism, stating on p. 388:

> ... the vast majority of the complexity we see in biological systems actually has its origin in the purely abstract fact that among randomly chosen programs many give rise to complex behavior"

Wolfram thus seems to attribute "the vast majority" of the functional structure in living organisms to the basic laws of physics that are operating. Our universe (luckily) has a set of physical laws that naturally gives rise to daffodils, fruit flies and crocodiles, in something like the same way that Rule 30 generates inverted white triangles, and structures resembling the letter 'J'.

Goldenfeld and Kadanoff (1999) also seem to take a constructionist view, albeit more tentatively. Referring to dissipative structures, such as vortices and convection cells, they suggest that living bodies are similarly products of physics.

> As we have seen from the examples quoted here and many others, in nonequilibrium situations many-particle systems can get very complicated indeed. It is likely that this tendency is the basis of life.

How strongly do such empirical examples support constructionism, however? Two points are salient here. The first is that, while some of the products of self-organization are intricate, and even beautiful, the known examples are not themselves functional in any way. They do not even slightly resemble either machines or organisms, although organisms do include some patterns that can be created this way. For example, the pattern on a cone shell is similar to that generated by Wolfram's Rule 30, and the familiar stripes of a zebra can be duplicated fairly closely by a 2D cellular automaton. (Camazine ref?)

The second point is that each set of laws produces the same particular pattern (or limited set of patterns) endlessly, with no further development. The history of technology, by contrast, is *cumulative*: each new invention lays a foundation for further developments, which in turn enable even better devices, and so on. For example, the internal combustion engine went through almost 200 years of refinement before it was ready to be used in an automobile. Also, a device developed for one application can be modified and used for something else. A broadly similar history of cumulative innovation can be seen in the fossil record. Self-organized structures, by contrast, display no such gradual and cumulative improvements; once formed, they are basically static.

The second point is widely recognised, for example by Edis (2007, p.71):

> [Self-organization] sets the stage for evolution, but a fourth step remains. We need a mechanism to create information. Self-organization far from equilibrium can give us impressive and intricate structures, but these are physically constrained. Life requires more flexibility. …

The consensus among biologists is that, while evolution certainly *harnesses* the self-organizing tendencies of matter (such as the tendency of phospholipids to form into a bilayer membrane) the open-ended and creative aspect of evolution cannot come from self-organization. Something very different is required.

2.3.3 Limitations of self-organization

The problems raised for constructionism so far may not appear to be too daunting. The first objection may merely show a limitation of human intuition, and the second is an argument from ignorance that is potentially vulnerable to new data. Is there any general theoretical argument to prove that the laws of physics cannot strongly favour functional structures? There is indeed such an argument, which I will now summarise.[4]

To appreciate the limitations of any local transition rule to create functional structures, imagine for a moment that the state of the system is to be controlled by a person, rather than a LIM law. This person is given the task of making the grid of the cellular automaton evolve toward a target state that is provided to him. In other to do so, he can at each time *t* make the state of each cell at time *t* anything he chooses. This task will be very easy, if he can see the grid all at once, but it becomes difficult if his knowledge of the grid is restricted to local information. An extreme case of such a restriction is where, in making a decision about the state of a particular cell at time *t*, the person is allowed to see only the state of *that cell* at $t - 1$, and is given no information about its location in the grid. Success in such a case will be extremely unlikely unless the target state is highly uniform. If every cell of the target is in state 0, for example, then the person can simply put each cell in the grid into state 0 as well. No position information for the cell being decided upon is then needed. If, on the other hand, the target state has cells in states 0 and 1, in equal numbers, then the person cannot succeed except by luck. Even though he can see the entire target (it may be a checkerboard pattern, perhaps) he has no reliable way to get the 0 and 1 cells on their grid into the same arrangement. He sees a given cell in the grid that is in state 0, say, but has no idea what state it should be in, because *he does not know where it is* in the grid. There is no better strategy here than to simply to flip a coin to determine the new state of the cell.

The above case of a person who is allowed to see only the cells singly, not all together, is exactly analogous to a dynamical law that is even more local in its operation than a LIM law, in that the next state of each cell depends only on its own present state, and is independent of even the surrounding cells. Such extreme locality makes each cell a causally isolated system, unable to interact at all with its neighbours. It is obvious that no particular non-uniform state, and hence no functional structure[5], is likely to emerge under such a law, say from a random initial state.

To create a case that is analogous to a LIM law, suppose that the person has a little more information about each cell in the grid when they have to make a decision about its next state.

[4] These arguments are based on some ideas in Johns (2011). I believe however that his concept of 'salience' is not too helpful, and I use stationary probability instead.
[5] Functional structures cannot be perfectly uniform and homogenous – contrary to some 19th century views of the cell as a sphere of homogeneous protoplasm.

When choosing the state of a given cell at time *t*, they are allowed to see not just its own state at *t*–1, but also the states of the eight neighbouring cells at *t*–1. However, they are still given no information about the position of the cell within the grid, beyond what they can infer from that 3x3 'local block' of states. With this extra information about the grid, they have more control over its evolution, and can now favour some simple structures. If the target is a 'checkerboard' of 0 and 1 states, for example, then their own grid when correct will contain just two types of local block. The person can then adopt a local rule that favours these two kinds of local block, and after a small number of time steps the grid will match the target perfectly.[6]

Even with this extra information, however, the person's control over the grid is extremely limited. Suppose, for example, that the target state is 'irregular' in the sense that each of the 512 possible local block types is present in the target, and with about the same frequency. In that case, an optimal local rule cannot favour any local blocks over the others – it must be neutral at the local level. Such a locally neutral law essentially ignores the surrounding neighbourhood of cells, and so has the same results as the extreme locality case discussed above, where each cell is causally isolated.

In other words, it is impossible for any LIM law to favour a particular irregular state of the grid. Another argument for this conclusion is as follows. A regular state of the grid is self-similar or repetitive – 'the same thing over and over'. If the target state is regular, therefore, there is no need to know the absolute position of the cell whose state is being chosen, in order to choose it correctly. An irregular state, by contrast, is heterogeneous, or non-self-similar, and so different in each part. Matching the grid to an irregular target thus requires some knowledge concerning the absolute location of the cell under consideration. Such information about position could only come from knowledge of the local block, since that is the only information available. Yet, if the initial state of the system is simple and uniform, for example, then the local blocks are the same everywhere and contain no such positional information. The same will be true for a random initial state: even though the local blocks will be different from one another, the state of a block tells you nothing about its position. In fact, the state of a local block will provide positional information only when the grid is already in a *known* irregular state. Getting

[6] I have verified this fact using a simple computer program. Around each cell is both a St. George's cross and a St. Andrew's cross of cells. The algorithm adds up the number of St. George cells that are in state 1, and compares it to the number of St. Andrew cells that are in state 1. If the St. Andrew number is higher, then the central cell becomes 1. If the St. George number is higher, it becomes 0, and if they're equal it has a 50% chance of each.

the grid into such a state, using a local updating rule, is however exactly the original task! Thus we have a chicken-and-egg situation, where producing a specific irregular state requires positional information, which requires producing a specific irregular state, and so on.

A LIM law, which is limited to favouring structures through formation by self-organization, cannot therefore favour any specific irregular structures. We know from experience, however, that functional objects are always irregular and very specific. This is true both of biological structures and human artifacts. A single organism requires thousands of distinct proteins, for example, and individual proteins are also highly irregular sequences of amino acids. Thus, if a LIM law cannot favour a specific irregular structure, then it cannot favour a specific functional structure either. From this it follows that functionality in general cannot be favoured by a LIM law, since functional arrangements are extremely sparse in the set of all possibilities.

2.3.4 Constructionism competes with evolutionary biology

Constructionism claims that the bias toward functionality seen in evolutionary history arises from a basic preference in the laws of physics for functional structures, whereas biologists give the credit for this to natural selection. Moreover, biologists regard natural selection as the *only* source of functional bias in evolutionary history, not merely as one source among others. As Johnson (2010) says, "selection is the only mechanism by which functional relationships evolve". Historically, this role for natural selection was needed to rid evolutionary biology of the apparent need for a rational creator. As Ayala puts it, "It was Darwin's greatest accomplishment to show that the complex organization and functionality of living beings can be explained as the result of a natural process—natural selection—without any need to resort to a Creator or other external agent." Of course, if a functional bias existed in the laws of physics themselves, then this would in itself be a sufficient explanation of "the complex organization and functionality of living beings", so that Darwin's idea of selection would be redundant.

3. Neutralism

The basic idea of neutralism is that the laws of physics are neutral toward the formation of functional objects, so that those laws do not favour functional structures over their non-functional counterparts. According to neutralism, the functional bias in evolution arises not from the laws of physics themselves, but rather from specific biological conditions, such as those appealed to in the modern synthesis. Under abiotic conditions, for example, no such functional bias would exist.

3.1 The Basic Problem with Neutralism

In Section 2.3 we discussed four problems with constructionism. Neutralism on the other hand faces just one difficulty, but it is extremely straightforward and apparently devastating. Kurt Gödel, as reported by Wang (1995), states the problem as follows:

> The formation within geological time of a human body, by the laws of physics (or any other laws of similar nature), starting from a random distribution of elementary particles and the field, is as unlikely as the separation by chance of the atmosphere into its components…The complexity of living bodies has to be present either in the material or in the laws.

Unfortunately Wang doesn't report the details of Gödel's reasoning, but it seems easy to reconstruct. Gödel compares the formation of a human being (or any other living body) by the laws of physics to the spontaneous separation of the atmosphere into its components. Why does Gödel mention gases here? Presumably because gas particles move independently of each other, bar the occasional collision, so that the laws of physics are neutral toward the arrangement of molecules in a gas. If the gas begins in a random initial state, then that state is very likely to be one in which the gases are well mixed, since the separated states are extremely rare. Also, as the state then evolves according to the laws of physics, a separated state is almost impossible to occur, due the rarity of these states together with the neutrality of the laws with respect to the arrangement of the molecules. For the system to produce such a separated state, with more than a miniscule degree of probability, the laws would have to strongly favour such states.

The situation here is analogous to finding a needle in a haystack, where the haystack represents the state space of the system, and the needle represents any rare type of state, such as one in which the atmosphere is separated into its components. If one has a long-range metal detector, or any device that guides one toward the needle, then there is no problem. But if one's movements are neutral with respect to the needle's location, then the sheer size of the haystack in comparison to the needle makes the task very difficult.

Gödel's argument draws the same conclusion about the formation of living bodies, and thus seems to assume two analogous premises about living organisms:

(i) The laws of physics are neutral toward the formation of living organisms, and
(ii) The living-organism arrangements of matter are extremely sparse in the set of all possible arrangements.

I am not sure why Gödel would assume that the laws of physics are neutral toward the formation of living bodies, but he certainly appears to do so. With that premise in place, the two arguments are perfectly analogous in these relevant respects, and the conclusion about living organisms follows in exactly the same way. A formal version of this argument is given in the Appendix.

So if neutralism is true, then the stationary probability of the set of all functional arrangements is exceedingly low, like the probability of the atmosphere separating into its components. In that case, however, we have no good explanation of why living bodies exist, since a good explanation of some phenomenon must, at the very least, be a story according to which the phenomenon has a reasonably high probability. If neutralism is true, then the appearance of life is so improbable as to constitute a miracle.

3.2 What about evolution?

But what about evolution? We know that structures whose formation is extremely improbable, generally speaking, can become probable under the right conditions. Lego blocks, as discussed above, have no inherent tendency to form themselves into robots, but robots will form with high probability when eight-year-old boys are present. In a similar way, one might think, the formation of living bodies by physical laws may be highly improbable generally, but it becomes probable once the conditions for biological evolution exist. After all, the great achievement of Darwin (and Wallace) was to find a physical mechanism – natural selection – that has a functional bias.

This idea is nicely illustrated by Ronald Fisher, whom Julian Huxley (1953) quotes as saying, "Natural selection is a mechanism for generating an exceedingly high degree of improbability." Huxley clarifies this apparent paradox by saying that natural selection produces certain structures with high probability, even though they would be improbable by other means. In a similar way, Richard Dawkins (1996) pictures advanced organisms as sitting atop the peaks of a high mountain, apparently surrounded by unscalable cliffs. But on the other side of the mountain one finds "gently inclined grassy meadows, graded steadily and easily toward the distant uplands". The evolution of life, in Dawkins view, is highly improbable by *almost* all means, but not when Darwin's mechanism has the opportunity to work its magic.

In considering this issue it is important to recognise that, even if neutralism is true, there can be a strong bias toward functionality under the right conditions. Just as a sloping soccer field can have an eastward bias, despite the laws being isotropic, so too a room full of engineers tasked with a problem will have a functional bias, even if the laws are functionally neutral. The presence of a fertilized egg also creates a strong bias toward the formation of a particular functional organism, and if Fisher and Dawkins are right, then the conditions of biological evolution create a strong bias toward function in general. Such pro-functional circumstances may themselves be difficult to form, however, and indeed this must be the case, if the argument in the Appendix is correct, and the absolute probability of functionality is very low. Let the event F be the formation of a functional structure, and event E be some conditions which create a strong functional bias. We then have $P(F) = \delta$, say, where δ is very small, but $P(F \mid E) = q$, where q is reasonably high. The probability calculus then entails that $P(E) \leq \delta/q$. In other words, if an event E renders improbable things probable, then E itself must be improbable. Thus, in a system whose laws are functionally neutral, any conditions that would make the formation of advanced life probable must themselves be absurdly unlikely.

4. Objections

Evolutionary theory is well established in biology, and present versions are believed to be compatible with physicalism. The thesis of this paper, that physicalism cannot account for the origin of biological forms, is therefore likely to be resisted. In this section I will present two possible objections to my thesis and briefly respond to them.

4.1. What about intermediate views?

Functionalism and neutralism are stark alternatives, and a spectrum of intermediates could exist. For example, the laws of physics might have a slight functional bias, though one that is too small by itself to account for the functional bias shown in the fossil record. On this view, the further functional bias needed would come from elsewhere, such as from natural selection.

Such intermediate views cannot be successful, however, since they face the problems (though to a lesser degree) of both constructionism and neutralism. First, the claim that LIM laws have even a slight functional bias is problematic. The same arguments in Section 2.3 apply, especially 2.3.3, so that LIM laws must be completely neutral between states that are locally equivalent. Second, the fact remains that special circumstances (such as the conditions required for evolution by natural selection) cannot increase the probability of advanced life, unless those circumstances are themselves improbable to the same degree.

4.2 The conditions for natural selection arise easily

Let us review the probability argument of Section 3.2, where 'E' (for 'evolution') represents special circumstances that render 'F' (functional structures) probable. It was shown there that, for E to substantially increase the probability of F, E itself must be highly improbable. One may object here that this is simply not the case: the conditions for evolution by natural selection are actually rather modest, especially by comparison with the enormous difficulty of producing advanced life *without* natural selection. The main requirement for natural selection is a population of 'self-replicators', to use Dawkins' term, competing for limited resources, and this does not seem fantastically unlikely to occur -- especially in a universe containing a vast number of suitable planets, and with billions of years of time available. You only have to get lucky once, after all.

I am not qualified to judge the probability of self-replicators arising by physical processes, somewhere in the universe, in the time available. Dawkins and others may be right in thinking that the probability of this is not absurdly small. But if P(E) is not too small, then according to my argument P(F | E) must be very low instead. In other words, *if* self-replicators arise easily, then evolution by natural selection cannot have a very strong functional bias. In other words, my arguments against constructionism and neutralism commit me to denying a conjunction, namely that P(E) *and* P(F | E) are both reasonably high. I prefer not to say which of these two claims I reject, since I am not entirely sure, but if forced I would wager that P(F | E) is much lower than is usually supposed.

Peter Atkins, for example, seems to have great confidence in the functional bias of natural selection, for he states (1981, p. 3):

> A great deal of the universe does not need any explanation. Elephants, for instance. Once molecules have learnt to compete and to create other molecules in their own image, elephants, and things resembling elephants, will in due course be found roaming around the countryside ... Some of the things resembling elephants will be men.

Such confidence in the power of natural selection to create functional beings is not unusual, yet I know of no evidential basis for it. For, while Darwin *proposed* natural selection (together with heritable variation, etc.) to be the main mechanism for evolution, he was not able to demonstrate its causal adequacy. In fact, such a demonstration would be an extremely difficult theoretical exercise, and is still not even remotely possible, except in extremely simple models whose biological application is dubious. There are no observed cases, either in the real world or in computer models, of non-trivial complexity being produced by such processes. Nor is there a mathematical demonstration that this is possible in principle. Instead, selection is presumed to be a creative force, capable of producing life as we find it, on the grounds that (i) this is at least superficially plausible, (ii) attempts to prove that natural selection is *not* adequate have failed, and (iii) we cannot think of another physical process that might accomplish this feat.

The third argument in favour of natural selection, on the basis of being the only physicalist game in town, must be dismissed as circular. For there are no direct *a priori* arguments in favour of physicalism, or materialism, itself. As BonJour (2010) states,

> As far as I can see, materialism is a view that has no very compelling argument in its favor and that is confronted with very powerful objections to which nothing even approaching an adequate response has been offered.

Instead, the evidence for physicalism is empirical, and results from the enormous explanatory success of the physical sciences over the past few centuries. Given this impressive record, it is certainly reasonable to conjecture, in accordance with Ockham's razor, that all phenomena may be ultimately explicable in physical terms. But the use of this argument requires us to take apparent counter-examples to physicalism very seriously – otherwise it becomes the circular argument: "Physicalism is true, because there are no counter-examples. And we know that such-and-such isn't a counter-example to physicalism, because physicalism is true." In the face of an argument that physicalism cannot explain the origin of biological form, it would be circular to argue that natural selection *can* explain this phenomenon on the basis that no other physical mechanism can.

Now let us turn to the first and second arguments. It is true that many scientists and philosophers find selectionism superficially plausible, and that many attempts to prove its inadequacy have failed to gather much support.[7] These facts provide very little evidence for selectionism, however. Concerning the plausibility of selectionism, we know from the history of science that intuitive judgements of plausibility are often mistaken. Moreover, there are also many scientists[8] who find it rather implausible, and this has led to attempts to refute the theory, as noted above. In addition, while arguments against selectionism have not been very persuasive, one cannot infer the truth (or probable truth) of a theory from the fact that it has not yet been

---

7 Such criticisms began with Mivart (1879) and Wallace (1910), but continue to the present day with such authors as Margulis and Sagan (2002), Shapiro (2011) and Nei (2014).

8 For example, those on the list above.

falsified. Some theories remain unfalsified simply because they are unfalsifiable, which is hardly a badge of honour in science.[9]

The weakness of the first and second arguments is also shown by the following analogy, based on similarities between the selectionist mechanism for evolution and the alleged perpetual motion machine known as the Brownian ratchet, or Feynman-Smoluchowski ratchet. This theoretical device has a shaft that is free to rotate, with a paddle wheel at one end that is buffeted randomly by the air molecules striking the paddles. This random buffeting is not biased in either direction, and so by itself is very unlikely to cause any significant net rotation of the shaft in the long run. But at the other end of the shaft is a ratchet and pawl mechanism that allows it to turn in (say) the clockwise direction only. Every now and again, the air molecule impacts will produce a small clockwise motion that is enough to slide the pawl up the gently-sloping side of a tooth, and drop it into the valley beyond. This gain is then fixed and permanent, since counter-clockwise motions are prevented by the steep side of the pawl teeth, and the pawl will remain there until (by chance) another small clockwise turn occurs. Over long periods of time, the shaft will turn very slowly clockwise (so the argument goes) and can be used to do work, such as winding up a weak spring.

The Brownian ratchet would be a perpetual motion machine in the case where the whole device is at the same temperature – hence no present scientist believes that the device could work at thermal equilibrium, since perpetual motion would contradict the second law of thermodynamics. However, I suggest that if the second law were unknown, then scientists might well find the Brownian ratchet very plausible. This would be true especially for a scientific community that strongly believed (for whatever reasons) that perpetual motion must be possible *somehow*. The argument would then be made, I believe, that something like the Brownian ratchet *must* be the cause of perpetual motion, since no other mechanism seems conceivable.

No doubt a few outspoken scientists would criticise the Brownian ratchet, attempting to show that it could not work as advertised. They might, for example, think of Feynman's (1962) argument that since the pawl is at the same temperature as the paddles, it will suffer thermal fluctuations of its own that will occasionally allow the shaft to slip backwards, and that these will

[9] Indeed, as Caplan (1978: 261) reports, selectionism itself has been criticised on the grounds that it "seems to possess a disquieting amount of elasticity or flexibility with regard to explaining organic phenomena. Anything and everything in the empirical biological world seems to be compatible with evolutionary explanations".

exactly balance its forward motion in the long run. This kind of analysis is easy to contest, however, and indeed Abbott et al (2000: 214) state, for example, "It is now well known that Feynman's analysis was flawed". Another difficulty is that all such specific arguments can be circumvented by modifying the device in some way. For example, the use of multiple pawls, whose thermal motions are uncorrelated, would seem to almost eliminate the instances of reverse motion. Now of course the viability of such modified devices can also be criticised, yet it is not possible to criticise all devices simultaneously except using a general principle such as the second law. Therefore, absent an intuition against the device, the hypothetical scientists who are unware of the second law would have little reason even to take such criticisms seriously. When Abbott et al (2000: 215) discuss the "many pawl paradox", for example, they begins by stating, "It would appear *prima facie* that … the wheel will rotate in one direction. This cannot be correct as it would then be possible to construct a machine that would disobey the Second Law." Actual criticisms of the Brownian ratchet are thus driven by prior knowledge that it cannot work.

The counterfactual scientists would believe that their Brownian-ratchet theory of perpetual motion must be correct, on the grounds that it is the only plausible way for perpetual motion to occur, and that criticisms of it are not clearly successful. But they would of course be badly mistaken, and so we should be extremely wary of arguing for selectionism on similar grounds.

5. Conclusion

In this paper I have investigated the possibility of explaining the origin of biological form within a physicalist framework, according to which all the ultimate constituents of the world are physical, and follow physical laws (either deterministic or probabilistic). The difficulty for physicalism is to account for the strong bias toward functionality that is displayed in evolutionary history. Physicalism allows only two possible theories that concerning this functional bias, constructionism and neutralism, which are here defined precisely in terms of the stationary probability distribution of an irreducible Markov process. Both constructionism and neutralism face very serious objections, however, and I conclude that physicalism cannot solve the problem of biological form.

References

Abbott, D., Davis, B. R., Parrondo, J. M. R. (2000) "The Problem of Detailed Balance for the Feynman-Smoluchowski Engine (FSE) and the Multiple Pawl Paradox" in *Unsolved Problems of Noise and Fluctuations*, (2000) ed. D. Abbott and L. B. Kish, American Institute of Physics.

Atkins, P. (1981) *The Creation*, Oxford: Freeman and Co.

BonJour, Laurence (2010), "Against Materialism", in Bealer and Koons (eds.) *The Waning of Materialism*, 2010.

Camazine, S. et. al. (2001), *Self-Organization in Biological Systems*, Princeton University Press.

Caplan, A. (1978) "Testability, disreputability, and the structure of the modern synthetic theory of evolution", *Erkenntnis*, vol. 13, 261-78.

Churchland, P. M. (1988), *Matter and Consciousness*, MIT Press.

Dawkins, R. (1986), *The Blind Watchmaker*, New York: W. W. Norton & Company, Inc.

--------- (1996) *Climbing Mount Improbable*, New York: W. W. Norton & Company, Inc.

Descartes, R. (1644), *Principles of Philosophy*, trans. Jonathan Bennett, available at www.earlymoderntexts.com, accessed May 14, 2012.

Goldenfeld, N. and Kadanoff, L. (1999) "Simple Lessons from Complexity", *Science*, 284: 87-89.

Goldenfeld, N. and Woese, C. (2011), "Life is Physics: Evolution as a Collective Phenomenon Far From Equilibrium," *Annu. Rev. Condens. Matter Phys.* 2:375–99

Gould, S. J. (1977) *Ever Since Darwin: Reflections in Natural History*, New York: Norton.

Gould, S. J. (1990), *Wonderful Life: The Burgess Shale and the Nature of History*, New York: Norton.

Johns, R. (2011), "Self-organisation in dynamical systems: a limiting result", *Synthese*, 181: 255- 75.

Margulis, L. and Sagan, D. (2002) *Acquiring Genomes: A Theory of the Origins of Species*, Basic Books.

Mivart, S. G. J. (1871) *On the genesis of species*, London.

Nagel, T. (2012) *Mind and Cosmos*, OUP.

Polanyi, M. (1968) "Life's Irreducible Structure", *Science* 160, 1308-1312.

Shannon, C. E. (1948) "A Mathematical Theory of Communication", *The Bell System Technical Journal*, 27, 379–423, 623–656.

Shapiro, J. (2011) *Evolution: A View from the 21st Century*, New Jersey: FT Press Science.

Wallace, A. R. (1910) *The World of Life*, London: Chapman and Hall.

Wang, Hao (1995). 'On "computabilism" and physicalism: Some problems', in J. Cornwell, ed., *Nature's Imagination*, 1995.)

White, R. (2007) "Does Origins of Life Research Rest on a Mistake?", *Noûs* 41:3, 453-477.

Wolfram, S. (2002) *A New Kind of Science*, Wolfram Media.