

Dynamical Complexity and Regularity

(Some ideas for discussion)

Richard Johns
University of British Columbia
johns@interchange.ubc.ca
June 11, 2001

ABSTRACT

The aim of this paper is to provide a mathematical basis for the plausible idea that regular dynamical laws can only produce (quickly and reliably) regular structures. Thus the actual laws, which are regular, can only produce regular objects, like crystals, and not irregular ones, like living organisms.

1. The dynamical complexity of an object is defined. This is something like its algorithmic information content, but I use a dynamical system in place of a universal Turing machine. A dynamically-complex object is, roughly speaking, one that the dynamical laws have little or no tendency to produce from a random initial state. A “GIGO” theorem is proved, that an object with dynamical complexity n bits requires time 2^n to be generated, so that highly-complex objects effectively cannot be generated spontaneously.

2. The term “regular dynamical law” is defined, in such way that the known dynamical laws of physics are regular. The irregularity of an object is defined, and then I *try* to show that the dynamical complexity of an object s , with respect to a regular law, always exceeds the irregularity of s . It seems clear that living organisms are highly irregular, in the sense defined, so (if this result holds) they must be dynamically complex as well. It would then follow that living organisms could not have been produced, from a random initial state, by the known dynamical laws.

1. Overview

1.1 Dynamical Complexity

It is generally recognised that living organisms are highly complex. Since the work of von Neumann (1966) on self-reproducing automata, some attempts have been made to understand what biological complexity is, and how (or if) it can be produced naturally.

The main difficulty with this project, at present, is that we do not have a general definition of complexity that can be applied to biological systems. While there are a number of such definitions, they are all rather arbitrary. Any definition of complexity requires some sort of “reference frame”, relative to which complexities are evaluated – complexity cannot be defined in a vacuum, so to speak. The choice of reference frame seems arbitrary, so that no complexity measure is uniquely correct.

The two main kinds of reference frame are (i) a language, and (ii) a universal Turing machine. Once a language L has been chosen, for example, then the (linguistic) complexity of an object s relative to L can be defined as the length of the shortest complete description of s in L . Or, if a universal Turing machine M has been chosen, then the (algorithmic) complexity of s is the length of the shortest program that, given to M , produces the output s .

The rough idea of these definitions is that the complexity of an object is the amount of information needed to specify it. This is the right idea, I believe, but the question is: Which language, or Turing machine, do we choose? Each language, or Turing machine, gives rise to a different complexity measure. The answer seems to be that there is no single complexity measure that is useful in every context; rather, one chooses the reference frame according to the problem at hand.

In evolutionary theory, the problem is how the world managed to produce complex living organisms. Thus, for this context, the best “reference frame” is surely the dynamical laws of the real world. The real world is not a Turing machine, so that any choice of Turing machine will be arbitrary in this context. Also, the physical world does not provide us with any language. The idea, then, is to use the dynamical laws of physics in something like the same way that Solomonov, Kolmogorov and Chaitin (SKC) used a universal Turing machine.

This notion of complexity I shall call *dynamical complexity*. It should depend only on the object, and the dynamical laws of the system. The rough idea, as always, is that the complexity of an object is the amount of information needed to specify the object, within the reference frame. As with the SKC definition, the object's being "specified" has something to do with its being produced by the system. Dynamically complex objects are, roughly speaking, ones that have little or no tendency to be produced by the system.

1.2. Complexity and Irregularity

There is an intuitive notion of complexity that is not exactly captured by formal definitions.¹ We think of a complex object as heterogeneous, aperiodic, elaborate, or patternless. Complexity is opposed to simplicity. A simple object is something like a crystal, which has a small number of basic parts, arranged in a way that is easily specified. The more symmetry, self-similarity, or repetition an object contains, the simpler it is. A complex object, on the other hand, has little or no pattern to it. Let us call this intuitive kind of complexity *irregularity*. Note that an irregular object must be composed of many parts; no object with few parts is elaborate, no matter how the parts are arranged. Also, an irregular object need not do anything useful or interesting—its parts may even be arranged randomly.

There is a link between irregularity and the formal definitions of complexity, but they are not exactly equivalent. Consider linguistic complexity, for example. The sequence $\langle 0,0,0,0,\dots,0 \rangle$ (all zeros) strikes us as simple, as it is highly regular. It is invariant under many different transformations, including all translations. It is also easily specified in English, e.g. "All zeros". So linguistic complexity matches irregularity in this case. On the other hand, consider a sequence that strikes us as complex, such as the binary form of *War and Peace*. It may seem that this is linguistically complex, as it is a very long string of English words, but what if the language includes a name for the string, such as "*War and Peace*"? In that case, the object is specified with only 13 letters, and so is linguistically simple. Thus, for linguistic complexity to coincide with irregularity, it seems that we require a "simple" language, containing no short names for irregular objects.

¹ See McShea (1991).

The same situation exists with algorithmic complexity. A long, irregular string may be built into a Turing machine, so that it is produced by a very short program, one that effectively says “print out the stored string”. In that case, the irregular string has low algorithmic complexity for that machine. Again, we seem to require the Turing machine to be “simple”, in order for algorithmic complexity to approximate irregularity.

In a similar way, the dynamical complexity of an object will depend upon the set of dynamical laws in question, so that no rigid link exists between dynamical complexity *in general* and irregularity. Fortunately, however, we need only look at the actual laws of physics, so that there is no need to consider all possible sets of laws. The question then arises: “What is the relation between dynamical complexity, *for the actual dynamical laws*, and irregularity?”

This question is more easily answered than one might think, since the actual laws of physics have certain general properties, which are (I think) sufficient to show that complexity and irregularity are linked in the following way:

Regularity Principle Dynamical complexity \geq Irregularity

Thus any object that is irregular is also dynamically complex, but the converse does not hold. A dynamically-complex object may actually be quite regular, if the regularity is of the “wrong kind”, i.e. the pattern is one that is not easily produced by the laws. For example, a cubic crystal of carbon atoms is a very regular object, but the laws of physics do not allow its formation, so that it is dynamically complex. The only real carbon crystals, i.e. graphite and diamond, have quite different structures.

What are the general properties of the laws of physics that lead to the regularity principle? They are (i) causal locality, and (ii) symmetries (of various kinds). Causal locality means that causal processes form a continuous world line of time-like separated points. Roughly speaking, every event is directly caused only by neighbouring events, in its immediate past, so that there is no “action at a distance”. The symmetries include invariance under spacetime translation and spatial rotation, among others.

These properties of locality and symmetry effectively mean that the laws cannot directly control the *global* structure of the state of the system. They can directly control only the local structure, which is not always sufficient to determine the global structure. An irregular object

has a global structure that is largely independent of its local structure, so that a law with these properties cannot (reliably) produce any *particular* irregular object.

The concept of irregularity is relevant to biology, since living organisms are generally held to be highly irregular. From the regularity principle, it would then follow that they are dynamically complex.

1.3 The GIGO Theorem

Dynamical complexity is related to chance and time via a fairly trivial theorem, which I call the GIGO (Garbage In, Garbage Out) theorem. This theorem basically says that, if an object has dynamical complexity n bits for a given set of laws, then those laws require a time of order 2^n to produce the object, with a reasonable chance, from a random initial state. In a short time, the chance of producing the object is only about 2^{-n} .

Thus, if the dynamical complexity of an animal is one million bits, or greater, and the time available is only a few billion years, then its production from a random initial state is effectively impossible. The time required is greater than this by many thousands of orders of magnitude.

2. Sketchy Details

2.1 Dynamical Complexity Defined

The rough idea of dynamical complexity is that $Comp(s)$ is the number of bits required to specify the object s , relative to the dynamical laws. This will be defined in terms of the probability that s appears in the history of the system, and how quickly it is likely to appear. The first question is: What do we choose as the initial state of the system? The choice of initial state will greatly affect the history of the system, and thus the objects that are produced. Now, since we wish the complexity of an object to depend only on the dynamics of the system, we shall set the initial state at random, i.e. have a uniform probability distribution over the possible states. The idea

here is that a random initial state is “informationless”, i.e. by setting the initial state at random we give the system no information.

The random initial condition, and the dynamical laws (which may be either deterministic or stochastic), together provide a probability function on the set of possible histories of the system. For simplicity, we shall first just look at the *objects* that exist in each history, supposing that it is clear what should count as an object. These objects, we suppose, are ordered by the times at which they appear in the history. Thus, each history is regarded as a well-ordered sequence of objects.

A complexity is usually (minus) the logarithm of a probability, so the obvious idea is to define the dynamical complexity of \mathbf{s} , relative to a set of laws, as (minus) the log of the probability that those laws produce \mathbf{s} from a random initial state. This approach, unfortunately, may give every object low complexity. The reason for this is easy to see: many systems, even very simple ones, will eventually produce every possible object (for which sufficient material exists). For example, consider a binary counter, with (say) 100 bits.² The dynamical law for this system is very simple, but it eventually produces every 100-bit sequence, regardless of its initial state. Thus the probability of any sequence \mathbf{s} appearing in the history is 1, which makes the dynamical complexity of \mathbf{s} equal to zero.

To overcome this problem, I have developed the *salience* function, which is not exactly a probability function.³ The idea is that objects that tend to be among the first produced by the system are more prominent, or salient, than those that tend to appear much later, if at all. To quantify salience, I consider a method for generating a *single* object, using a dynamical system, which is as follows. We assign the system a random initial state, and then let the system produce r objects, according to its own dynamical laws. One of these r objects is then selected at random.

We can define the *r-salience* of an object \mathbf{s} as the probability of generating the object \mathbf{s} by this method. More precisely:

Definition Let the proposition $F^r \mathbf{s}$ say that \mathbf{s} is among the first r objects in the history.

Then $Sal^r(\mathbf{s}) = P(F^r \mathbf{s})/r$.

² This system’s state is characterised by a 100-bit binary number. At each time step, this number increases by one.

³ The salience of \mathbf{s} is in fact an *upper* probability. See Howson and Urbach (1989).

This concept of r -salience seems promising. If the object \mathbf{s} tends to be produced fairly quickly, from most initial states, then its r -salience will be quite high for some (low) values of r . If \mathbf{s} is rarely produced, on the other hand, for almost all initial states, then its r -salience will be low for all r . This fact suggests that we define salience from r -salience in the following way.

Definition $Sal(\mathbf{s}) = \max_r \{Sal'_r(\mathbf{s})\}$.

In other words, we define the *salience* of \mathbf{s} as its *maximum* r -salience, over all values of r . Thus, if a state tends to be produced quickly by the dynamics, so that its r -salience is quite high, for some small r , then its salience will also be quite high. A state that is unlikely to be produced in any short time will have low salience.

The salience of \mathbf{s} is then something like the best probability of finding \mathbf{s} using the dynamics of the system in question. So we can define the complexity of the state as minus the log of its salience.

Definition $Comp(\mathbf{s}) = -\log Sal(\mathbf{s})$.

With this definition, each n -bit state of a binary counter has n bits of dynamical complexity, as each has r -salience 2^{-n} for every r .

2.2 The GIGO theorem

Before we can state and prove the GIGO theorem, we must first define the concept of a *program*. If the object \mathbf{s} exists in some possible state of a dynamical system, then it is possible to “program” the system to produce \mathbf{s} , simply by giving the system a suitable initial state.⁴ Thus manipulation of the initial condition can greatly reduce (or increase) the time required to produce a particular object. To investigate how much manipulation is need to bring about a significant reduction in the required time, we define a *program* as follows.

⁴ One may put the system in a state where \mathbf{s} already exists, or in a state which quickly evolves to one where \mathbf{s} exists.

Definition A *program* Π is a restriction of the initial state to a subset of the state space. (The subset is also called Π .)

Since the initial state is set at random, each program Π has a probability $P(\Pi)$, which in the case of a finite state space is simply the proportion of states that are in Π . The more restrictive the program, the lower its probability is. We also define the *length* of a program as follows:

Definition The *length* of Π , also written $|\Pi|$, is $-\log P(\Pi)$.

We now have a useful lemma.

Basic Lemma $Comp(\mathbf{s}) = \min_{r, \Pi} \{\log r + |\Pi| - \log P(F^r \mathbf{s} | \Pi)\}$.

Proof: Let the O_i^r be all the possible output sets of length r . Then

$$P(F^r \mathbf{s}) = \sum_i P(F^r \mathbf{s} | O_i^r) P(O_i^r).$$

Now $P(F^r \mathbf{s} | O_i^r) = 1$ if $\mathbf{s} \in O_i^r$, and is 0 otherwise. Thus:

$$P(F^r \mathbf{s}) = \sum_{\mathbf{s} \in O_i^r} P(O_i^r); \text{ Also } P(F^r \mathbf{s} | \Pi) = \sum_{\mathbf{s} \in O_i^r} P(O_i^r | \Pi).$$

Further, $P(O_i^r) = P(O_i^r | \Pi)P(\Pi) + P(O_i^r | \neg\Pi)P(\neg\Pi)$, so $P(O_i^r) \geq P(O_i^r | \Pi)P(\Pi)$. But, if Π is the entire state space, then $P(O_i^r) = P(O_i^r | \Pi)P(\Pi)$. Hence $P(O_i^r) = \max_{\Pi} \{P(O_i^r | \Pi)P(\Pi)\}$.

Substituting this in the previous equation gives:

$$P(F^r \mathbf{s}) = \sum_{\mathbf{s} \in O_i^r} \max_{\Pi} \{P(O_i^r | \Pi)P(\Pi)\}$$

$$\begin{aligned} \text{Thus } Sal(\mathbf{s}) &= \max_r \left\{ \frac{1}{r} \sum_{\mathbf{s} \in O_i^r} \max_{\Pi} \left\{ P(O_i^r | \Pi) P(\Pi) \right\} \right\} \\ &= \max_r \left\{ \max_{\Pi} \left\{ \frac{P(\Pi)}{r} \sum_{\mathbf{s} \in O_i^r} P(O_i^r | \Pi) \right\} \right\}. \end{aligned}$$

$$\text{But } \sum_{\mathbf{s} \in O_i^r} P(O_i^r | \Pi) = P(F^r \mathbf{s} | \Pi),$$

$$\begin{aligned} \text{So } Sal(\mathbf{s}) &= \max_r \left\{ \max_{\Pi} \left\{ \frac{P(\Pi)}{r} P(F^r \mathbf{s} | \Pi) \right\} \right\} \\ &= \max_{r, \Pi} \left\{ \frac{P(\Pi) P(F^r \mathbf{s} | \Pi)}{r} \right\}. \end{aligned}$$

$$\begin{aligned} \text{Then } Comp(\mathbf{s}) &= -\log \left[\max_{r, \Pi} \left\{ \frac{P(\Pi) P(F^r \mathbf{s} | \Pi)}{r} \right\} \right] \\ &= \min_{r, \Pi} \left\{ -\log \left(\frac{P(\Pi) P(F^r \mathbf{s} | \Pi)}{r} \right) \right\} \\ &= \min_{r, \Pi} \left\{ \log r - \log P(\Pi) - \log P(F^r \mathbf{s} | \Pi) \right\}. \end{aligned}$$

Then, putting $|\Pi| = -\log P(\Pi)$, we get:

$$Comp(\mathbf{s}) = \min_{r, \Pi} \{ \log r + |\Pi| - \log P(F^r \mathbf{s} | \Pi) \}. \blacksquare$$

This lemma immediately yields the GIGO theorem.

GIGO Theorem Suppose $Comp(\mathbf{s}) = n$, and $|\Pi^*| < n$. Then $P(F^{r^*} \mathbf{s} | \Pi^*) \leq r^{* \cdot 2^{|\Pi^*| - n}}$.

Proof: From the basic lemma, $n = \min_{r, \Pi} \{ \log r + |\Pi| - \log P(F^r \mathbf{s} | \Pi) \}$. Then consider some particular program Π^* and some value r^* . It is then clear that:

$$n \leq \log r^* + |\Pi^*| - \log P(Fr^*s | \Pi^*),$$

And therefore $P(Fr^*s | \Pi^*) \leq r^* \cdot 2^{|\Pi^*| - n}$. ■

The GIGO theorem is not too surprising, in view of the way dynamical complexity is defined. A complex object is defined as one that is hard for the system to produce, so it is not surprising that complex objects are hard to produce! There are some features that are worthy of note, however.

First, it should be noted that time is measured not in seconds, but as the number of objects produced. Thus the size of the system is relevant, as a large system can produce many objects at once, and will thus produce objects at a fast rate. I am not sure how quickly the physical universe produces objects, as I'm not sure quite what to count as an "object". This is something that has to be sorted out. But since the universe contains only about 10^{80} basic particles (physicists tell us) and is only about 10^{18} seconds old, it seems that the number of objects produced by now cannot be too much greater than 10^{100} .

Second, a slight restriction on the initial state of the system only has a slight effect on the probability of producing the object, in a given time. Each bit of the program can only double (at most) the probability of getting the object. This result essentially shows the stability of defining dynamical complexity using a random initial state. An approximately-random initial state would give roughly the same values for dynamical complexities.

Third, the relation in the GIGO theorem between complexity and time is not completely obvious. The theorem shows that the relation is exponential, so that each added bit of complexity doubles the time required for a good chance of producing the object. In other words, there can be no special mechanisms that are able to produce dynamical complexity quickly, in linear or even polynomial time. If a mechanism existed, within the system, to produce s quickly and reliably from a random initial state, then s would *ipso facto* be dynamically simple.

Fourth, the exponential relation between complexity and time is exactly what one has in the case of random guessing. If one tries to guess a sequence of n bits, then the probability of success is 2^{-n} each time, and the number of guesses needed for a good chance of success is of the order 2^n . Thus no physical mechanisms can exist to produce a dynamically-complex object

quicker than pure chance would. This seems directly to contradict claims of Richard Dawkins (1986), and perhaps of neo-Darwinists in general. Dawkins acknowledges that organisms are highly complex (in some sense), having complexities of at least millions of bits. Yet he claims that organisms are generated, quickly and reliably, by the physical mechanisms of mutation and natural selection. Is Dawkins' claim refuted by the GIGO theorem?

It is not that simple. Dawkins may say that *his* notion of complexity differs significantly from dynamical complexity, as I define it. He may claim, in other words, that living organisms are dynamically simple—indeed, neo-Darwinism is surely committed to this. Neo-Darwinists may point out that some objects, like the decimal expansion of pi, appear to be complex but are not (in the sense of algorithmic complexity). The digits have no pattern or regularity that is easily perceived, but the sequence is generated by a fairly simple algorithm. Perhaps living organisms are like pi? They appear to be complex, but are actually generated by simple mechanisms like mutation and natural selection. They are highly irregular, but dynamically simple.

In the next section we shall look at the regularity principle, which entails that all irregular objects are dynamically complex, relative to the actual laws of physics. If this principle is true, then it seems that living organisms are all highly complex, with the dynamical complexity of a vertebrate probably exceeding one million bits. The conjunction of the GIGO theorem with the regularity principle thus would create a serious problem for Darwinism.

2.3 The Regularity Principle

The regularity principle depends on two basic claims: (i) The laws of physics are “regular” in some sense, and (ii) Only regular objects can have high salience with respect to regular laws, so that regularity begets regularity. We shall look at these in turn.

As stated in Section 1.2, the dynamical laws of physics are “regular” by virtue of two properties, namely causal locality and (various kinds of) symmetry. Causal locality means that the behaviour of the world at a spacetime point is directly caused only by *neighbouring* events in the immediate past. For one event A to exert influence over a distant event B, there must be a “causal process” of events connecting A to B, where each event in the process causes its

immediate successor. Moreover, causal processes cannot travel faster than the speed of light c , so that causes and effects must be timelike separated.

Causal locality has the consequence that *information* cannot be transmitted faster than light. This means that, when the laws determine the state of the world at some point (or the probabilities over the possible states) they only have information about the local situation. The “decision”, if you will, has to be based solely on the local situation, and is independent of everything else.

The laws of physics are thought to possess many kinds of symmetry. For example, they are believed to be the same in all parts of space, and not to vary with time. Space is said to be isotropic, in the sense that the laws are the same in every spatial direction. These symmetries are usually expressed by saying that the laws are invariant under certain transformations. Thus the laws are supposed to be invariant under (static) spatial translation and rotation. They are also invariant under time translation, i.e. a switch from t to $t'=t+\delta$, for any δ . There are many other symmetries, such as Lorentz invariance and time-reversal invariance, but we need only consider the more basic ones of spatial translation and rotation, and time translation. We therefore have to following definition.

Definition A *regular* law is causally local, and invariant under spatial translation and rotation, and time translation.

An example of a regular law may be useful. For a very simple model, consider Conway’s well-known game of Life. This world consists of a rectangular grid, each cell of which is occupied by a ‘0’ or a ‘1’. The initial state of the world is filled in by hand, or at random, and thereafter it evolves, in a series of discrete steps, according to its dynamical law. The law is deterministic, and applies separately to each individual cell. The state of every cell at time $t+1$ depends on its state at t , and the states of its (eight) immediate neighbours at t , in the way stated by the following table.

Previous State	Number of Surrounding '1's	New State
0	3	1
0	0,1,2,4,5,6,7 or 8	0
1	2 or 3	1
1	0,1,4,5,6,7,8	0

Thus a '0', surrounded by three '1's will change to a '1', a '0' surrounded by five '1's will remain a '0', and so on.

The causal locality of this law consists of two facts, that (i) the state of each cell at $t+1$ depends (directly) only on states at t , the immediate past, and (ii) the state of each cell depends (directly) only on the states of its immediate neighbours.

The law is invariant under spatial translation, since the same law is applied to each cell in the grid. It is also invariant under rotation, since the law only appeals to the *number* of '1's in neighbouring cells, and not to their positions. The law does not distinguish between the four basic directions.

The question now is: What kinds of object are produced by (have high salience relative to) such regular laws? The short answer to this question is: "regular objects", where object regularity is defined in terms of regular laws.

Regular laws have a certain kind of "blindness", in that they cannot distinguish between some pairs of states. For example, since regular laws are invariant under spatial translation, they cannot distinguish between a state and any of its spatial translations. More importantly, a regular law is blind to the *global structure* of a state. It sees only the little pieces of local structure, and is unable to step back and see the big picture, so to speak. The basic strategy for defining the regularity (or rather, irregularity) of a state is then as follows:

(i) We partition the state space into equivalence classes, where members of a single class are indistinguishable by any regular law.

(ii) The irregularity of a state is then defined as the number of bits of information required to specify the state, within its class. (This is just the base-2 logarithm of the size of the class.)

It should be noted that irregularity thus defined is quite different from dynamical complexity, although we shall later show that the two are related. The irregularity of a state does not depend on what the dynamical law is, whereas the dynamical complexity of a state depends very strongly on what the law is. To examine the relation between irregularity and dynamical complexity, let us look at a very simple example—cellular automata in one dimension.

Each state of the cellular automaton is a sequence of n bits, so that the world is a one-dimensional spatial array. (The model can easily be generalized to three spatial dimensions, however.) One may think of the world as having n spatial cells, in a circle⁵, each of which is occupied by either a 0 or a 1. (One may easily increase the number of states per cell, also.) The world is discrete in time as well as space, so that there are states for times 0, 1, 2 and so on. The state of the world at $t = 0$ is set at random, and thereafter the state evolves according to the dynamical law. The state at $t = 1$ is derived directly from the state at $t = 0$, the state at $t = 2$ from the one at $t = 1$, and so on, indefinitely.

The causal locality constraint says that the state at cell (t, x) is directly derived from the states at $(t-1, x-1)$, $(t-1, x)$, and $(t-1, x+1)$, *and no others*.⁶ In other words, the state of each spacetime cell is caused only by the three cells closest to it, in its immediate past. The constraint of translation invariance says that the same law is used to determine the probabilities for the contents of each spacetime cell. The law need not be deterministic, in that it might only provide probabilities for the possible states of each cell.

Any dynamical law L that satisfies these two constraints may be represented by eight probabilities, as explained below. Consider some spacetime cell, (t, x) . There are a total of eight possible states for the three earlier cells that cause the contents of this one, namely 000, 001, 010, 011, 100, 101, 110 and 111. We shall call each of these eight states a *triple*, and name them 0, 1, 2, ..., 7, i.e. in octal (base 8) form. For each triple type i , there is a probability q_i that it causes a '1' to appear in each cell for which it is responsible, and a probability $(1-q_i)$ that it produces a

⁵It makes everything much easier if the world “wraps around”, so that there are no edges to space, and thus every point can be treated alike.

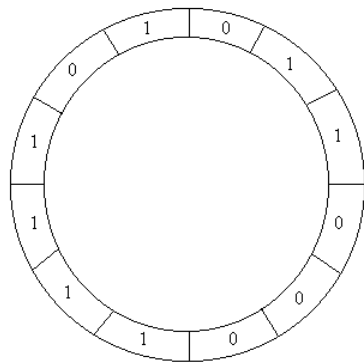
⁶Since space wraps around, the state at $(t, 0)$ depends on the states at $(t-1, n)$, $(t-1, 0)$ and $(t-1, 1)$.

'0'. Now, in view of translation variance, these probabilities q_i cannot vary across spacetime, so that only one set $\{q_0, q_1, \dots, q_7\}$ is needed to specify the dynamical law exactly. If the law should happen to be deterministic, then each of the q_i will be either zero or one. In many situations it is more convenient to think in terms of the probability of a "toggle", i.e. a change from '0' to '1', or from '1' to '0', rather than the probability of a '1'. These "toggle" probabilities will be written $\{p_0, p_1, \dots, p_7\}$. Each p_i gives the probability that the middle bit of a triple of type i will toggle; i.e. $p_i = q_i$ for $i = 0, 1, 4, 5$, and $p_i = 1 - q_i$ for $i = 2, 3, 6, 7$.

A regular law, in a one-dimensional cellular automaton, is actually specified by only six probabilities rather than eight, owing to the rotational invariance of the law. This means that triples 1 and 4 (001 and 100) and triples 3 and 6 (011 and 110) are indistinguishable, so that $p_1 = p_4$ and $p_3 = p_6$.

Now let us see which pairs of states are indistinguishable by a regular law. It is obvious that each state is indistinguishable from all others obtained from it by translation and "rotation". (A "rotation", in one dimension, simply reverses the order of the bits.) Such pairs of states differ only trivially, however, as they contain the same bits in essentially the same order. More interesting cases of indistinguishable states are ones where the state is "shuffled", so that the ordering of the bits differs between the two states. This shuffling must be carried out on the triples, rather than the bits, as the following example explains.

Let the world consist of only 12 spatial cells, so that one possible state is 011000111101. This can be visualised as:



This state can be broken up into “triples”, i.e. sequences of three consecutive bits. The triples are not disjoint but overlap, so that each cell is the centre of one triple. In this way, starting at the top and going clockwise, we get the triples 101, 011, 110, 100, 000, 001, 011, 111, 111, 110, 101, and 010. To write this down more briefly, we can write each triple in octal (base 8) form, which yields 536401377652, as separating commas are no longer needed. Overall we see that this state is composed of the following triples:

Type of triple	0	1	2	3	4	5	6	7
Number of triples of that type	1	1	1	2	1	2	2	2

This table gives what we might call the “triple composition” of the state, and also specifies the *local structure*. Each triple gives that part of the state that is local to a particular cell, and so the table gives all the little bits of local structure. The table does not specify how the triples are arranged in the state, however—this information is called the *global structure*. In general, we have the following definitions.

- Definitions**
- (i) The *local structure at x* of a state is that part of the state that exists in the neighbourhood of the point x .
 - (ii) The local structure of a whole state is the unordered collection of local structures at all spatial points.
 - (iii) The *global structure* of a state is the arrangement of the local structures.

We might think of the local structure as that knowledge of the state which can be obtained by a short-sighted person, who can only get one triple into focus at a time, and is unable to step back and look at the whole state.

One might wonder, in this example, whether there are any other states that share the same local structure (triple composition) as this one. There are indeed such states, such as

001011011110, 001011101110, 001011110110, and 001110101110, which in triple form are 012536537764, 012537653764, 012537765364 and 013765253764 respectively. These can be called *triple shuffles* of the original state, as they are not mere translations. In this example, therefore, the local structure does not determine the global structure, so that the state is somewhat irregular.

For other states, the local structure determines the global structure entirely. The state 000000000000, for example, is the only state that consists of twelve triples of type 0. For other states, the local structure determines the global structure up to translation. Consider the state 101010101010, for example, which is 252525252525 in triple form. The only other state with this triple composition is 010101010101, which is just a translation of the original state. Thus the local structure of this state virtually determines its global structure, which means that the state is highly regular. There is but one ordering of these triples, so that the only choice is whether to begin with a '1' or a '0'. In this respect the state is like a crystal, the structure of whose unit cell determines the whole structure.

I have suggested that a regular law cannot distinguish between states with the same triple composition. More precisely, my claim is GBRL, as stated below. We first have to define a few terms, however.

Definition A *triple shuffle* is any transformation that maps one state to another that has the same triple composition.

Definition Two states are *locally similar* iff one can be transformed to the other by triple shuffles and rotations.

I claim that a regular law cannot “distinguish” between locally-similar states. To distinguish them means to assign them different probabilities, at some time.

Definition Let P_t be the probability function for the state of the system at time t , where the system evolves according to law L from a random initial state. Then L *distinguishes* between \mathbf{s} and \mathbf{s}' just in case, for some time t , $P_t(\mathbf{s}) \neq P_t(\mathbf{s}')$.

Here then is my central claim, called GBRL for *Global Blindness of Regular Laws*.

GBRL If \mathbf{s} and \mathbf{s}' are locally similar, then they are indistinguishable by any regular law.

We can use GBRL (if it is true) to prove the regularity principle, that the dynamical complexity of an object is greater than its irregularity. First, however, we have to define the irregularity of an object. To do this, we first define the *global freedom* of a state, as follows.

Definition The global freedom $G(\mathbf{s})$ of a state \mathbf{s} is the number of states (including itself) that are locally similar to it.

Definition The *irregularity* of a state \mathbf{s} , $Irreg(\mathbf{s})$, is $\log G(\mathbf{s})$.

It is easy to see that some regular states have high dynamical complexity for some laws. For example, consider the state $\langle 0,0,0,0,\dots,0 \rangle$, which is as regular as they come. (Suppose that the world still has 12 cells.) Then the law which puts $p_i = 1/2$, for all i , gives this state probability 2^{-12} for appearing at each time, so that the salience of the state is 2^{-12} . This means that the dynamical complexity of the state is 12 bits, i.e. the same as every other state under this law. Note that another law, namely $q_i = 0$ for all i , gives this state a salience of $1/2$ and thus dynamical complexity 1.

The more interesting question is whether an irregular state can have low dynamical complexity, for some regular law. I conjecture that this is impossible, and that the following relation holds:

Regularity Principle $Comp(\mathbf{s}) \geq Irreg(\mathbf{s})$.

The regularity principle is easily shown if we assume GBRL, as follows:

1. Locally-similar states are equally likely at all times. From GBRL.
2. Let $Irreg(\mathbf{s}) = n$. Then $G(\mathbf{s}) = 2^n$, so that there By definition
are 2^n states locally similar to \mathbf{s} .

are 2^n states locally-similar to \mathbf{s} .

3. $P(\mathbf{s}$ is the state at $t) \leq 2^{-n}$, for all t .

From 1 and 2 by additivity, as there are 2^n equally-likely possible states at t .

4. $P(F^r \mathbf{s}) \leq r \cdot 2^{-n}$, for all r .

From 3, by additivity.

5. $Sal^r(\mathbf{s}) \leq 2^{-n}$, for all r .

From 4, by definition.

6. $Comp(\mathbf{s})$, i.e. $-\log Sal(\mathbf{s})$, is at least n .

From 5.

7. $Comp(\mathbf{s}) \geq Irreg(\mathbf{s})$.

From 2 and 6. ■

Is GBRL true, however? If we imagine that the law's job is carried out by a human being, call him Fred, then we see that Fred does not have to be able to tell the difference between locally-similar states. He works on one triple at a time, flipping a coin with a suitable bias to decide if the middle bit is to be toggled. He can do his job perfectly without ever knowing the ordering of the triples. Also, Fred doesn't have to be able to tell the difference between a 1 and a 4, or between a 3 and a 6.

Fred cannot distinguish between two locally-similar states just by looking at them, but perhaps he can do it by means of their relations to other states? To rule this out, we need a certain symmetry on the class of states, which is specified below.

Definition The *toggle distance* $Dist(\mathbf{x}, \mathbf{y})$ between two states \mathbf{x} and \mathbf{y} is the number of cells whose bits differ between \mathbf{x} and \mathbf{y} , i.e. the number of cell toggles required to turn \mathbf{x} into \mathbf{y} or vice-versa.

Symmetry Principle Let \mathbf{s} and \mathbf{s}' be locally similar, and let S be any equivalence class of locally-similar states. Then there exists a bijection $f, S \rightarrow S$, such that $Dist(\mathbf{x}, \mathbf{s}) = Dist(f\mathbf{x}, \mathbf{s}')$, for all $\mathbf{x} \in S$.

This principle says that, overall, \mathbf{s} and \mathbf{s}' are equally distant from the members of S . Toggle distances, in other words, cannot be used to distinguish between locally-similar states. The symmetry principle entails GBRL, as is shown below. First we define *toggle functions*.

Definition Suppose there are n spatial cells. Then a toggle function assigns the value *keep* or *toggle* to each cell; i.e. is a function $c: \{1, 2, \dots, n\} \rightarrow \{\text{keep}, \text{toggle}\}$.

In the work below, we actually use toggle functions that are defined on random spatial coordinates. This may seem like an odd manoeuvre, but it mirrors the fact that the dynamical law acts on a triple independently of its spatial position, and thus (as if) in ignorance of its position. If you know only a cell's random coordinate, then you have no idea where that cell is.

Definition A random coordinate system is an assignment of the coordinates $\{1, 2, \dots, n\}$ to the spatial cells, in such a way that each possible assignment has the same probability as the others.

Let the set $\{c^j\}$ contain all the different toggle functions defined on the random coordinates. (The letter ' c ' stands for *control*, as these toggle functions are the means by which the law controls the evolution of the state.) Let the random variable X_t be the state of the system at time t . Individual states will be represented by lower-case bold letters \mathbf{x} , \mathbf{s} and so on. At each iteration, one of the c^j takes place, chosen by the previous state X_t and the toggle probabilities p_1, p_2, \dots, p_8 . Let the toggle function chosen to convert the state at time t to the state at $t+1$ be the random variable C_t , and let S be the equivalence class of locally-similar states to which X_t belongs. We then have the following lemmas.

Blindness Lemma $P(C_t = c^j | X_t = \mathbf{x} \ \& \ X_t \in S) = P(C_t = c^j | X_t \in S)$

This is obvious, as knowing the exact state $X_t = \mathbf{x}$ just tells you the positions of the triples. And since c^j is defined on random coordinates, knowing the positions of the triples is irrelevant.

Non-Convergence Lemma

Let $S = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N\}$ be an equivalence class of locally-similar states. If the symmetry principle holds, and \mathbf{s} and \mathbf{s}' are locally similar, then

$$\sum_{k=1}^N P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x}^k \ \& \ C_t = c^j) = \sum_{k=1}^N P(X_{t+1} = \mathbf{s}' \mid X_t = \mathbf{x}^k \ \& \ C_t = c^j).$$

Proof: The coordinates of the toggled cells in c^j are unimportant, since they are random coordinates that give no information about the positions of the toggled cells. The only relevant aspect of c^j is the total *number* of toggled cells. Suppose, for example, that the world has 12 cells and c^j has 5 toggles. Then, to calculate the left-hand sum, we need consider only the members of S that are 5 toggles from \mathbf{s} . If \mathbf{x} is such a state, then $P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x} \ \& \ C_t = c^j)$ is just the probability that c^j is the unique toggle function (on the cells) that maps \mathbf{x} to \mathbf{s} . Since there are $\frac{12!}{5!7!} = 792$ toggle functions with 5 toggles, which are equally likely given c^j , this probability is just 1/792. The left-hand sum is then just 1/792 times the *number* of states in S that are exactly 5 toggles from \mathbf{s} . Now, by the symmetry principle, the number of \mathbf{x}^k that are 5 toggles from \mathbf{s} equals the number that are 5 toggles from \mathbf{s}' . It follows that the sums are equal. ■

It should be noted that the symmetry principle is not directly used to prove the following theorem, but only via the Non-Convergence Lemma. The idea of the theorem is that, in certain cases, control over a system's evolution requires knowledge of its present state. If one is "blind", i.e. ignorant of the present state, and a fixed value of the control variable produces no convergence, then the system evolves randomly.

Knowledge/Control Theorem If the symmetry principle holds, then GBRL holds.

Proof: We will use mathematical induction on the discrete time t . Assume that, within the arbitrary equivalence class S of locally-similar states, all states are equally likely at t . I.e. if there are N states $\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^N$ in S , then $P(X_t = \mathbf{x}^k \mid X_t \in S) = 1/N$. Let \mathbf{s} and \mathbf{s}' be arbitrary states that

are locally similar (they may or may not be members of S). The theorem of total probability then gives us:

$$\begin{aligned} P(X_{t+1} = \mathbf{s} \mid X_t \in S) &= \sum_{k=1}^N P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x}^k \ \& \ X_t \in S) P(X_t = \mathbf{x}^k \mid X_t \in S) \\ &= \frac{1}{N} \sum_{k=1}^N P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x}^k \ \& \ X_t \in S). \end{aligned}$$

We then apply the theorem of total probability again, using the partition $\{C_t = c^j\}$.

$$P(X_{t+1} = \mathbf{s} \mid X_t \in S) = \frac{1}{N} \sum_{k=1}^N \sum_j P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x}^k \ \& \ X_t \in S \ \& \ C_t = c^j) P(C_t = c^j \mid X_t = \mathbf{x}^k \ \& \ X_t \in S).$$

Using the Blindness Lemma, and the fact that $\mathbf{x}^k \in S$, this simplifies to

$$P(X_{t+1} = \mathbf{s} \mid X_t \in S) = \frac{1}{N} \sum_{k=1}^N \sum_j P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x}^k \ \& \ C_t = c^j) P(C_t = c^j \mid X_t \in S).$$

We then change the order of summation, which allows us to bring the term $P(C_t = c^j \mid X_t \in S)$ outside the k sum.

$$P(X_{t+1} = \mathbf{s} \mid X_t \in S) = \frac{1}{N} \sum_j P(C_t = c^j \mid X_t \in S) \sum_{k=1}^N P(X_{t+1} = \mathbf{s} \mid X_t = \mathbf{x}^k \ \& \ C_t = c^j).$$

We can then use the Non-Convergence Lemma to put \mathbf{s}' in place of \mathbf{s} in the inner sum, so it is clear that

$$(*) \quad P(X_{t+1} = \mathbf{s} \mid X_t \in S) = P(X_{t+1} = \mathbf{s}' \mid X_t \in S).$$

Now let S_1, S_2, \dots be all the equivalence classes of locally-similar states. Then

$$P(X_{t+1} = \mathbf{s}) = \sum_i P(X_{t+1} = \mathbf{s} \mid X_t \in S_i)P(X_t \in S_i),$$

and
$$P(X_{t+1} = \mathbf{s}') = \sum_i P(X_{t+1} = \mathbf{s}' \mid X_t \in S_i)P(X_t \in S_i).$$

From (*) it then follows that

$$P(X_{t+1} = \mathbf{s}) = P(X_{t+1} = \mathbf{s}').$$

Now, since \mathbf{s} and \mathbf{s}' are arbitrary locally-similar states, it follows that, if all pairs of locally-similar states are equally likely at time t , then they remain so at time $t+1$. Also, since they are equally likely at $t=1$, it follows by induction that they are equally likely at all times. Thus all pairs of locally-similar states are indistinguishable, so that GBRL holds. ■

2.4 Conclusions

1. If GBRL (Global Blindness of Regular Laws) holds, then the regularity principle holds, so that all irregular states are dynamically complex. Moreover, since the symmetry principle entails GBRL, it follows that the symmetry principle entails the regularity principle. The situation may be represented thus:

$$\text{Symmetry Principle} \Rightarrow \text{GBRL} \Rightarrow \text{Regularity Principle}.$$

I do not have a proof of the symmetry principle, however, and am not even convinced that it holds. Where does this leave us?

I am very confident that GBRL holds, at least approximately, so that the regularity principle holds as well, at least approximately. An approximate result would be perfectly satisfactory, in view of the very large irregularities of living organisms, and the very short time in which they appeared in the world.

The basic intuition behind GBRL is that the law of motion acts as if it has no knowledge of the global structure of the state, since its action on a given cell is stochastically independent of the global structure. I call my main theorem the Knowledge/Control theorem, since it is based on the idea that knowledge is needed for control. Since the global structure is not known by the law, it is not controlled by the law either. Thus, to the extent that the global structure is unconstrained by the local structure, the global structure evolves randomly, which is what GBRL says.

Let me explain how the proof of the Knowledge/Control theorem works. The Blindness Lemma is so called because it states that the “control” variable C_t is set independently of the state of the system, given knowledge of the local structure, so that the control variable might as well be ignorant of the global state of the system. The Non-Convergence Lemma basically says that there is no automatic convergence to any particular (global) state of the system. For the evolution of the system to converge to a particular state, the control variable C_t has to be sensitive to the state of the system X_t . Any fixed value of C_t is neutral between locally-similar states.

The symmetry principle is used only to guarantee the Non-Convergence Lemma, so that the situation is more accurately as follows.

Symmetry Principle \Rightarrow Non-Convergence Lemma \Rightarrow GBRL \Rightarrow Regularity Principle.

Thus, if some other proof of the Non-Convergence Lemma is found, then the symmetry principle will be redundant.

2. Is there any reason to suppose that the symmetry principle holds? There is the following rough argument. Consider an equivalence class S of locally-similar states, and two locally-similar states \mathbf{s} and \mathbf{s}' , which may or may not be members of S . Is there any way for S to “pick out” \mathbf{s} (say) from the pair $\{\mathbf{s}, \mathbf{s}'\}$? S cannot point to \mathbf{s} using its triple composition, as \mathbf{s}' has exactly the same triple composition. So, if S is able to pick out \mathbf{s} , then it must be by means of the ordering of \mathbf{s} 's triples. The problem here is that S is essentially a disordered object, as it contains every state with some particular triple composition. S is actually determined by the (single)

triple composition of its states. So how can a disordered object like S distinguish s from s' , using the ordering of s ? It seems that S should be blind to such variations of ordering.

The difficulty with this reasoning is that the triples in S determine their *own* ordering, to some extent. A ‘6’ triple, for example, is really ‘110’, and so can be followed only by ‘100’ or ‘101’, i.e. a ‘4’ or a ‘5’. The triples thus have a certain ability to arrange themselves. Hence S is not totally disordered, but merely has no ordering beyond what is implicit in its triples. I just can’t see if the implicit ordering in S is enough to distinguish between s and s' . I feel that it isn’t enough, but I have no proof.

3. The global blindness of regular laws first became apparent to me through computer simulations of cellular automata. One of the first things I tried was to produce the “crystal” structure 0101010101... from a random initial state. This is fairly easy, being accomplished by the following law (among others).

Triple (octal)	Triple (binary)	Toggle probability p_i .
0	000	1
1	001	0.1
2	010	0
3	011	0.1
4	100	0.1
5	101	0
6	110	0.1
7	111	1

These toggle probabilities are easy to understand. The “target state” is 0101010101..., which consists of the triples 010 and 101, i.e. 2 and 5. Thus, if we manufacture these two triples, then we will automatically make the whole state. The triple 000 can easily be turned into 010, just by toggling the middle bit, so this toggle probability is 1. In a similar way, we want to turn every 111 into 101. If the present state is the target state, then we want to leave it unchanged, so that the toggle probabilities for 101 and 010 are both zero. The other triples, namely 001, 011, 100 and 110, are not correct, but may be adjacent to correct triples. These have toggle probabilities

that are smallish, but not zero, so that they will be toggled rarely. If this probability is too small, then errors in the crystal persist for a long time before being changed, so that progress to the target is slow. If this probability is too large, then the correct triples are messed up before they can become established, so that the crystal never emerges. I found by experience that this probability should be between about 0.01 and 0.2.

During the evolution of the system, under this law, from a random initial state, the basic structure emerges quite quickly, in about 10 iterations. Usually the crystal is not perfect, however, as it contains a few errors, such as "...0101010010101...". These errors slowly move, to the left or to the right, as a triple such as '100' or '001' is occasionally toggled. When two errors collide, they annihilate each other, so that eventually all these errors disappear, and the crystal is complete.

This collision and mutual annihilation of errors occurs quite rapidly at first, when there are many errors in the state. After a while, however, there are just two errors left, which (if the world is large) may be far apart. Moreover, they seem to be moving independently of each other, which means that it can take a long time for them to meet. The point I wish to make here is that each error is, in effect, *ignorant* of the position of the other. One might try to modify the toggle probabilities p_i so that these errors will tend to move toward each other, and so annihilate each other more quickly. This is ruled out, however, by the causal locality of the law. Since, in iterating the state at one error, the law only sees the local situation, it cannot take the position of the other error into account. The two errors *have* to move independently of each other. This was the source of the idea that the law has to be blind to the global state.

4. I defined the irregularity of a state as the number of bits of information needed to specify the state, given its local structure. It may be unclear what this has to do with irregularity as defined in Section 1.2. There I defined an irregular object as heterogeneous, aperiodic, patternless, and so on. I can say two things here.

First, it does seem that my definition is a reasonable way to capture the intuitive idea, as it seems to give plausible results. Periodic states, for example, with a short period, all come out as regular. The state 011011011011..., for example, contains only the triples 011, 101 and 110, which can only be joined in that order. Thus the global freedom of the state is very low. States

that strike us as irregular have to contain more kinds of triple, which means that the global freedom (and hence the irregularity) is much higher.

Second, my overall argument does not require that “irregularity” as defined exactly coincide with the intuitive concept. The aim, after all, is simply to prove that living organisms are dynamically complex, via the regularity principle. The important thing, therefore, is that “irregularity” can be *calculated*, or at least estimated, for actual physical objects. If it can be shown that living organisms are irregular in this sense, then it will follow that they are dynamically complex, which is all we need.

At present I am not able to calculate a value for the irregularity of a living organism, but this does seem feasible. In fact, it may even be fairly straightforward to calculate the irregularity of a genome, as this is a mere sequence of terms. (This will give us a lower bound for the dynamical complexity of the organism, if we can assume that the organism is no less complex than its genome.)

In the case of a one-dimensional cellular automaton, in binary, the irregularity of a state is determined merely by its triple composition. The most irregular states seem to be ones where all eight kinds of triple are present in equal numbers, at least roughly. (It should be easy enough to determine if the genome of a particular organism has this property.) In such cases, the irregularity of the state, if its size is n bits, seems to be close to n bits, for large n .

5. A regular law iterates the state of the system one cell at a time, using knowledge of the local triple only. When I was experimenting with different laws, i.e. different sets $\{p_i\}$, trying to generate different structures, it seemed that all I could do was favour some triples over others. The law that generates 0101010101..., for example, merely favours 010 and 101 over the rest.

What if this is true in general? What if a law can make a particular state salient only by favouring the triples it contains? It immediately follows that some states cannot be made salient, by any regular law. If a state contains all kinds of triple in equal numbers, then which triples should you favour? All triples should be given equal treatment, which means that all the toggle probabilities should be equal. It is trivial to show that *all* states are equally likely, at all times, for such a law, so that all states have very low salience.

More generally, if to favour a state you can only favour the triples it contains, then GBRL is immediate. If two states contain the same triples, then no regular law can favour one over the

other. It is this idea that I have tried to capture, using the definitions and arguments of Section 2.3.

Bibliography

- Bennett, C. (1996) “Logical depth and physical complexity”, unpublished manuscript.
- Chaitin, G. (1975) “A theory of program size formally identical to information theory”, *Journal of Assoc. Comput. Mach.* 22, 329-340.
- Darwin, C. (1859) *On the Origin of Species by Natural Selection*, London: John Murray.
- Dawkins, R. (1986) *The Blind Watchmaker*, Reprinted by Penguin, 1988.
- Gärdenfors, P. (1988) *Knowledge in Flux*, Cambridge, Mass: MIT Press.
- Hinegardner, R. and Engleberg, J. (1983) “Biological complexity”, *Journal of Theoretical Biology* 104, 7-20.
- Howson, C. and Urbach, P. (1989) *Scientific Reasoning: The Bayesian Approach*, 2nd ed., La Salle: Open Court, 1993.
- Johns, R. (2002) *A Theory of Physical Probability*, forthcoming, University of Toronto Press.
- Kampis, G. and Csányi, V. (1987) “Notes on order and complexity”, *Journal of Theoretical Biology* 124, 111-21.
- Kolmogorov, A. N. (1968) “Logical basis for information theory and probability theory”, *IEEE Transactions on Information Theory* IT-14, No. 5, 662-664.
- Lewis, D. (1980) “A subjectivist’s guide to objective chance”, Reprinted in Lewis (1986b), 83-113.
- (1986b) *Philosophical Papers Volume II*, New York: Oxford University Press.

- Livingstone, D. N. (1987) *Darwin's Forgotten Defenders*, Edinburgh: Eerdmans and Scottish Academic Press.
- McShea, D. W. (1991) "Complexity and evolution: what everybody knows", *Biology and Philosophy* 6, 303-24.
- Ramsey, F. P. (1931) "Truth and probability", in *The Foundations of Mathematics and Other Logical Essays*, London: Routledge and Kegan Paul.
- Shannon, C. E. (1948) "The mathematical theory of communication", *Bell System Technical Journal*, July and October.
- Solomonov, R. J. (1964) "A formal theory of inductive inference", *Information and Control* 7, 1-22
- Von Neumann, J. (1966) *Theory of Self-Reproducing Automata*, Urbana Illinois: University of Illinois Press.