Euro Jnl Phil Sci (2013) 3:33–67
DOI 10.1007/s13194-012-0056-8

ORIGINAL PAPER IN PHILOSOPHY OF SCIENCE

What is a complex system?

James Ladyman · James Lambert ·
Karoline Wiesner

Received: 19 February 2010 / Accepted: 12 May 2012 / Published online: 19 June 2012

Abstract Complex systems research is becoming ever more important in both the natural and social sciences. It is commonly implied that there is such a thing as a complex system, different examples of which are studied across many disciplines. However, there is no concise definition of a complex system, let alone a definition on which all scientists agree. We review various attempts to characterize a complex system, and consider a core set of features that are widely associated with complex systems in the literature and by those in the field. We argue that some of these features are neither necessary nor sufficient for complexity, and that some of them are too vague or confused to be of any analytical use. In order to bring mathematical rigour to the issue we then review some standard measures of complexity from the scientific literature, and offer a taxonomy for them, before arguing that the one that best captures the qualitative notion of the order produced by complex systems is that of the Statistical Complexity. Finally, we offer our own list of necessary conditions as a characterization of complexity. These conditions are qualitative and may not be jointly sufficient for complexity. We close with some suggestions for future work.

Keywords Complexity · Statistical complexity · Information · Complex system

J. Ladyman (✉) · J. Lambert
Department of Philosophy, University of Bristol, Bristol Avon, UK
e-mail: james.ladyman@bristol.ac.uk

K. Wiesner
Department of Mathematics and Centre for Complexity Sciences,
University of Bristol, Bristol Avon, UK

1 Introduction

The idea of complexity is sometimes said to be part of a new unifying framework for science, and a revolution in our understanding of systems the behaviour of which has proved difficult to predict and control thus far, such as the human brain and the world economy. However, it is important to ask whether there is such a thing as complexity science, rather than merely branches of different sciences, each of which have to deal with their own examples of complex systems. In other words: is there a single natural phenomenon called complexity, which is found in a variety of physical (including living) systems, and which can be the subject of a single scientific theory, or are the different examples of complex systems complex in ways that sometimes have nothing in common?

Hence, the fundamental foundational question in the domain of the complexity sciences is: What is complexity? Assuming that there is an answer to the latter question and that 'complexity' is not just an empty term, we should then ask more particularly whether there is one kind of complexity for all the sciences or whether complexity is domain-specific?

In the next section, we review various attempts to characterize complex systems, and consider a core set of features that are widely associated with complexity by scientists in the field. We argue that some of these features are not necessary for complexity after all, and that some of them are too vague or confused to be of an analytical use. In Section 3 we briefly discuss the connections between complexity and probability and information, before explaining the Shannon entropy and the algorithmic complexity, and arguing that neither is adequate as a measure of complexity. Having introduced a distinction between 'deterministic' and 'statistical' complexity measures, in Section 4 we review some standard measures of complexity from the scientific literature. In Section 5 we consider the popular idea that complexity lies between order and randomness and argue that the measure that best captures the qualitative notion of the order produced by complex systems is the Statistical Complexity. In Section 6, we scrutinise the alleged necessity and sufficiency of a number of conditions that have been proposed in the literature. We offer our own list of necessary conditions for complexity. These conditions are qualitative and may not be jointly sufficient for complexity. We then briefly discuss the very important idea of a hierarchy of organisation in complex systems and close with some suggestions for future work. We do not arrive at a definition of a complex system in the sense of necessary and sufficient conditions but engage in explication of the notion of complex system, so that what is implicit in the science is made more explicit though not completely so. We can hardly think of any important scientific concepts that admit of an analysis into necessary and sufficient conditions but that ought not to dissuade us from introducing such clarity and precision as we can.

2 Complex systems and their features

Complexity science has been comparatively little studied by analytic philosophers of science, however, it has been widely discussed by social scientists and philosophers of social science. Rather than begin our investigations with what they have said about it we think it best to start with what practicing complexity scientists say about what a complex system is. The following quotations (apart from the last one) come from a special issue of *Science* on "Complex Systems" featuring many key figures in the field (*Science* 2 April 1999).

1. "To us, complexity means that we have structure with variations." (Goldenfeld and Kadanoff 1999, p. 87)
2. "In one characterization, a complex system is one whose evolution is very sensitive to initial conditions or to small perturbations, one in which the number of independent interacting components is large, or one in which there are multiple pathways by which the system can evolve. Analytical descriptions of such systems typically require nonlinear differential equations. A second characterization is more informal; that is, the system is "complicated" by some subjective judgment and is not amenable to exact description, analytical or otherwise." (Whitesides and Ismagilov 1999, p. 89)
3. "In a general sense, the adjective "complex" describes a system or component that by design or function or both is difficult to understand and verify. [...] complexity is determined by such factors as the number of components and the intricacy of the interfaces between them, the number and intricacy of conditional branches, the degree of nesting, and the types of data structures." (Weng et al. 1999, p. 92)
4. "Complexity theory indicates that large populations of units can self-organize into aggregations that generate pattern, store information, and engage in collective decision-making." (Parrish and Edelstein-Keshet 1999, p. 99)
5. "Complexity in natural landform patterns is a manifestation of two key characteristics. Natural patterns form from processes that are nonlinear, those that modify the properties of the environment in which they operate or that are strongly coupled; and natural patterns form in systems that are open, driven from equilibrium by the exchange of energy, momentum, material, or information across their boundaries." (Werner 1999, p. 102)
6. "A complex system is literally one in which there are multiple interactions between many different components." (Rind 1999, p. 105)
7. "Common to all studies on complexity are systems with multiple elements adapting or reacting to the pattern these elements create." (Brian Arthur 1999, p. 107)

8. “In recent years the scientific community has coined the rubric ‘complex system’ to describe phenomena, structure, aggregates, organisms, or problems that share some common theme: (i) They are inherently complicated or intricate [...]; (ii) they are rarely completely deterministic; (iii) mathematical models of the system are usually complex and involve non-linear, ill-posed, or chaotic behavior; (iv) the systems are predisposed to unexpected outcomes (so-called emergent behaviour).” (Foote 2007, p. 410)
9. “Complexity starts when causality breaks down” (Editorial 2009)

The last citation well illustrates the difficulties of this field. Clearly many people will have a sufficiently permissive idea of causality to allow that there are causal relationships in complex systems, indeed many people will claim that complexity science’s main job is to understand them. (1) may be true but is hardly informative unless we define what we mean by structure and variations. (2) asks us to choose between the conflation of complexity science with chaos and nonlinear dynamics, or the conflation of complexity with having a lot of components, or the conflation of complexity with a system with different possible histories on the one hand, and a completely subjective answer to our question. (3) and (4) takes us to more interesting territory. The computational notions of data structures, conditional branches and information processing are central to complexity science, and they will be of central importance in Sections 3, 4, and 5. (5) introduces the central idea of nonlinearity. We argue in the next section that while many complex systems are subject to nonlinear dynamics, this is neither a necessary nor a sufficient condition for complexity. We endorse the view of (6) and (7) that a system cannot be complex unless there are many components interacting within it, but we argue this condition is not sufficient, and that it is of limited interest in so far as it is left vague what ‘many’ means. (8) introduces the idea of emergence that we argue is too confused to be part of an informative characterization of a complex system.

Abstracting from the quotations above and drawing on the culture of complexity science as expressed through a wide range of popular as well as academic sources, we arrive at the following list of properties associated with the idea of a complex system:

2.1 Nonlinearity

Nonlinearity is often considered to be essential for complexity. A system is linear if one can add any two solutions to the equations that describe it and obtain another, and multiply any solution by any factor and obtain another. Nonlinearity means that this superposition principle does not apply.

The interesting consequences of nonlinearity are often when the divergence of the system from the superposition principle is particularly extreme with respect to properties other than those specified by the microstate: such as, for example, the property of being alive or dead of an organism whose fate is determined by a nonlinear dynamical system like a skidding car. We often consider systems where we take fine-grained states, such as the positions and momenta of particles for example, as the inputs for dynamical equations, but we are really interested in the values of physical quantities that are coarse-grained with respect to the microstate. Non-linearity in the equations of motion may have the consequence that small differences in the values of the initial conditions may make for radically different macrostates.[1]

In the popular and philosophical literature on complex systems a lot of heat and very little light is liable to be generated by talk of linearity and non-linearity. For example, Klaus Mainzer claims that "[l]inear thinking and the belief that the whole is only the sum of its parts are evidently obsolete" (Mainzer 1994, p. 1). It is not explained what is meant by linear thinking nor what non-linearity has to do with the denial of ontological reductionism. Furthermore, an obvious response to this kind of claim is to point out that it is perfectly possible to think in a linear way about systems that exhibit non-linear dynamics. Unfortunately the discussion of complexity abounds with non-sequiters involving nonlinearity. However, nonlinearity must be considered an important part of the theory of complexity if there is to be one, since certainly many complex systems are also non-linear systems.

Nonetheless, being subject to non-linear dynamics is not a necessary condition for a complex system. For example, there are structures involving linear matrices that describe networks which are a substantial part of what is studied by the complexity sciences; and there are complex systems subject to game-theoretic and quantum dynamics all of which are subject to linear dynamics (MacKay 2008). In general, feedback can give rise to complexity even in linear systems. Neither non-linear dynamics nor linear dynamics can be necessary conditions for complexity, since complexity scientists also study static structures. One may of course argue that such a complex synchronic structure could only come about through a dynamics that is nonlinear. This is the motivation for some of the conceptions of complexity discussed in Section 4.

Non-linearity is also not sufficient for complexity not least because a simple system consisting of say a single chaotic pendulum can be subject to non-linear dynamics but it is not a complex system.

Complexity is often linked with chaos; and as noted above, it may be conflated with it. There are systems that exhibit complexity in virtue of being chaotic. On the other hand, a completely chaotic system is indistinguishable from one behaving randomly. Robert MacKay argues for a definition of complexity as the study of systems with many interdependent components and excludes low-dimensional dynamical systems (MacKay 2008) and hence many chaotic systems. Furthermore, since chaotic behaviour is a special feature

[1]The relationship between macrostates and microstates is key to the complex sciences because very often what is interesting about the system is the way that a stable causal structure arises that can be described at a higher level than that of the properties of the parts (see Section 2.5 on emergence below).

of some deterministic systems, any dynamical system that is stochastic will by definition not be chaotic; and yet complexity scientists study many such systems. So it seems that chaos and nonlinearity are each of them neither necessary nor sufficient for complexity.[2]

However, we may suppose perhaps that in many cases non-linearity in some guise, usually as of dynamics, is at least a necessary part of some set of conditions that are jointly sufficient for complexity. (But there may be more than one such set.)

2.2 Feedback

Feedback is an important necessary condition for complex dynamical systems. A part of a system receives feedback when the way its neighbours interact with it at a later time depends on how it interacts with them at an earlier time. Consider a flock of birds. Each member of the group takes a course that depends on the proximity and bearing of the birds around it, but after it adjusts its course, its neighbours all change their flight plans in response in part to its trajectory; so when it comes to plan its next move, its neighbours' states now reflect in part its own earlier behaviour.

The presence of feedback in a system is not sufficient for complexity because the individuals need to be part of a large enough group to exhibit complexity, and because of how the feedback needs to give rise to some kind of higher level order, such as, for example, the behaviour of ants who are able to undertake complex tasks such as building bridges or farms even though no individual ant has any idea what they are doing, and left to their own they will exhibit much simpler behaviour. The ants behave as they do because of the way they interact with each other.

An abstract way of representing the prevalence of feedback in a complex system is provided by the theory of causal graphs. A chain of causal arrows indicates no feedback while a graph with loops of causal arrows shows feedback. In many contexts feedback is used by a control system, the paradigm of which is the Watt steam regulator, where the speed of rotation of the device interacts in a feedback loop with the steam engine to control the speed of the engine. However this is not a complex system because it has a central controller who sets up the machine. Control theory is importantly related to complexity theory because another central idea associated with complex systems is that of order, organisation and control that is distributed and locally generated (as with the ants) rather than centrally produced (as with the steam regulator). Feedback can also be used for error correction, for example, in motor systems in the brain. We return to this in Section 2.4.

[2]One anonymous referee claimed that it is not possible to define chaos, but on the contrary unlike complexity chaos can readily be defined as systems that exhibit so-called strong mixing. Moreover, recently, Charlotte Werndl has shown that there is a kind of unpredictability unique to chaos (2008). Note that chaos as in chaos theory is always deterministic chaos.

2.3 Spontaneous order

Given the above it is clear that a fundamental idea in complex systems research is that of order in a system's behaviour that arises from the aggregate of a very large number of uncoordinated interactions between elements. However, it is far from easy to say what order is. Notions that are related include symmetry, organisation, periodicity, determinism and pattern. One of the most confusing issues is how order in complex systems relates to the information content of states and dynamics construed as information processing. The problem is that the interpretation of states and processes as involving information may be argued to be of purely heuristic value and based on observer-relative notions of information being projected onto the physical world. We return to the role of information theory in answering our central questions in Section 3. For now we note that the notion of order may mean so many things that it must be carefully qualified if it is to be of any analytical use in a theory of complex systems, but we note also that some such notion is central because pure randomness is sufficient for no complexity whatsoever. On the other hand, total order is also incompatible with complexity. The fact that complex systems are not random but also not completely ordered is of central importance in what follows. Nonetheless, it is a necessary condition for a complex system that it exhibit some kind of spontaneous order.

2.4 Robustness and lack of central control

The order in complex systems is said to be robust because, being distributed and not centrally produced, it is stable under perturbations of the system. For example, the order observed in the way a flock of birds stay together despite the individual and erratic motions of its members is stable in the sense that the buffeting of the system by the wind or the random elimination of some of the members of the flock does not destroy it. A centrally controlled system on the other hand is vulnerable to the malfunction of a few key components. Clearly, while lack of central control is always a feature of complex systems it is not sufficient for complexity since non-complex systems may have no control or order at all. A system may maintain its order in part by utilizing an error-correction mechanism (we return to this in Section 2.4). Robustness seems to be necessary but not sufficient for complexity because a random system can be said to be robust in the trivial sense that perturbations do not affect its order because it doesn't have any. A good example of robustness is the climatic structure of the Earth's weather where rough but relatively stable regularities and periodicities in the basic phenomena of wind velocity, temperature, pressure and humidity arise from an underlying non-linear dynamic. Note that these latter properties are coarse-grainings relative to the underlying state-space. That such properties exist and enable us to massively reduce the number of degrees of freedom that we need to consider is the subject of the next section.

Note that robustness may be formulated in computational language as the ability of a system to correct errors in its structure. In communication theory error correction is achieved by introducing some form of redundancy. This redundancy need not be explicit such as a copy of the string or its parts. It may be more subtle, for instance, exploiting parity checking (Feynman 2000) which is more computationally intensive but also more efficient (the message is shorter) than simple duplication. In his account of complexity discussed in Section 4.1.1 Charles Bennett specifically mentions error correction:

> "Irreversibility seems to facilitate complex behavior by giving noisy systems the generic ability to correct errors."

The situation can be viewed like this: A living cell is arguably a paradigm complex object and does indeed have the ability to repair itself (correct errors), for instance a malfunctioning component may be broken down and released into the surrounding medium. Contrast the cell with a non-complex object such as a gas in a box, a small perturbation of this gas is rapidly dispersed *with no limitations* to the many billions of degrees of freedom within the gas. The cell on the other hand has a one-way direction for this dispersal, errors within the cell are transported out, and errors outside the cell are kept out (assuming, in both cases, the errors are sufficiently small).

2.5 Emergence

Emergence is a notoriously murky notion with a long history in the philosophy of science. People talking about complexity science often associate it with the limitations of reductionism. A strong, perhaps the strongest, notion of emergence is that emergent objects, properties or processes exhibit something called 'downwards causation'. Upwards causation is uncontroversial in the following sense: a subatomic decay event may produce radiation that induces a mutation in a cell that in turn causes the death of an organism. The biological, chemical, economic and social worlds are not causally closed with respect to physics: economic effects may have physical causes. On the other hand, many people take it that the physical world is causally closed in the sense that all physical effects have physical causes. This immediately raises the question as to how complexity relates to physicalism, and whether the latter is understood in terms of causal completeness or merely in terms of some kind of weak asymmetric supervenience of everything on the physical.

There is a sense in which approximately elliptical orbits emerge over time from the gravitational interaction between the sun and the planets. This is not the notion of emergence at issue here. Rather we are concerned with the kind of emergence exemplified by the formation of crystals, the organisation of ant colonies; and in general, the way that levels of organisation in nature emerge from fundamental physics and physical parts of more complex systems. There is much controversy about how this happens and about its implications. Again we conclude that the notion of emergence would need to be very precisely characterized to avoid simply adding to our confusion about the nature of complex systems. Indeed, it is attractive to define emergence in terms of increase in complexity, and so it is preferable not to use the concept in the definition of complexity itself as both concepts seem to be at a similar level of generality.

Emergence is either purely epistemological, in which case it can be strong or weak depending on whether the lack of reduction is in principle or merely in practice; or it is ontological. In its latter form no consensus exists about how to understand it; although it is truistic to say that there is a very important sense in which the physical interaction of atoms and molecules with light and electricity and magnetism and all other physical entities have led to the emergence of the immensely complex and structured system of life on earth including the human brain and the complexity of human culture and social life. Unless one is prepared to say with some metaphysicians that all this enjoys only some kind of second class existence or even none at all, then it seems one must embrace ontological emergence in some sense. The problem then is with resisting or avoiding the argument that one's position requires that one accept downwards causation on pain of making most of reality causally inert or abstract rather than concrete. While non-reductive physicalism without the violation of physicalism may be defensible these issues are not to be solved *en passant* in characterizing complexity. Certainly we must say that emergence in all epistemological senses is necessary for complex systems. If a system doesn't exhibit higher-level order as discussed above, then it is not complex. However, emergence is not sufficient because, for example, an ideal gas exhibits emergent order but is not a complex system.

2.6 Hierarchical organisation

In complex systems there are often many levels of organisation that can be thought of as forming a hierarchy of system and sub-system as proposed by Herbert Simon in his paper 'The Architecture of Complexity' (Simon 1962). Emergence occurs because order that arises from interactions among parts at a lower level is robust. It should be noted of course that such robustness is only ever within a particular regime; the interactions among our neurons generate an emergent order of cognition but only in an operating temperature range of up to 5 degrees Celsius above its normal temperature.

The ultimate result of all the features of complex systems above is an entity that is organised into a variety of levels of structure and properties that interact with the level above and below and exhibit lawlike and causal regularities, and various kinds of symmetry, order and periodic behaviour. The best example of such a system is an ecosystem or the whole system of life on Earth. Other systems that display such organisation include individual organisms, the brain, the cells of complex organisms and so on. A non-living example of such organisation is the cosmos itself with its complex structure of atoms, molecules, gases, liquids, chemical kinds and geological kinds, and ultimately stars and galaxies, and clusters and superclusters.

2.7 Numerosity

Philip Anderson in his paper about complexity 'More is Different' (Anderson 1972) famously argues against reductionism and also emphasises the importance of considering hierarchies of structure and organisation to understand complex systems as just discussed. His title alludes to the fact that many more than a handful of individual elements need to interact in order to generate complex systems. The kind of hierarchical organisation that emerges and gives rise to all the features we have discussed above, only exists if the system consists of a large number of parts, and usually, only if they are engaged in many interactions. We call this numerosity and return to it in Section 6.

2.8 Remarks

The above discussion makes it clear that the definition of complexity and complex systems is not straightforward and is potentially philosophically interesting. The notions of order and organisation introduced above and the idea of feedback are suggestive of an information-theoretic approach to complexity, since complex systems can be often helpfully be construed as maintaining their order and hierarchical organisation by the exchange of information among their parts. Many people think that it is fruitful to think of complex systems as characterized by the way they process information, as well as by the information-theoretic properties of the data we obtain by sampling them. Since the notion of information is itself a philosophically problematic one, in the next section we explain the fundamentals of information theory so as to approach a mathematical theory of complexity.

3 Complexity, information, and probability

Complex systems are often characterized in information-theoretic terms. This is based on the idea that the order of a complex system can be understood as maintained by the internal processing of information. Whether or not it makes sense to talk about physical systems containing and processing information independently of our representations of them is a vexed issue. It has recently been argued by Chris Timpson in the context of quantum information theory that talk of information flowing in a system is based on a confusion between abstract and concrete nouns (Timpson 2006). His view is that information is not a physical quantity. On the other hand, there are radical proposals such as that popularized by the great physicist John Wheeler to take information as the primitive "component" of reality from which other physical properties are derived. We will not take a stand on these issues here but rather address the question of whether or not the standard measure of information content can be used to measure complexity. We remain neutral about the ontological status of information itself and in particular about how if at all information theory relates to the semantic conception of information.

Claude Shannon was the first to formalise the statistical reliability of communication channels and from an engineering perspective invented a measure of information that can be applied to any probability distribution. Hence, Shannon information theory is closely connected with probability theory.

Another way to define information comes from algorithmic complexity theory. It is based on the concepts of computing machines and programmes with input and output. Here, the size of a programme which runs on a universal computing machine and generates a particular output is a measure of information (of that output).

We will review both definitions of information and their use in devising measures of complexity. Most importantly we will in the following point out how they both in their original form measure randomness rather than complexity (or structure).

3.1 Shannon entropy

Claude Shannon was concerned with quantifying information of a message so he could devise an optimal method for sending it over a communication channel. He writes: "Frequently the messages have meaning; [...] These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages." (Shannon 1948, p.1) The Shannon entropy is a measure of the probability distribution over the messages only and is not concerned with the content of the messages. Loosely speaking, the Shannon entropy is higher the more uniform the probability distribution; it reaches its maximum when all messages have the same probability. Formulating the following three axioms Shannon showed that there is a unique function H which satisfies them (up to a multiplicative constant), see Eq. 1 (Shannon 1948, p. 49). Consider a probability distribution $\Pr(X) = \{\Pr(x_1), \Pr(x_2), \ldots, \Pr(x_n)\}$

1. H should be continuous in the probabilities $\Pr(x_i)$.
2. If all the probabilities $\Pr(x_i)$ are equal, $\Pr(x_i) = \frac{1}{n}$, then H should be a monotonic increasing function of n.
3. If the choice [of x_i] is broken down into two successive choices, the original H should be the weighted sum of the individual values of H.

The resulting unique function H, now called the Shannon entropy, is formally defined for a random variable X as follows, with probability distribution $\Pr(X)$ and alphabet $\mathcal{X} \ni x$, where the alphabet corresponds to the "messages":

$$H(X) := -\sum_{x\in\mathcal{X}} \Pr(x) \log \Pr(x) \ . \tag{1}$$

The log is usually taken to the base 2. One can see that Shannon entropy is a measure of randomness of a probability distribution by considering the example of tossing a coin. The coin represents a binary random variable. If the coin is fair it has maximum entropy ($H(X) = 1$ bit), each outcome is equally likely. In contrast, when the coin has probability of heads equal to, say 0.9 and that of tails equal to 0.1, it has much lower entropy ($H(X) = 0.47$ bit), its outcomes are less random or, equally significant, more predictable.

Because Shannon entropy is a function of a probability distribution it does not measure information of a single object (unless this object is characterised by a trivial probability distribution in which case it has zero Shannon entropy). Instead, Shannon entropy measures the information of one object out of a set in the sense of the amount of "surprise" when this message is selected; "surprise" is inversely proportional to its probability.

3.2 Kolmogorov complexity

Shannon entropy cannot express the notion of randomness, order, or complexity of a single object. It can only express properties of a total set of sequences under some distribution. Combining notions of computability and statistics one can express the complexity of a single object. This complexity is the length of the shortest binary programme from which the object can be effectively reconstructed (see e.g. Li and Vitnyi 2009). It is known as *algorithmic complexity* or *Kolmogorov complexity* and was formulated independently by Solomonoff, Kolmogorov and Chaitin in the mid-sixties (Kolmogorov 1965, 1983). It has been claimed to be more fundamental than information theory in so far as a key result of the latter can be derived from it Kolmogorov (1983).

It is easiest to think in terms of binary strings and to associate every object (numbers, vectors, lists, etc) with a string that describes it. In the case of numbers this will no doubt be just the binary expansion of the number, in the case of a list it might be the ASCII code. The algorithmic complexity of a random string, for example, is the length of the string itself (plus a constant independent of the string). Any repetition of digits or symmetry in the string will allow the programme that outputs the string to be shorter than the string itself. The more compressible the string is the shorter is the programme. A random sequence is maximally incompressible while a string of all zeros, say, is maximally compressible. It is important that the algorithmic-complexity measure is independent of the source of the string. For example, a string of apparently random numbers might in fact have been generated from the decimal expansion of π according to some rule, but its Kolmogorov complexity will be maximal because nothing in the string itself tells us where to find it in the decimal expansion of pi and hence we are unable to use the latter as a shortcut to outputting the string.

The concept of algorithmic complexity discussed so far does not allow for the *difficulty* of running the progamme and generating its output. In particular we will want to consider the amount of time it takes for the programme to generate its output. This is called the *running time* or "time complexity" of the programme. We will get back to this in Section 4.1.1.

3.2.1 Lempel-Ziv complexity

From a practical point of view it is important to note that the Kolmogorov complexity is uncomputable. As a way to approximate the Kolmogorov complexity, Lempel and Ziv invented an algorithm for compressing a sequence by taking advantage of repetitions of a specified maximum length in the sequence (see e.g. Cover and Thomas 2006). The length of the resulting programme is considered to be a good estimator of the Kolmogorov complexity for many practical purposes. The *Lempel-Ziv complexity* is defined as the ratio of the length of the programme and the length of the original sequence.

3.3 Deterministic versus statistical conceptions of complexity

One of the most intriguing ideas is that complexity lies between order and randomness. The latter two notions are here meant in the sense of algorithmic complexity theory. Systems of interest to complexity scientists are neither completely ordered nor completely random. Neither entirely ordered, nor entirely random processes qualify as 'complex' in the intuitive sense since both admit of concise descriptions. Ordered processes produce data that exhibit repetitive patterns describe the string '0101 . . . 0101 consisting of 1000 bits concisely as:

`'01' 500 times`

Whereas random 'noise' such as the outcomes of 1000 coin-toss trials, while lacking periodic structure, affords a concise statistical description:

`'0' with probability 0.5, '1' otherwise`

What is needed then would seem to be a measure of complexity different from the algorithmic complexity and we make this argument more explicit below. We first note that measures of complexity may be deterministic or statistical in the following sense. A *deterministic* measure of complexity treats a completely random sequence of 0s and 1s as having maximal complexity. Correspondingly a *statistical* measure of complexity treats a completely random sequence as having minimal complexity. Although the Shannon entropy is a function over a probability distribution and the algorithmic complexity is a function of an instantiation, both are deterministic in the above sense.

If we are interested in a measure that tells us when we are dealing with a complex physical system we clearly do not want a measure that is maximal for completely random data strings. Note also that the algorithmic complexity and the Shannon entropy of the data produced by a biased coin would be lower than that of a fair coin although there is no sense in which the former is more complex that the latter. Both functions are monotonically increasing with randomness whereas we want a function that is unimodal and peaked between randomness and complete order. Hence, neither Shannon entropy nor algorithmic complexity of a data string are an appropriate measure for the complexity of a physical system. We need a measure that is statistical in the sense defined above.[3] In the next section we review some important measures of complexity in the literature. Note that such measures of complexity may be applied to three targets, namely, to the methods used to study certain systems, to data that are obtained from certain systems, or to the systems themselves. Some say that the complexity sciences are simply those that use certain characteristic methods. In this view it makes little or no sense to speak of complex data sets or complex systems. Call this a pragmatic account of complexity. We set this point of view aside. At the other extreme, some measures of complexity are supposed to be applied directly to systems to tell us whether they are complex or not. Call this a physical account of complexity. Among physical accounts of complexity are the theories of Logical Depth, Thermodynamic Depth and Effective Complexity reviewed in Sections 4.1.1, 4.1.2, and 4.2.1. These views allow for derivative notions of complex data (the data produced by complex physical systems), and complex methods (the methods appropriate to studying such systems). Finally, a third kind of account of complexity applies primarily to data and derivatively to physical systems and to the methods appropriate to their study. Call this a data account of complexity. Among data accounts of complexity are the Statistical Complexity and the Effective Measure Complexity reviewed in Sections 4.2.2 and 4.2.3.

4 Measures of complexity

Before reviewing several measures of complexity from the literature we consider the following categorisation suggested by Seth Lloyd on conditions for a good measure of complexity:

> "Three questions that are frequently posed when attempting to quantify the complexity of the thing [...] under study are (1) How hard is it to describe? (2) How hard is it to create? (3) What is its degree of organization?" (Lloyd 2001, p. 7)

Lloyd continues to categorise measures of complexity into three different groups according to whether they measure the *difficulty of description*, the *difficulty of creation*, or the *degree of organisation*. Examples of measures of difficulty of description are the measures in the previous Section 3—entropy and algorithmic complexity. In this section we will look at some proposed measures of the difficulty of creation and degree of organisation. We adopt Lloyd's classification because it illustrates some of the differences. Note that it is similar to our tripartite distinction of Section 3.3 in so far as the complexity

[3]Note that we are not here talking about whether the system that produces the data is deterministic or not. Of course, the Shannon entropy of a probability distribution is insensitive to whether that probability distribution was produced by deterministic or an indeterministic system. Our point is just that a good measure of complexity will not be maximal for random data strings.

of the data (the product) corresponds to the difficulty of description, and the complexity of the system corresponds to the difficulty of creation. Of course, these degrees of difficulty are related in so far as there is a correlation between how hard things are to create and how hard they are to describe; for example, a living organism is both difficult to create and to describe. On the other hand, some things that are easy to describe are difficult to create such as for example a perfectly spherical and homogenous ball. So the relations among them are not straightforward.

4.1 Difficulty of creation

> "Such a measure [...] should assign low complexity to systems in random states and in ordered but regular states [...]. Complexity is then a measure of how hard it is to put something together." (Lloyd and Pagels 1988, p. 189)

This quote, again from Lloyd, suggests that it might be possible to define and measure the complexity of a system by considering its history. This is the approach taken by Bennett with his measure of Logical Depth (Bennett 1988), as well as by Lloyd and Pagels (1988) with their notion of Thermodynamic Depth.

We now look at these measures in some more detail.

4.1.1 Logical depth

1. Complex objects lie somewhere between complete order and complete disorder.
2. Complex or logically deep objects cannot be produced quickly so any adequate complexity measure should obey a *slow growth law*.
3. The history of a complex object is not only long, but non-trivial, that is, the most plausible explanation for the object's origin is one that entails a lengthy computation/causal process.

A book on Number Theory might seem very difficult or "deep". However, it has very low Kolmogorov complexity, all of its theorems and proofs follow from a few axioms. Clearly, this does not capture our sense of difficulty which is based on the experience that it takes a long time to reproduce all these theorems and proofs. A proper definition of the "depth" of an object is going to be a compromise between the programme size and computation time.

Charles Bennett's approach to finding a measure of this "depth" begins with the intuition that some objects, such as, for example, the human body contain "internal evidence of a non-trivial causal history" (Bennett 1988, p. 227). He formalizes the idea of an object that has such a history in the language of algorithmic information theory. Bennett's key idea is that long causal processes are usually necessary in order to produce complex or 'deep' objects; deep objects are produced by short processes only with very low probability.

Suppose, following Paley (2006, p. 7–8), you find a pocket watch on a deserted heath, and consider all possible explanations for the watch's presence. Some explanations will involve a watchmaker who, after training for many years finally manufactures the watch in question, plus some further story about how the watch came to be on the heath, such as that the chain snapped while the owner was out walking. However, other explanations could be altogether more unusual, for instance, perhaps the moment you set foot on the heath the impact of your foot caused atoms in the sand grains to leap into the exact configuration of a pocket watch awaiting your discovery. Intuitively some of these explanations are more plausible than others. Bennett's intuition is that complex objects are those whose most plausible explanations describe long causal processes, and that hypothesising the objects to have had a more rapid origin forces one to adopt drastically less plausible stories. Here 'explanation' means a complete description of the object's entire causal history.

Bennett's challenge now is to say exactly how we grade various causal histories in terms of plausibility; why is the watchmaker hypothesis a better explanation than the springing-into-existence hypothesis? To this end Bennett turns to algorithmic information theory. Bennett proposes that the shortest programme for generating a string represents the a-priori most plausible description of its origin, while a 'print' program on the other hand offers no explanation whatsoever, and is equivalent to saying 'it just happened' and so is effectively a null-hypothesis. If we view ad-hocness in an hypothesis as the use of arbitrary unsupported assertions that could or should be explained by a more detailed theory, then a hypothesis represented by a programme s-bits longer than the minimal programme is said to suffer from s-bits of 'ad-hocness' when compared with the hypothesis represented by the minimal programme.

The Logical Depth of the system then, represented as a string, is dependent on the running time of the programmes that produce it. Bennet's example of the digits of π illustrates the difference between algorithmically complex and logically 'deep'. A 'print' programme has high algorithmic complexity but is logically very 'shallow' because it runs very fast. Whereas an algorithm which computes the digits of π is a comparatively short programme but which has a longer running time. The former would constitute a lot of 'ad-hocness' while the latter is based on a theory. Bennett defines Logical Depth as follows:

Logical Depth — A string x is d, b *deep* [...] if and only if d is the least time needed by a b-incompressible program to print x. (from Li and Vitnyi 2009, p. 512)

Logical Depth seems to be a reasonable measure of the difficulty of creation and to that extent is a good measure of complexity. In terms of our threefold distinction at the end of Section 3.3 it is applied to the system itself and so may be thought of as a physical measure. However, the Logical Depth is not computable because it is defined in terms of the algorithmic complexity which is provably non-computable. We argue at the end of Section 4.3 that this makes it unfit for purpose. We also argue in Section 5 that one of Bennett's motivating intuitions for his definition of Logical Depth, namely that complexity is maximal for systems that are neither completely ordered nor completely random is misleading despite its initial appeal and popularity among commentators.

4.1.2 Thermodynamic depth

The notion of Thermodynamic Depth introduced by Lloyd and Pagels (1988) shares much of its informal motivation with Logical Depth.

> "[Complexity] is a property of the evolution of a state and not of the state itself, it vanishes for ordered and disordered states, it is a universal, physical quantity, and it corresponds to mathematical complexity in problem solving. (Lloyd and Pagels 1988, p. 187)"

Their description of complexity is motivated by the fact that Lloyd and Pagels do not consider multiple copies of a complex object to be significantly more complex than a single complex object, since, as they point out, producing a single specimen of a plant takes billions of years of evolution, but seven more specimens of the same plant can be easily produced with a single one. The process from no specimens to several is only minutely harder than the process from none to one. Putting aside concerns about the biological possibility of a single specimen in isolation, the idea, much like Bennett's, is that complexity has something to do with how difficult it is to produce something.

Lloyd and Pagels introduce three requirements that any reasonable measure of complexity defined over these trajectories should obey. (1) The function must be continuous in the probabilities, (2) if all trajectories are equally likely then the measure should be monotonically increasing in n, the number of trajectories, and (3) additivity: the measure associated with going from state a to b to c should be equal to the sum of the measures for going from a to b and from b to c. In other words the complexity of building a car from scratch is the complexity of building the components from scratch and building the car from the components.

These are effectively the same demands made by Shannon on the entropy function (see Section 3.1). It will therefore not come as a surprise later on that Thermodynamic Depth is not a measure of order as it claims to be but a measure of randomness.

Like Bennett, Lloyd and Pagels consider the set of histories or *trajectories* that result in the object in question. A trajectory is an ordered set of macroscopic states $a_i, b_j \dots c_k$. If we are interested in the complexity of a state d, then we consider all trajectories which end in state d and their probabilities.

The depth of state d is determined by the probability of the trajectory $a_i, b_j \dots c_k \dots d$ by which it was reached:

$$\mathcal{D}(d) = -k \ln \Pr\left(a_i, b_j \dots c_k \dots d\right) \ . \tag{2}$$

where k is an arbitrary constant which is set equal to the Boltzmann constant to make the connection between information theory and statistical mechanics.

How the probabilities are to be interpreted is unclear but Lloyd and Pagels reject a subjective reading.

> "[A]lthough the complexity depends on the set of experiments that determine how a system reaches a given state, the measure as defined is not subjective: two different physicists given the same experimental data will assign the state the same complexity." (Lloyd and Pagels 1988, p. 190)

Because their discussion draws heavily on concepts from statistical mechanics, such as those of phase spaces, microscopic/macroscopic states and so on, it is reasonable to suppose the probabilities used in defining a system's depth are of the same sort as those found in statistical mechanics. But the correct interpretation of these probabilities is still controversial with some prominent authors, notably Jaynes (1957a, b), arguing for a purely epistemic interpretation. If the biggest problems Thermodynamic Depth faces were those it inherits from statistical mechanics then Lloyd and Pagels could be relieved because, while statistical mechanics faces many philosophical problems like most (if not all) successful physical theories, the subject has hardly been brought down by them. However, bigger problems are highlighted by Crutchfield and Shalizi (1999). First we are not told how long the trajectories we consider should be. Complexity is claimed to be a property of a process but how do we identify when a process starts? To remedy this Crutchfield and Shalizi suggest looking at the rate at which depth increases, rather than its absolute value. The rate of increase of depth turns out to be the rate at which the system's entropy increases. Entropy rate, however, is a measure of randomness and does not capture complexity in the here desired way. This renders the Thermodynamic Depth itself unsuitable as a measure of complexity. Lloyd and Pagels specifically state that a reasonable complexity measure should not award high complexity to a purely random process but Thermodynamic Depth does exactly this. An N body Ising spin system being cooled below its critical temperature assumes a frozen magnetic state 'up','down'... which has very many probable predecessors which results in a very large Thermodynamic Depth of this frozen state again showing that it is not a good measure of complexity.

4.2 Degree of organisation

4.2.1 Effective complexity

> "A measure that corresponds [...] to what is usually meant by complexity [...] refers not to the length of the most concise description of an entity [...], but to the length of a concise description of a set of the entity's regularities." (Gell-Mann 1995, p. 2)

Effective complexity, introduced by Gell-Mann (1995), Gell-Mann and Lloyd (2004) (see also Gell-Mann and Lloyd 1996), is a statistical measure based on the (deterministic) Kolmogorov complexity. To find the complexity of an object Gell-Mann considers the shortest description, not of the entity itself, but of the ensemble in which the entity is embedded as a typical member. Here, 'typical' means that the negative logarithm of its probability is approximately equal to the entropy of the ensemble. This assumes ways of estimating what the ensemble is. The Kolmogorov complexity (Gell-Mann uses the equivalent term *algorithmic information content*) of the ensemble is the length of the shortest programme required to list all members of the ensemble together with their probabilities.

The resulting definition of *Effective Complexity* of an object (represented as a string) is the algorithmic information content of the ensemble in which the string is embedded. Effective complexity is designed to measure the regularities of a string as opposed to the length of a concise description. This makes Effective Complexity a measure of *degree of organisation* as opposed to *difficulty of description*. The problem with this measure, as with Logical Depth, it that it is not computable, as Gell-Man states himself: "There can exist no procedure for finding the set of all regularities of an entity." (Gell-Mann 1995, p. 2). From a conceptual point of view, however, it is consistent with the above discussed notions of complexity. A random string is assigned zero Effective Complexity. Maybe more surprising, the digits of π have very low Effective Complexity.

Logical depth vs effective complexity Gell-Man argues that Logical Depth and Effective Complexity are complementary quantities. The Effective Complexity of a set of data might be very high such as a detailed diagram of the Mandelbrot set until we learn that it can be computed from a very simple formula and so its Logical Depth is the determining factor to understand it rather than its Effective Complexity. It is often not clear whether something which is apparently complex is really the combination of an underlying simplicity and a certain amount of Logical Depth. Gell-Mann argues against Logical Depth being responsible for the intricateness of structure observed in nature calling on the "relentless operation of chance". These chance events Gell-Man attributes to the inherent probabilistic nature of quantum mechanics. The resulting Effective Complexity of the universe is a result of many "frozen accidents" accumulating over time generating the many regularities we observe today rather than a predetermined algorithm with a large Logical Depth.

4.2.2 Effective measure complexity

> "Quantities which qualify as measures of complexity of patterns [...] (1) they are measure-theoretic concepts, more closely related to Shannon entropy than to computational complexity; and (2) they are observables related to ensembles of patterns, not to individual patterns. Indeed, they are essentially Shannon information needed to specify not individual patterns, but either measure-theoretic or algebraic properties of ensem-

bles of patterns arising in a priori translationally invariant situations." (Grassberger 1986, p. 907)

Introduced by Peter Grassberger, the *Effective Measure Complexity* (EMC) is measuring the average amount by which the uncertainty of a symbol in a string is decreasing due to knowledge of previous symbols (Grassberger 1986). For a completely random string the measure is zero since no information is gained about one random symbol by looking at another random symbol. The more statistical dependencies there are, i.e. the more *order* there is, the larger the Effective Measure Complexity is going to be. The mathematical definition is based on the Shannon entropy (Eq. 1). For strings x^N of length N $H(X^N)$ is the Shannon entropy of the probability distribution $\Pr(x^N)$ (replacing $\Pr(x)$ in Eq. 1). Considering strings of length $N+1$ from the same system one can define the increase in entropy as the difference $H(X^{N+1}) - H(X^N)$ and call this *entropy rate* h_N. The limit for $N \to \infty$ exists under certain assumptions and is denoted by h. The entropy rate is a measure of persistent randomness which can not be eliminated by observing the system over a long time or space. It is randomness *inherent* to the system. Grassberger then defines the Effective Measure Complexity as the accumulated difference between the "believed" entropy rate h_N and the "true" entropy rate h.

$$EMC = \sum_{N=0}^{\infty} (h_N - h) \; . \tag{3}$$

This difference quantifies the perceived randomness which, after further observation, is discovered to be order. The more observations are required to identify this order the higher the Effective Measure Complexity.

4.2.3 The statistical complexity

"The idea is that a data set is complex if it is the composite of many symmetries." (Crutchfield and Young 1990, p. 227).
"When we use the word 'complexity' we mean degrees of pattern, not degrees of randomness." (Shalizi and Crutchfield 2001, p. 824)

The Statistical Complexity was introduced by James Crutchfield and Karl Young to measure information processing of nonlinear dynamical systems (Crutchfield 1994; Crutchfield and Young 1989). Behind this measure lies a technique of causal-state reconstruction further developed by Shalizi and Crutchfield (2001). From either empirical data or from a probabilistic description of behaviour one infers a model of the hidden process that generated the observed behaviour.

The inference proceeds as follows. From empirical data probabilities for successive observations conditioned on previous observations are gathered. Possible such paths (sequences of observation types) are put into the same equivalence class when they have the same conditional probability for future observations. The resulting sets each define predictions for future observations and, hence, each constitute a regularity. Note that this procedure applied to a sequence of coin flips with an unbiased coin would result in merely one set containing the probabilities of observing a head or a tail conditioned on any previous observation which is $1/2$ and independent on the previous observations. What we have put here in layman's terms is the construction of a so-called "minimal sufficient statistic", a well-known concept in the statistics literature. Crutchfield and Shalizi introduced the term *causal states* following Hume's weak sense of causality: "one class of event causes another if the latter always follows the former" (Shalizi and Crutchfield 2001, p. 824). This set of causal states $\mathcal{S}$ and probabilistic transitions between them, summarised in the so-called ϵ-machine, represents our ability to predict the process's future behaviour. And it is the minimal and optimal such representation (Shalizi and Crutchfield 2001). The mathematical structure of the ϵ-machine is that of a hidden Markov model or stochastic finite-state automaton. The Statistical Complexity, denoted C_μ, is then the Shannon entropy over the stationary distribution over the causal states $\mathcal{S}$:

$$C_\mu = H(\mathcal{S}) \tag{4}$$

Hence, the Statistical Complexity is a derivative quantity of the ϵ-machine which in itself contains all the available information about the process's organisation. For a system to have high Statistical Complexity it must have a large number of causal states and hence a large amount of regularities. Successful applications of the causal-state reconstruction technique to complex systems range from protein configuration (Li et al. 2008) and atmospheric turbulence (Palmer et al. 2000) to self-organisation (Shalizi et al. 2004).

While the Kolmogorov complexity (see Section 3.2) and Thermodynamic Depth (see Section 4.1.2) award random strings high complexity; the former by defining a random string as one whose minimal program is approximately the length of the string itself, the latter by measuring a systems rate of change of entropy, the Statistical Complexity awards zero complexity to random strings and high complexity to strings with many statistical dependencies.

Bennett considers the computational resources (programme size, space-time complexity etc.) needed to reproduce the given string verbatim. In contrast Crutchfield's approach is to consider the simplest computational model capable reproducing the statistical properties of the string. In this sense the Statistical Complexity is a data-driven account of complexity.

4.3 Computable versus non-computable measures

In this section it is argued that only computable measures are suitable as measures for the complexity of a physical system.

A good measure of complexity needs to be statistical as opposed to deterministic. This rules out Thermodynamic Depth and leaves us with Effective Complexity, Logical Depth, Effective Measure Complexity, and Statistical Complexity. If we, in addition, impose computability as a requirement, which seems reasonable in so far as we are interested in using our measure of complexity to tell us which systems in the world are complex, then among the measures we have considered we are left with Effective Measure Complexity and Statistical Complexity as the ones that meet both requirements of being computable and statistical. Note, that Effective Measure Complexity always bounds from below the Statistical Complexity of any given process (Shalizi and Crutchfield 2001, Theorem 5). Furthermore, as we have outlined above, with the Statistical Complexity we have access to not only a measure of the degree of organisation but also to a description of that organisation. Hence, we come to the conclusion that the measure that best captures the qualitative notion of the order produced by complex systems is that of the Statistical Complexity.

The Statistical Complexity is applied to data sets, as opposed to systems or to methods. A data set with a high Statistical Complexity is not necessarily produced by a complex system. For example, highly structured data may be produced by a centralised control mechanism rather than by emergence. However, when looking at large data sets gathered from real world systems, high Statistical Complexity is probably produced by complex systems. In principle, there is no reason to suppose that there could not be some true property of systems that measures their complexity even though we cannot compute it. However, since the Statistical Complexity can be computed and used in practice to infer the presence of complex systems, it is the best candidate we have considered for a measure of the order produced by complex systems.

4.4 Remarks

We have not argued exhaustively by considering every putative measure of complexity in the literature, rather we have taken a representative sample of complexity measures and considered at least one that falls into each of the four categories of measure generated by our two binary distinctions among different kinds of measure. We have not attempted an exhaustive survey of such measures, since, that would be a Herculean task that would take up a whole long paper in itself, so we offer only a guide to further reading. A discussion of complexity measures in the context of dynamical systems can be found in Wackerbauer et al. (1994). In it the authors suggest a classification scheme of dynamical systems into four categories similar to the scheme presented here. The scheme distinguishes between order and chaos using the notions of generating and homogeneous partitions. A discussion of complexity measures in the context of physical systems can be found in Badii and Politi (1999). This book surveys different kinds of complex system and measures of complexity. Of the ones, mentioned there the Grammatical Complexity deserves discussion. Opposed to the statistical measures reviewed so far Grammatical Complexity is concerned with higher level organisation while discarding finite-length correlations as not complex. Hence, it is worth considering as a measure when the investigator is interested in higher-level order and infinite correlations. A survey of measures of complexity and their criticism can be found in Grassberger (1989). In it the author concludes that no unified measure can be found nor should there be one since complexity is intricately linked to meaning which is subjective and unmeasurable.

5 The "peaked" complexity function

It is generally agreed upon that a complex system is structured. One thing that seems certain is that complex systems are not ones in which all the parts are performing random walks, and this is enough to distinguish complexity from just "being complicated". A gas at equilibrium is an incredibly complicated system but it is not complex just because the behaviour of its parts is effectively random in the sense that the behaviours of individual molecules are largely uncorrelated. However, the one point of agreement in the complex-systems community is that a measure of complexity should assign its highest value to systems which are neither completely random nor completely ordered. Usually this is justified by giving examples which are intuitively not complex. At the random end of this scale a sequence of coin flips or an ideal gas are often mentioned. At the ordered end of the scale a perfect crystal or a checker board pattern are examples. Following this logic one would have to assume that a disordered crystal would score higher in complexity than a perfect crystal. Equally if we start to insert random bits into a string of ones we get a more "complex" sequence. This form of randomness in order does not seem to capture what we mean when we call a cell complex or when we consider a bee hive complex. Of course, the latter are mostly complex because of their dynamics not because of their static structure. A pattern in the sand on a beach might be considered complex because it has many regularities and yet is not perfectly symmetric. The reason it is complex is not because it is imperfect but rather because such regular patterns form no matter how the wind blows. So there is an interplay between regularities and a form of robustness which is important here.

What is the reason for randomness to be stated as a necessary ingredient by so many authors? An answer to this is given by Edgar Morin in his treatise "Method" (Morin and Belanger 1992). In his view the role of randomness is crucial to generate and sustain complexity. To summarise his view in our terms: Complexity could "be" without randomness, but it could not become nor sustain itself; randomness acts as a source of interaction which is necessary for the build-up of correlations. Because of the second law of thermodynamics, however, these correlations will decohere and the initial random state will be populated once more. Only, if there is a balance between random encounters and, literally, constructive interactions can decoherence be avoided and order maintained. Thus, there is no complex system which is static. Complex systems are in a continuous process of maintaining their complexity. And hence randomness and order are intricately linked to complexity.

Careful investigation shows that the Statistical Complexity does increase monotonically with the orderliness of the system.[4] Hence, it is a good measure of structure, even in the presence of noise, but it does not indicate any of the other features of a complex system such as robustness and adaptivity. The ubiquitous single-humped curve as the "ideal measure" of complexity (zero for perfect order and zero for complete randomness) is not fulfilled by the Statistical Complexity, since it has a maximum at the point of perfect order. This "ideal curve" is misleading in the following sense; a crystal with many defects is just that, ordered and noisy at the same time while not being a complex system—while being perfectly robust to the end of being completely inflexible it is not adaptive to signals from the environment. A fractal is a complex pattern because it has many regularities and symmetries superposed. If natural systems were perfect fractals they would still cause wonder. The fact that the fractals observed in nature are not perfect is not the reason that they are considered more complex. Rather, the fact that fractals are generated under very noisy circumstances illustrates the combination of two goals in a natural systems: The generation of structure, the more perfect the better, and the robustness of that structure (and hence its generation mechanism). Clearly, these are opposites and the system needs to optimise the two. An ideal gas is perfectly robust in its structure simply because it does not have any in the first place. A crystal is also very robust in its structure but once it is broken it cannot "mend itself". A school of fish is very structured though clearly not a perfect grid of fish. But once a predator has disturbed the school formation it can then reform immediately. So rather than requiring a good measure of complexity to have its maximum value for systems which are neither completely random nor completely ordered this suggests that a measure of order is not sufficient. A complex system is expected to score high on a measure of order. And it also scores high on a measure of "adaptivity" (Holland 1992) or, as Sandra Mitchell calls it, "responsiveness" (Mitchell 2009). The latter is then responsible for the robustness of the order under perturbations and outer influences or, in other words its ability to correct errors. In the light of this we conclude that the "peaked" form of complexity functions is not a prerequisite for a measure of complexity but it is a *consequence* of a complex system's adaptiveness while wanting to maintain its structure. Hence, we conclude that the Statistical Complexity does not measure complexity as such but rather order. But because it measures order in the presence of noise it can be used to measure the order produced by complex systems since these systems necessarily exhibit noise.

[4]For a proof consider the following. For a given number of causal states the Statistical Complexity (Eq. 4) has a unique maximum at uniform probability distribution over the states. This is achieved by a perfectly periodic sequence with period equal to the number of states. As soon as deviations occur the probability distribution will likely not be uniform anymore and the Shannon entropy and with it the Statistical Complexity will decrease. Hence, the Statistical Complexity scores highest for perfectly ordered strings.

6 The features of complex systems revisited

6.1 Necessary and sufficient conditions for complexity

In this section we revisit some of the features of complex systems discussed in Section 2 and discuss to what extent they are necessary or sufficient for complexity. We start by giving the following *tentative definition of complexity* which we consider as a physical account of complexity:

Complex System (physical account)	A complex system is an ensemble of many elements which are interacting in a disordered way, resulting in robust organisation and memory.

6.1.1 Ensemble of many elements

Most definitions or descriptions of complexity mention 'many elements' as a characteristic. For interactions to happen and for pattern and coherence to develop, the elements have to be not only many but also similar in nature. This is the prerequisite for the condition of interaction. For systems to be able to interact or *communicate* (in the broadest sense) with each other they have to be able to exchange energy or matter or information. Physical systems have to be particles comparable in size and weight, subject to the same physical laws. In biology, cells before they form multi-cellular organisms are indistinguishable / identical so they can maximally communicate (exchange genetic information) with each other. Non-physical systems, e.g. social structures have to be similar in character, behaviour, or rules obeyed. Other examples are the brain consisting of neurones, an ant colony consisting of ants, a financial market consisting of agents. While it is a necessary condition for a complex system that there are many similar parts of some kind it should be noted that not all the parts have to be similar and of the same kind. Weather systems are composed of many similar parts in so far as oxygen, nitrogen and water are all gases composed of molecules, yet of course these gases are importantly different in their properties and these differences give rise to important features of the climate. Moreover, weather systems also include geological features and radiation too. Also weather systems are composed of many similar parts in so far as there are many small volumes of the atmosphere that are similar and which interact to give rise to the weather. Furthermore, the hierarchical structure of complex systems reflects a necessity for some kind of similarity. Ensembles of similar elements at one level form a higher-level structure which then interacts with other similar higher-level structure. As an example consider a society. Many cells make up a human body, many human bodies make up a group, many groups make up a societal structure. Every complex system has at least some parts of which there are many similar copies interacting. Of course, the idea of a large number of components is vague, but we note that such vagueness is ubiquitous in the sense that it is vague whether, for example, a quantum system consists of a large number of particles so that it reduces to classical behaviour in accordance with the correspondence principle, or whether there are sufficient birds flying together to be regarded as a flock.[5]

6.1.2 Interactions

The second condition is for the elements of the system to have the means of interacting. Interaction can be either exchange of energy, matter, or information. The mediating mechanism can be forces, collision or communication. Without interaction a system merely forms a "soup" of particles which necessarily are independent and have no means of forming patterns, of establishing order. Here we need to emphasise that interaction needs to be direct, not via a third party, a so-called common cause. Thus, we require not merely probabilistic dependence but direct dependence. Locality of interaction is not a necessary condition, neither is it sufficient. Interactions can be channelled through specialised communication and transportation systems. Telephone lines allow for non-local interactions between agents of a financial market, nerve cells transport chemical signals over long distances. It is important that the idea of interaction here is not necessarily one of dynamics. What is important is the idea of the dependence of the states of the elements on each other. Non-linearity of interactions is often mentioned as necessary condition for a complex system. This claim can be easily refuted. Complex networks, which are clearly complex in their structure and behaviour, can be defined by matrices which are inherently linear operations. The fact that non-linearity is not necessary illustrates a commonly mentioned feature of complex systems. They can be perceived as complicated yet be defined by simple rules, where complicated means difficult to predict.

6.1.3 Disorder

Disorder is a necessary condition for complexity simply because complex systems are precisely those whose order emerges from disorder rather than being built into them. Interaction is the basis for any correlations to build up and hence for order to arise from disorder. An example of order rising out of disorder is thermal fluctuations which are necessary for most biological processes to take place such as protein binding to DNA which is driven by the energy provided through thermal fluctuations. The result is the transcription of DNA to make new proteins.

6.1.4 Robust order

The above three conditions are all necessary for a complex system to emerge but they are not sufficient because many similar elements interact in a dis-

[5] We are grateful to an anonymous reviewer whose criticisms based on the example of the climate forced us to clarify these points.

orderly way in a gas but they are not complex systems. However, a system consisting of many similar elements which are interacting in a disordered way has the potential of forming patterns or structures. An example are the Rayleigh-Bénard convection patterns. On an appropriate time scale the order is robust. This means that although the elements continue to interact in a disordered way the overall patterns and structures are preserved. A macroscopic level arises out of microscopic interaction and it is stable (on that time or length scale). This kind of robust order is a further necessary condition for a system to be complex.

6.1.5 Memory

> "A system remembers through the persistence of internal structure."(Holland 1992)

Memory is a straightforward corollary of robust order.

6.2 Statistical complexity revisited

Since we have come to the conclusion that the Statistical Complexity best captures the qualitative notion of the order produced by a complex system, we offer the following data-driven qualitative definition of a complex system:

Complex system (data-driven account)	A system is complex if it can generate data series with high *Statistical Complexity*.

How do the data-driven account of complexity and the physical account of complexity from Section 6.1 represent the features of complex systems with which our discussion began, namely, nonlinearity, feedback, emergence, local organisation, robustness, hierarchical organisation, and numerosity? It is difficult to say how well the Statistical Complexity captures *nonlinearity* and *feedback*. Certainly some systems of high Statistical Complexity are nonlinear and exhibit feedback, for instance, see Crutchfield's investigation into the logistic map (Crutchfield and Young 1990). However, neither is necessary or sufficient for high Statistical Complexity. Even strong forms of ontological *emergence* are consistent with high Statistical Complexity. However, high Statistical Complexity is also prima facie compatible with strong forms of ontological reduction of high level phenomena to lower level phenomena, in which case it would be construed as a useful measure solely for epistemological purposes. It is not yet clear what, if any, light Statistical Complexity sheds on emergence. There is an interesting suggestion to define emergence using the predictive efficiency of a process which is defined in terms of Statistical Complexity (Shalizi and Moore 2003, p. 9).

As discussed earlier, *local organisation* and *robustness* are related, and the former appears necessary for the latter because order that is centrally produced may be easily disrupted by perturbing the point of control. Locally produced order, on the other hand, involves many distributed copies which serve as back-ups for each other. A statistically complex system is one that exhibits a diverse range of behaviour over time i.e. it has many causal states. In contrast a system with no memory, say a coin being tossed repeatedly, can be modelled accurately with only a single causal state, and thus has zero Statistical Complexity. Complex systems, unlike a coin toss, must possess some memory, some record of their past.

Local organisation and robustness appear to be related by this idea of memory; both memory and robustness involve stability over time, and for this stability we need some form of back-up or redundancy with which to correct errors. This may be provided when we have a system whose order is not centrally controlled.

6.3 The hierarchical nature of complex systems

We now turn to the underlying architecture—by which we mean the relative organisation of the elements—which leads to properties such as robustness or de-centralised control. As pointed out by Simon (1962), hierarchical organisation involves features which are characteristic of paradigmatic complex systems.

A system is organised hierarchically when it is composed of subsystems that, in turn, have their own subsystems, and so on. Many social systems, such as the military, corporations, universities and legal systems are hierarchies. A university, for example, consists of staff which are grouped in departments which are grouped in faculties. The highest level of a university is the vice-chancellor but one can go up the hierarchy even further to networks of universities and so on. It is not only social systems that exhibit hierarchical organisation. Biological organisms are composed of systems, organs and cells that are made of nuclei, cell membranes, and mitochondria, etc. The parts of a cell are made of molecules, which are made of atoms, and so on. Clearly, natural and social hierarchies are ubiquitous, however, there is an important difference between social and physical systems in terms of the nature of the hierarchical levels to which we return at the end of this section.

A core problem in complex systems research is to understand how a complex structure can have evolved out of a simple one. In particular, if the underlying processes that give rise to evolution (by which we do not mean biological evolution exclusively) are following a random walk through the space of possibilities the likelihood of anything complex and hence non-random evolving is negligible. This mystery of how complexity can evolve is partly explained by hierarchical organisation. Indeed, 'selective trial and error' can lead to complex forms on a geologically meaningful time scale. As Simon points out, in such a "selective" process each intermediate level forms a stable configuration which gets "selected" for further levels to build on top. In this way, the hierarchical nature of complex systems is explained by the efficiency and stability of hierarchical building processes. Here, natural evolution and engineering tasks are very similar. Neither would be very successful if the final product was to be built from scratch every time (recall the above discussion of the watchmaker). This process of assembling stable intermediates into new stable configurations at a higher level does not rely upon any particular nature of the subsystems. Hence, the way that stable intermediates lead to hierarchically organised complex systems is essentially the same for physical, biological, and social systems. This is why stable democratic states cannot be created overnight; they need to have a hierarchy of civic structures as well as the high-level democratic processes for electing the government. It is apparent that hierarchical architecture and the evolution and stability of complex systems are intricately linked.

So the robustness of many complex systems arises from the hierarchical process which generates them and their hierarchical organisation. Many other common features of complex systems can be similarly explained. The absence of central control, for instance, may also be a consequence of hierarchical organisation. For example, a cell is composed of molecules which are composed of atoms etc. Thus, an organism is composed of a multitude of identical particles. None of these has any particular role at the outset. The particles are equipped with the same interaction forces and dynamical degrees of freedom. The different levels of the hierarchy are made up of regroupings of lower levels, of a redefining of the system's boundaries. Physical systems do not have a central control because the highest level of the hierarchy is merely the sum of the levels below.

Here we find a general difference between physical or biological systems on the one hand and social systems on the other that is not mentioned in Simon's discussion. In a social system an element at a higher level does not necessarily consist of elements at the level below. The CEO of a company, for example, does not 'consist' of her employees. She is, indeed, a novel central control element. This being said, many social systems do exhibit complex features. For example, the stock market is highly complex in the sense that there are many traders interacting, equipped with buying and selling capacity, and this gives rise to feedback, as in a market panic. The resulting dynamic is highly unpredictable though and it is often claimed that the market as a whole behaves like a random walk. Hence, whether or not a company is a complex system in the full sense of the term is an open question. Although the interactions between many companies make for a good candidate.

The connection between social and physical hierarchies can be regained to some extent when we consider the additional element of representation. A head of a political party, for example, not only controls the party but also represents its members. Social hierarchies can combine to larger and larger systems through the mechanism of representation. An organisation like the United Nations could not exist if we didn't allow for representative governance. There is no such thing as "representation" in most physical systems and this makes the robust communication between millions of molecules even more impressive.

The hierarchical architecture of complex systems is linked to the Statistical Complexity as a measure of a complexity. Crutchfield writes that "[...] a data set is complex if it is the composite of many symmetries." (Crutchfield and Young 1990). A hierarchical structure possesses exactly the architecture which can generate many symmetries. Such symmetries are, for example, the low- and high-frequencies which are associated with the upper and lower levels of the hierarchy, as described by Simon. Thus, the definition of a complex system as one which has high Statistical Complexity does overlap if not coincide with the definition of a hierarchy. It is an open question whether any non-hierarchical, non-engineered structure generates data with high Statistical Complexity.

It is fair to ask how the natural sciences were so successful before they began to take the hierarchical structure of nature into account. Clearly, physical and not just biological reality is hierarchical. Subatomic particles form atoms, they in turn form molecules, crystals, condensed matter and ultimately rocks, planets, solar systems, galaxies, clusters and superclusters. However, the natural sciences can generate meaningful and predictive theories for higher levels without knowing much about the lower levels, because often a higher level can be explained with only a coarse picture of the level underneath it or indeed with no picture at all as with the gas laws and phenomenological thermodynamics, both of which were developed independently of how we now understand their microstructural underpinnings. Similarly, the statics and dynamics of fluids and solids was studied long before the existence of atoms and molecules was known, and an atomic physicist does not need to know about subatomic particles to master her field.

However, in contemporary science, understanding the levels above and below is becoming more and more important to the science of a particular level. Many of the phenomena now studied in the natural and social sciences are beyond the scope of any one discipline. For example, understanding human diseases requires knowledge of the physics of electromagnetism, the chemistry of molecular bonding, the biology of cellular organisms, and the psychology of the human mind. Hence, the increasing success of complex systems research shows that our scientific knowledge has reached the point where we are able to put pieces together that we had previously been gathering separately. Complexity science contributes to and is only possible because of the unity of science.

7 Conclusion

The right measure of the order exhibited by complex systems must be computable and not be maximal for randomness. Statistical Complexity has these features, and is also in other ways illustrative of a good complexity measure. Inferring the minimal ϵ-machine from a data sequence means identifying some form of pattern or structure, statistical or deterministic in the data. It is time to raise an important philosophical issue for complexity science, or, indeed, any scientific theory, namely the question of realism versus instrumentalism. What are the prospects for adopting a realist position toward the patterns represented by ϵ-machines as opposed to the view, say, that the patterns are merely useful tools for predicting the system's behaviour but not real features of the world?

In nature patterns are everywhere: the familiar arrow-head flight formation of geese; bees which assemble honeycomb into an hexagonal tessellation; the spiral formations of sunflowers and galaxies. Cicadas of the genus *magicicada* exhibit the interesting temporal pattern that their life-cycles are prime numbers of years to avoid synchrony with potential predators' lifecycles. The scientific study of naturally occurring patterns requires both a suitable means for formally representing patterns and a method of inferring patterns from data that picks out objective features of the world. The reconstruction of ϵ-machines meets the former challenge by representing patterns via the smallest computational model capable of statistically reproducing them. The latter challenge is an instance of the classic problem of natural kinds articulated by Plato:

> ...[T]o be able to cut up each kind according to its species along its natural joints, and to try not to splinter any part, as a bad butcher might do. (Cooper 1997, p. 542)

Patterns are doubly challenging for the scientific realist because not only are physical theories concerning patterns subject to the usual anti-realist arguments (namely the pessimistic meta-induction, the underdetermination argument and so on), but also all examples of patterns, such as the ones given above, are dependent on the existence of some underlying matter to exhibit them so there is always the objection that all that really exists is the matter not the pattern. A related but stronger motivation for antirealism about the patterns studied by the complexity sciences is reductionism. The reductionist begins with the plausible premise that the causal power of a high level phenomenon, say a brick, is nothing over and above the sum of its micro constituents' causal powers. Next she argues that only those objects whose causal interactions are necessary to produce the observed phenomena should be regarded as the real ones. The conclusion is that bricks and windows are superfluous causal terms in our explanation of observed phenomena. In other words, all phenomena arise from the low level knocking about of objects, and these are the only objects one needs to take to exist. An example of this strategy is eliminative materialism in the philosophy of mind, which seeks to replace folk-psychological discourse about minds, beliefs and emotions etc. with neuroscientific discourse about neural activity within the central nervous system. Contemporary metaphysics has also employed causal exclusion arguments against higher level phenomena, taking atomic collisions and suchlike as paradigm causal mediators and leading some authors (Merricks 2001) to dismiss talk of baseballs breaking windows, on the grounds that what *really* does the interacting are the brick's and the window's microstructures. Higher level phenomena such as bricks and windows are redundant causal terms and we are urged to view them as merely predictively useful abstractions, that is, we are urged to be instrumentalists about bricks and windows.

So how might we justify a realist attitude towards patterns? The most promising approach pioneered by Daniel Dennett in his paper "Real Patterns" (Dennett 1991) and endorsed in one form or another by Wallace (2003) and Ladyman et al. (2007) is to appeal to the remarkable amount of computational effort saved at the expense of very little predictive power when adopting a theory of patterns. Consider Wallace's example: suppose we are interested in predicting the movements of a tiger when hunting its prey; one approach would be to model the entire tiger, prey, environment system on the sub-atomic level and study its dynamics according the laws of quantum mechanics, since quantum mechanics is by far the most accurately verified physical theory ever devised we should be confident our predictions will be highly accurate. This strategy is obviously completely impractical because there is no realistic hope of obtaining an accurate quantum mechanical description of the entire system let alone computing its time evolution. Ascending to a higher level, we may view the tiger and its prey as biological systems and attempt to predict their behaviour through our understanding of biochemistry and neuroscience. This approach is vastly simpler than the previous approach but still way beyond our practical reach. Notice also that some accuracy will have been sacrificed. Highly improbable quantum fluctuations simply do not occur in the biological picture but this loss of accuracy is overwhelmingly insignificant in comparison to the reduction of the computational complexity of modelling the situation. Ascending to the even higher level of animal-psychology we have again compromised on predictive accuracy since the fine details of the individuals will not be modeled. However, as before this small loss in predictive power is more than compensated for by now having a tractable means for accurately predicting how the situation will play out. To quote Wallace's slogan:

> A tiger is any pattern which behaves as a tiger. (Wallace 2003, p. 7)

This sounds like realism: tigers exist and they are anything that behaves like a tiger. However the discussion so far has been distinctly instrumentalist in tone reporting how ascending to higher levels serves to increase our practical predictive capacity.

To sum up:

1. We are pursuing a realist account of patterns in the physical world.
2. It is noted that some patterns allow for an enormous simplification of our model of reality while trading off relatively little predictive power.
3. Identifying patterns via their predictive utility is suggestive of an instrumentalist and anti-realist approach. How can a pattern's predictive utility be given ontological weight?

The beginnings of a theory to answer the latter question were given by Dennett (1991) and refined by Ross (2000) and Ladyman et al. (2007) resulting in an ontological account of patterns called *Rainforest Realism*. The main thesis is that since computation is a physical process there is a determinate matter of fact about whether a pattern is predictively useful, namely, if it is possible to build a computer to accurately simulate the phenomena in question by means of said pattern, and if doing so is much more computationally efficient than operating at a lower level and ignoring the pattern. Since computation is a physical process, it is the laws of physics which determine whether such and such a computation can occur and hence whether a given pattern is real. In Section 5 of Shalizi and Crutchfield (2001) prove that ϵ-machines are the unique, optimal predictors of the phenomena they are meant to model.

> Our first theorem showed that the causal states are maximally prescient; our second, that they are the simplest way of representing the pattern of maximum strength; our third theorem, that they are unique in having this double optimality. Further results showed that ϵ-machines are the least stochastic way of capturing maximum-strength patterns and emphasized the need to employ the efficacious but hidden states of the process, rather than just its gross observables, such as sequence blocks. (Shalizi and Crutchfield 2001, p. 853)

The claim is that the patterns inferred and represented by ϵ-machines are the simplest, most accurate means for predicting the behaviour of the system they describe. According to rainforest realism it must be possible to construct a computer capable of simulating the phenomena under investigation by means of some pattern, in this case an ϵ-machine, in order for the pattern to be real. We know that in the vast majority of cases this will be possible. ϵ-machines are typically members of a simple class of computers such as finite state automata and we know from daily experience that we are adept at building computing devices capable of simulating these, namely, desktop computers. So long as we have enough memory, energy and time to run our computer the pattern described by the simulation is an objective feature of the simulated phenomena. The process of inferring ϵ-machines does indeed identify real patterns. Rainforest realism, despite its emphasis on the predictive utility of patterns, avoids instrumentalism through a criterion of pattern reality—the possibility or otherwise of such and such a computing device—whose truth is determined by objective facts about the physical world.

Acknowledgements We are extremely grateful to several anonymous referees for this journal and to the editor for very helpful comments and criticisms, and also to the students of the Bristol Centre for the Complexity Sciences doctoral programme over several years for their comments on our ideas. James Ladyman acknowledges the support of the AHRC Foundations of Structuralism project. Karoline Wiesner acknowledges funding through EPSRC grant EP/E501214/1.

References

Anderson, P. W. (1972). More is different: Broken symmerty and the nature of the hierarchical structure of science. *Science, 177*, 393–396.

Brian Arthur, W. (1999). Complexity and the economy. *Science, 284*, 107–109.

Badii, R., & Politi, A. (1999). *Complexity: Hierarchical structures and scaling in physics*. Cambridge University Press.

Bennett, C. H. (1988). Logical depth and physical complexity. In R. Herken, (Ed.), *The universal Turing machine, a half-century survey* (pp. 227–257). Oxford: Oxford University Press.

Cooper, J. M. (Ed.) (1997). *Phaedrus in Plato complete works*. Hackett.
Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd edn.). Wiley-Blackwell, September.
Crutchfield, J. P. (1994). The calculi of emergence: Computation, dynamics and induction. *Physica D: Nonlinear Phenomena, 75*(1–3), 11–54.
Crutchfield, J. P., & Shalizi, C. R. (1999). Thermodynamic depth of causal states: Objective complexity via minimal representations. *Physical Review E, 59*(1), 275–283.
Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Physical Review Letters, 63*, 105.
Crutchfield, J. P., & Young, K. (1990). Computation at the onset of chaos. *Entropy, Complexity and the Physics of Information, SFI Studies in the Sciences of Complexity, VIII*, 223–269.
Dennett, D. (1991). Real patterns. *The Journal of Philosophy, 88*(1), 27–51.
Editorial. (2009). No man is an island. *Nature Physics, 5*, 1.
Feynman, R. (2000). *Feynman lectures on computation*. Westview Press.
Foote, R. (2007). Mathematics and complex systems. *Science, 318*, 410–412.
Gell-Mann, M. (1995). What is complexity. *Complexity, 1*, 1.
Gell-Mann, M., & Lloyd, S. (1996). Information measures, effective complexity, and total information. *Complexity, 2*(1), 44–52.
Gell-Mann, M., & Lloyd, S. (2004). Effective complexity. In M. Gell-Mann, & C. Tsallis, (Eds.), *Nonextensive entropy – interdisciplinary applications*. The Santa Fe Institute, OUP USA.
Goldenfeld, N., & Kadanoff, L. P. (1999). Simple lessons from complexity. *Science, 284*, 87–89.
Grassberger, P. (1986). Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics, 25*(9), 907–938.
Grassberger, P. (1989). Problems in quantifying self-generated complexity. *Helvetica Physica Acta, 62*, 489–511.
Holland, J. H. (1992). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence* (new edn.). MIT Press, July.
Holland, J. H. (1992). Complex adaptive systems. *Daedalus, 121*(1), 17–30.
Jaynes, E. T. (1957a). Information theory and statistical mechanics. *The Physical Review, 106*(4), 620–630.
Jaynes, E. T. (1957b). Information theory and statistical mechanics, ii. *The Physical Review, 108*(2), 171–190.
Kolmogorov, A. N. (1965). Three approaches to the quantitive definition of information. *Problems of Information Transmission, 1*, 1–17.
Kolmogorov, A. N. (1983). Combinatorial foundations of information theory and the calculus of probabilities. *Russian Mathematical Surveys, 38*(4), 29–40.
Ladyman, J., Ross, D., Spurrett, D., & Collier, J. (2007). *Everything must go: Metaphysics naturalized*. Oxford University Press.
Li, C.-B., Yang, H., & Komatsuzaki, T. (2008). Multiscale complex network of protein conformational fluctuations in single-molecule time series. *Proceedings of the National Academy of Sciences, 105*(2), 536–541.
Li, M., & Vitnyi, P. M. B. (2009). *An introduction to Kolmogorov complexity and its applications* (3rd ed.). Springer, March.
Lloyd, S. (2001). Measures of complexity: A nonexhaustive list. *Control Systems Magazine, IEEE, 21*, 7–8.
Lloyd, S., & Pagels, H. (1988). Complexity as thermodynamic depth. *Annals of Physics, 188*, 186–213.
MacKay, R. S. (2008). Nonlinearity in complexity science. *Nonlinearity, 21*, T273.
Mainzer, K. (1994). *Thinking in complexity: The complex dynamics of matter, mind and mankind*. Springer.
Merricks, T. (2001). *Objects and persons*. Oxford University Press.
Mitchell, S. (2009). *Unsimple truths: Science, complexity, and policy*. University of Chicago Press.
Morin, E., & Belanger, J. L. R. (1992). *Method: Towards a study of humankind : The nature of nature: 001*. Peter Lang Pub Inc, November.
Paley, W. (2006). *Natural theology*. Oxford University Press.
Palmer, A. J., Fairall, C. W., & Brewer, W. A. (2000). Complexity in the atmosphere. *IEEE Transactions on Geoscience and Remote Sensing, 38*(4), 2056–2063.

Parrish, J. K., & Edelstein-Keshet, L. (1999). Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science, 284*, 99–101.

Rind, D. (1999). Complexity and climate. *Science, 284*, 105–107.

Ross, D. (2000). Rainforrest realism: A Dennettian theory of existence. In D. Ross (Ed.), *Dennett's philosophy: A comprehensive assessment* (Chapter 8, pp. 147–168). MIT Press.

Shalizi, C. R., & Crutchfield, J. P. (2001). Computational mechanics: Pattern and prediction, structure and simplicity. *Journal of Statistical Physics, 104*(3), 817–879.

Shalizi, C. R., & Moore, C. (2003). What is a macrostate? Subjective observations and objective dynamics. *cond-mat/0303625*.

Shalizi, C. R., Shalizi, K. L., & Haslinger, R. (2004). Quantifying self-organization with optimal predictors. *Physical Review Letters, 93*(11), 118701.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423; 623–656.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

Timpson, C. G. (2006). The grammar of teleportation. *British Journal of Philosophy of Science, 57*(3), 587–621.

Wackerbauer, R., Witt, A., Atmanspacher, H., Kurths, J., & Scheingraber, H. (1994). A comparative classification of complexity measures. *Chaos, Solitons & Fractals, 4*(1), 133–173.

Wallace, D. (2003). Everett and structure. *Studies In History and Philosophy of Modern Physics, 34*(1), 87–105.

Weng, G., Bhalla, U. S., & Iyengar, R. (1999). Complexity in biological signaling systems. *Science, 284*, 92–96.

Werner, B. T. (1999). Complexity in natural landform patterns. *Science, 284*, 102–104.

Whitesides, G. M., & Ismagilov, R. F. (1999). Complexity in chemistry. *Science, 284*, 89–92.