

Complexity and Homoplasy

Daniel W. McShea

January 1995
(Revised July 1995)

Mailing Address:
Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501

Phone: 505-984-8800
E-mail: dwm@santafe.edu

I. INTRODUCTION

II. COMPLEXITY

- A. Three Distinctions
- B. Eight Types
- C. Difficulties

III. HOMOPLASY

- A. Types of Homoplasny
- B. Relative Probabilities

IV. SUMMARY

I. INTRODUCTION

Consider the two objects in Figure 1. Which is more complex? Object A might seem more complex because it contains more parts. Or A might seem simpler, because its parts are all the same size and disconnected from each other, while B's vary in size and are nested within each other. On the other hand, the arrangement of parts in A might look more complex because it is less regular, less predictable, while in B the parts are neatly lined up along the midline of the object. The problem here is that complexity has various meanings. Both objects are complex, but in different senses.

Complexity has been the subject of a number of empirical and theoretical studies in recent years. At issue in some, for example, has been the existence of a general trend in complexity (McShea, 1991, 1993; Valentine et al., 1993); the relationship between complexity and such factors as ecological specialization (Flessa et al., 1975), morphological variability (Lauder, 1981), environmental variability (Hughes and Jackson, 1990), and species longevity (Boyajian and Lutz, 1992); and the probability of evolutionary homoplasy (Sanderson, 1993; see also Gould, 1970; MacBeth, 1980; Laurent, 1983). However, the significance of much of this interesting work, and its relation to other issues in evolutionary biology, has been at least partly obscured by the ambiguities and multiple meanings of the word "complexity."

Regarding trends, for example, many share the impression

that organisms have become more complex, on average, over the history of life. But more complex in what sense? Number of parts may have increased but irregularity of configuration may not, or vice versa. Nestedness of parts may have increased, but their numbers may have decreased, or vice versa. Is it complexity of structure or complexity of function that is thought to increase? We cannot begin to study trends until the various senses of complexity have been distinguished.

Likewise for complexity and homoplasy. Consider the powerful logic connecting complexity and evolutionary irreversibility, the logic of Dollo's Law (Gould, 1970; Gould and Robinson, 1994). A complex structure such as an organism, or a complex part of one, that has been significantly modified in evolution never returns precisely to its former state -- to the ancestral condition -- because the changes along the way have been too many, and the resulting pathway too improbable to be retraced (Sanderson, 1993, and references therein). Functional reversal is possible and simple morphological patterns may recur -- dolphins have regained the fusiform shape of their fish ancestors -- but if the transformation has been sufficiently complex, differences will be found in the details. Precise reversal is possible in principle, but exceedingly improbable in practice.

The logic is sound, but whether or not it is applicable depends on which sense of complexity is intended. One sort of complex evolutionary transformation might be a series of additions of independent steps to the developmental program of an

organism. In this case, for reversal-in-detail to occur, each would have to be removed (although not necessarily in the same order added). The greater the number of independent additions, that is, the greater the complexity of the transformation, the more improbable that complete reversal will occur. On the other hand, if the complexity is built up hierarchically, so that developmental steps added later are dependent on those added earlier, reversal may be a simple matter of eliminating the early steps. Take the lower supports out from under a wing of the developmental superstructure, and it collapses, restoring level ground. This example does not contradict Dollo's Law; the law remains valid, but only within a restricted domain defined by a particular sense of complexity.

In this chapter, I explore the various senses of complexity and develop a scheme for organizing and understanding them. Then I attempt to show, in a preliminary way, the manner in which each sense might be related to homoplasy, in particular, to the ease or probability of homoplasy. The relationship is important not only for the study of the development of complex structure (Wake and Roth, 1989), but also for systematics, where homoplasy is mainly considered an obstacle to the discernment of evolutionary relationships. One assumption of some cladistic methodologies is that homoplasy is improbable (see discussion in Sober, 1988), relative to other sorts of transformation, and that therefore the best-supported phylogeny is the one with the fewest homoplastic events. One purpose here is to analyze and evaluate this

assumption in the case of complex transformations. For example, one apparent implication of Dollo's law is that the assumption is especially likely to be valid for complex structures; as we shall see, the reasoning is correct for some senses of complexity, but not others.

II. COMPLEXITY

No single understanding of complexity has been proposed that encompasses all nuances and context-specific meanings. But a common theme has emerged from a number of theoretical treatments in biology (Hinegardner and Engelberg, 1983; Kampis and Csányi, 1987; Katz, 1986, 1988; Wicken, 1987; McShea, 1991; Slobodkin, 1992) and is implicit in recent empirical studies (Cisne, 1974; Boyajian and Lutz, 1992; McShea, 1992, 1993; Valentine et al., 1993). Complexity is heterogeneity. Most generally, the more diverse a system is, the more numerous its component parts and interactions and the more irregular their configuration, the more complex it is. But as Figure 1 dramatizes, complexity is itself a complex term, encompassing a number of different senses or types. I have identified eight types in all, based on three distinctions.

A. Three Distinctions

1. Differentiation and configuration. Differentiation has to do with numbers of parts (or interactions; see below) and the differences among them. For systems in which parts are discrete and types easily distinguished, differentiability complexity is

simply the number of different types (Valentine et al., 1993). Where variation among parts is continuous and types intergrade, complexity is some function of degree of differentiation, such as the variance or a variance analog (McShea, 1992, 1993). Configurational complexity is irregularity of arrangement. For objects in which the parts are arranged in a linear series, irregularity of configuration might be a function of degree of differentiation among adjacent parts, perhaps the number of changes in part-type along the series.

To make the contrast clearer: a parade is typically highly differentiated, consisting of many differently dressed individuals, but configurationally simple, in that individuals typically march in regular rows. Or more abstractly, in Figure 2, A is more differentiated, because it contains five different parts (shapes), while B has only four. On the other hand, the arrangement of parts in A is more regular, so B is more complex configurationally.

2. Objects and processes. In the present context, an object is a larger-scale entity which is composed of smaller-scale parts, and a process is a larger-scale activity or operation which is composed of smaller-scale interactions. Thus, object-complexity refers to differentiation or irregularity among the physical parts of a system, and process-complexity to the same properties of the interactions among the parts. It is the parts that interact, but their interactions can be considered in the

abstract, without reference to the parts themselves. Indeed, there is no necessary correspondence between the complexity of objects and processes. A single object may participate in essentially one process, as does the heart in vertebrates, or many, as does the liver. In Figure 2, C and D have the same object-differentiation, because they consist of precisely the same parts, but they differ in the number of distinct interactions among parts, shown by the arrows. D has more arrows and thus greater process-differentiation.

Organismal complexity has at least four components: molecular, developmental, morphological, and physiological. Molecular and morphological complexity refer to the heterogeneity of objects. For DNA molecules, the parts might be genes or nucleotides, and for whole organisms, cells or organs. Developmental and physiological complexity refer to processes. For development, the component interactions might be morphogenetic events, and for physiology they might be metabolic steps. Only developmental and morphological complexity will be considered here, although the logic of the argument can be easily adapted for molecules and physiology.

3. Hierarchical and non-hierarchical complexity. Non-hierarchical complexity refers to the number and configuration of parts at a given spatial scale, or interactions at a given spatial and temporal scale. For either parts or interactions, complexity is thus a scale-relative property; for example, the surface of a marble is smooth and featureless (i.e., simple) at

the usual scale of human perception but rough and irregular (i.e., complex) at the molecular scale. Non-hierarchical complexity also varies with scalar range; in Figure 2, E might be considered to have four parts at a single scale, or two parts at one scale and one at each of two other scales, depending on how the scalar spectrum is carved. Finally, scalar levels are independent to some extent (Allen and Starr, 1982; Salthe, 1985), and thus no object or process has a true or essential complexity. In other words, no scalar level has priority over any other, at least a priori. In particular, complexity at the genetic level is no more the "true" complexity of an organism than complexity at the scale of its cells or organs.

For objects, hierarchical complexity refers to the number of levels of nestedness of parts within parts. A classic object hierarchy in biology is the nested series: species, populations, individuals, organs or organ systems, cells, and so on. Upper level entities physically contain the lower and constrain their behavior to some extent (Salthe, 1985). Such hierarchies have been called scalar (Salthe, 199³), ecological (Eldredge and Salthe, 1984), or structural hierarchies (O'Neil et al., 1986). In Figure 2, E and F have the same parts, but E is hierarchically more complex, because its parts are physically nested within each other to some extent.

For processes, hierarchical complexity is the number of levels in a causal specification hierarchy (Salthe, 199³). An army chain of command is such a hierarchy, with the highest

ranking officers issuing the most general orders, causing the lower ranks to give more specific orders. Likewise, development is partly a causal specification hierarchy, with the more general features of the bauplan specified early and initiating interactions leading to elaborations of those features (Riedl, 1978; Wimsatt, 1986; Arthur, 1988; Salthe, 199³~~4~~).

In Figure 2, G and H have the same parts, but in G the command structure is hierarchical, while in H it is not. (The dashed lines in H are just a reminder that, in organisms at least, developmental pathways may tend to converge to one or a small number of initiating events, and to the extent that this occurs, an ontogeny as a whole must be at least partly hierarchical. Still, at shorter time scales, interactions may proceed independently, in parallel, as shown in the figure.)

Notice that physical nestedness of objects in a system is not crucial to the hierarchical relationship of its interactions (as in G). In an army, commands of increasing specificity flow from the highest levels, generals, to the lowest, privates, but privates are not physically nested within generals. Nor is a one-to-many relation between early- and late-occurring interactions crucial. The increasing specificity of interactions (orders) along the chain of command would still be hierarchical even if every level were occupied by just a single individual.

B. Eight Types

The three distinctions -- differentiation versus configuration, object versus process, and hierarchical versus

non-hierarchical -- are independent, and thus together, in their various combinations, define eight independent senses, or types, of complexity for every system. So for differentiation, every system would have a: 1) hierarchical process complexity; 2) non-hierarchical process complexity; 3) hierarchical object complexity; and 4) non-hierarchical object complexity. The list would be the same for configurational complexity (bringing the total to eight).

The configurational types are not listed, because configurational complexity will not be addressed here. Assessing irregularity of parts and interactions is more difficult -- even intuitively -- than assessing their differentiation, and appropriate metrics have not been developed for the most part (cf. Yagil, 1985; McShea, 1992). In Figure 2, the interactions in D seem configurationally more complex than those in C, but it is difficult to specify why or how.

C. Difficulties

1. Complexity and functionality. The scheme outlined above is purely structural in that the complexity of a system is a function only of its parts, its physical structure, or of its interactions, its dynamical structure. In either case, complexity is independent of function. So, for example, a live organism and a dead one of the same species would have about the same object-complexity, if the number and configuration of their parts were about the same. In common usage, however, complexity is closely connected with functionality, and to many this structural view

will seem somewhat thin or narrow.

The complaint is reasonable. However, one problem with integrating some notion of functionality into the structural view is that compound concepts are awkward in practice. Heterogeneity is difficult enough to quantify in organisms, and workable measures of complexity -- even qualitative measures -- would be doubly difficult to devise if complexity were simultaneously a measure of heterogeneity and functionality. One virtue of a purer or narrower view of complexity is that measurement requires knowledge of structure, and nothing else.

More important for present purposes is the fact that function undoubtedly has a key role in the origination and maintenance of complex structure, and may have a significant effect on the probability of homoplasy in certain complex structures (see below). And in order to examine the relationship between structure and function, the two must be kept conceptually distinct. In other words, if functionality and structural heterogeneity were collapsed somehow into a single variable -- which we would then call complexity -- investigation of their relationship using this variable would be impossible.

2. Overall complexity. Is Figure 1B more complex than Figure 1A *overall*? The question may not have an answer, even in principle. The reason is that the eight types of complexity are conceptually independent. The types or aspects of other compound measures, such as size, have this same sort of independence: a balloon might be larger than a cannonball in volume but smaller

in mass. Analogously, a trilobite might have fewer parts than a human being, and thus lower object-complexity, but more interactions among parts, and thus greater process complexity. Given this independence, it is hard to imagine how a meaningful notion of overall complexity could be devised. Of course, this is not to deny that the various types may turn out to be related empirically. Just as volume may be correlated with mass (as it is for organisms), so object- and process-complexities may turn out to be correlated in some domain.

3. Complexity and randomness. Three kinds of randomness are relevant here. First, systems can be random in the sense that a group of systems of the same type all seem the same. One random number string looks like another; one compost heap looks like another. However, randomness in this sense is relative to interests and use. To a sea gull, every compost heap may look very different (R. Thomas, in preparation). The narrow view avoids this subjectivity. Each compost heap is considered unique and its complexity is the heterogeneity of its specific, unique structure, whether or not that structure is considered significant in some context.

Second, the narrow view might seem to equate complexity with entropy, another sort of randomness (Wicken, 1987). A compost heap is entropic in that a large number of different microstates (possible combinations and configurations of parts) correspond to the same macrostate (the same compost heap). However, calling a compost heap complex for this reason would be a category mistake.

Entropy is a relationship between microstates and macrostate, while complexity in the narrow view is a property of a microstate, of one specific composition and configuration, and involves no such relationship (Wicken, 1987).

Finally, some have argued that systems contain both a "regular" and a "random" component (Crutchfield and Young, 1989; Crutchfield, 1991) and that only the differentiation of the regular portion ought to constitute its complexity. Here, random refers to the unique features of systems, such as the precise number of hairs on the arm of a particular human individual, while regular refers to shared features, such as a five-fingered hand. The intent is to restrict complexity to features that are "rule-based," in other words, to features that are regular, or consistent among systems, either as a result of simple natural laws acting in the present or of irrevocable, contingent events in the past (frozen accidents) (Gell-Mann, 1994). This approach is consistent with the narrow view, in that identifying types of parts is equivalent to discerning first-order regularities. For example, a decision that two cells are the same type can be construed as a decision that their similarities are rule-based and their differences not.

The above view is also consistent with certain information-theoretic approaches, including the notion that the complexity of a system is the length of the shortest complete description of it (Löfgren, 1977), provided that only the regularities of the system are considered (Papert, 1980, 1982). This approach seems

to require that an organism be understood as a kind of message; then, its complexity would be just the length of the shortest symbol string that constitutes or conveys the message. Number of different parts is the number of different symbols required, and thus would be a kind of zeroth-order measure of complexity.

4. Dependence on descriptive frame. The number of parts in a system depends on scale, as discussed, but also on descriptive frame, in other words, on how we define a part at a given scale (Kampis and Csányi, 1987). The choice may seem arbitrary, which will make complexity measures seem arbitrary as well. One solution is to define parts functionally, based on the plausible notion that distinctiveness of parts in organisms is the evolutionary result of selection for functionality (Wagner, 1995). The problem is that identifying function is difficult, especially in cases where function has been lost.

Alternatively, parts could be defined as regularities, in the sense discussed. Elegant algorithms have been devised for discovering regularities in bit strings (Crutchfield, 1991), and in principle these could be applied to organisms. However, the algorithms make assumptions, such as statistical stationarity, which have not yet been justified for organisms. At present, our best strategy is probably to proceed intuitively, that is, to discover regularities and identify parts using our precognitive perceptual skills. We are obviously sensitive to many cues, such as morphological regularities among parts. We can articulate only some of them, but this does not make the partitioning arbitrary.

III. HOMOPLASY

The issue here is the relative probability, in qualitative terms, of homoplasy in complex evolutionary transformations of organismal development and morphology. Probability is evaluated for the four differentiatinal types of complexity. Treatment of the problem is analytic; this is mainly an attempt to sketch the logical implications for homoplasy of the conceptual scheme outlined above. As will be seen, the logic is not impervious to particulars. For example, a particular selection regime can render a probable evolutionary reversal improbable, and vice versa (see examples in Bull and Charnov, 1985). Some of these confounding particulars are discussed.

Before proceeding, a clarification may help prevent confusion later on: the issue is not the probability of transformations leading to homoplasy in complex organisms as opposed to simple organisms. The complexity of the organisms themselves is undoubtedly an important factor in homoplasy, but as will be seen, the logic developed here is relevant only to the complexity of homoplastic *transformations*.

A. Types of Homoplasy

Like complexity, homoplasy has a number of senses or types: reversal, parallelism, and convergence (Wake, 1991). Here I adopt (and adapt for present purposes) Hennig's understanding of these terms (Hennig, 1966; cf. Haas and Simpson, 1946): reversal is the return in a single lineage of a developmental trajectory or a morphological structure to its ancestral condition, as in the

transition series: $A \rightarrow B \rightarrow C \rightarrow B \rightarrow A$, where A, B, and C are morphologies or developmental trajectories in an ancestor-descendant series or lineage. Parallelism is the independent production, simultaneous or not, in two or more lineages of similar trajectories or morphologies from similar starting points by similar morphological or developmental routes: $A \rightarrow B \rightarrow C$ and $A \rightarrow B \rightarrow C$. Convergence is the production, again simultaneous or not, in two or more lineages of similarity from different starting points and by different routes: $A \rightarrow B \rightarrow C$ and $D \rightarrow E \rightarrow C$.

Only reversal and parallelism can be treated in the current framework. In these two, the same transformations are repeated, either in the same direction or the reverse. If we can assume for reversal that the forward and backward probabilities of a transformation are about the same (although see below), and for parallelism that probabilities are about the same in both lineages, then it is possible to say something a priori about joint probabilities of occurrence, at least qualitatively. In convergence, however, the two lineages undergo very different transformations and we have no information at all a priori about the probability of their arriving at the same place, morphologically or developmentally.

Morphological or developmental transformations can be classified as additions, deletions, or modifications of parts or interactions (O'Grady, 1985). Abstractly, an addition might be represented as: $A \rightarrow AB$, a deletion as: $AB \rightarrow A$, and a modification

as $A \rightarrow B$. Only additions and deletions can be treated in the present framework. For these, the complexity of the transformation can be understood as the number of parts (in morphology) or interactions (in development) added or lost. For modifications, the complexity of the transformation presumably lies at a lower hierarchical level, in the gears or inner workings, so to speak, of A in the modification $A \rightarrow B$. In a modification, these lower-level workings are not specified, and probabilities cannot be evaluated. Where they are specified, the problem may reduce to a case of addition or deletion at the lower level (although see Alberch, 1985), and the analysis can proceed at that level.

B. Relative Probabilities

For each of the four types of differentiability complexity, I will consider the relative difficulty or relative probability of reversal of addition, reversal of deletion, parallel addition, and parallel deletion. Table 1 summarizes the findings.

1. Hierarchical developmental complexity: reversal. Figure 3A shows an idealized process hierarchy in which the later interactions are caused by, and are dependent upon, the earlier interactions. The circles represent morphological features present at various stages in the development of an organism and the arrows represent developmental steps, perhaps inductions (but not necessarily). Imagine that in the ancestral condition in an evolving lineage, only the structure represented by the darkly

shaded circle is present, but in the course of evolution, steps (arrows) are added as shown, with additional intermediate and terminal structures resulting as well (lightly shaded circles). The suggestion has been made that the build-up of such hierarchical dependencies may characterize developmental evolution generally, at least at higher taxonomic levels (Riedl, 1978; Wimsatt, 1986; Arthur, 1988; Kauffman, 1993).

Reversal of such a complex addition is relatively easy and therefore probable. Removal of the first step nips the developmental flowering at the stem. Possible examples include cases in which large or prominent complex structures have been lost entirely (Arnold et al., 1989), such as the loss of the elaborate pluteal larval stage in some sea urchin species (Raff, 1987) and the loss of the entire adult stage in some salamanders (Wake, 1991).

Reversal of a deletion would be much more difficult. Imagine now an ancestor starting with a complex developmental trajectory such as that in Figure 3A, and suppose that this trajectory is lost in evolution, perhaps by the action of a suppressor of some sort acting on step 1. Restoring the ancestral condition might be as simple as restoring step 1, which could in principle be as simple as suppressing the suppressor (Laurent, 1983). On the other hand, if a significant amount of time has passed, other developmental modifications will have occurred, some of which may have altered the latent later steps and structures. In other