Complexity, Complex Systems, and Adaptation

F.T. ARECCHI

Department of Physics, University of Firenze, and Istituto Nazionale di Ottica, Firenze, Italy

1. INTRODUCTION

A wealth of speculations has recently appeared on complexity and complex systems. The first term has been differently defined in formal languages,1 computer science,2 and nonlinear signal analysis,3 starting in the early 1980s with the intrinsic nonpredictability associated with chaotic time series.4

Complexity is associated with epistemic processes. Once a time series of data, coded in a given alphabet, has been assigned, can one retrieve the meaning of the message just by perusal of that sequence? Is “meaning” just (i) discovering the grammatical rules that allow some symbol sequences (words) and forbid some other ones or also (ii) attributing a likelihood of occurrence to each word and hence attempting predictions about the future of the time series? This complexity approach has received different formulations, with different solutions leading to automatic procedures on complexity assignment.5-9

On the other hand, natural scientists have rather focussed on the fact that reality unveils lots of complex structures. Even before any encoding into some alphabet and a consequent mathematical elaboration, any holistic perception of an event implies many possible, not equivalent, ways of describing it. The possibility of many irreducible points of view can be considered as the token of a complex system, independently from any quantitative indicator.

Two different tasks are then faced by investigators of (A) complexity and (B) complex systems problems, namely,

(A) Given an input, what is the optimal use we can make of it? We call “certitude” the subjective confidence that we have done the best in grasping the inner rules of the input.

(B) Facing a piece of world, can we express our knowledge of it in a suitable language, that is, encode phenomena into symbol sequences from some alphabet (which later will be analyzed as in (A))? As we see, (B) is “prior” to (A). It appears as the problem that any living being has to solve, and it is usually faced by adaptive strategies, which we can later formalize as linguistic procedures, but which arise at a prelinguistic level and even determine the same choice of the most appropriate language.10 This more fundamental problem is that of “truth” defined as adaequatio intellectus et rei, that is, “adjustment of our expectations to the changing world.”

2. COMPLEXITY OF SYMBOLIC SEQUENCES

In computer science, we define as complexity of a word (symbol sequence) some indicator of the cost implied in generating that sequence. There is a "space" cost (length of the instruction stored in the computer memory) and a "time" cost (the CPU time for generating the final result out of some initial instruction).
of each resolution box, yielding some ambiguity in the assignment of an event to a specific cell of state space.

At any time slot of width $\tau$, we extract $e^{-D}$ different space data that we can encode in a suitable one dimensional string of symbols of an alphabet (e.g., binary). The modeller $O_2$ is inputted by some sequence $s$, and it sends an explanation that should enable $O_2$ to reconstruct an output $s' = s$. Notice that $M = M(D, e, \tau, \beta)$ is a whole class of possible instruments, and different individuals will give rise to different data sequences (different words). The explanation consists of a theoretical guess (model in) whose validity is tested by simulating an output and comparing it with the actual input data $s$. The difference yields an error signal $e$. Observer $O_2$ is provided with both $m$ and $e$ and it can reconstruct $s' = s$ upon this information.

The virtue of the expression $X$ is to have a bit length $|s| = |m| + |e|$ much shorter than the sequence length $|s|$ this amounts to extracting a relevant semantics out of the redundant features of $s$.

The explanatory machine is complex in so far as it spans over a whole class of models in. If one had access to a complete probabilistic description of the modeling universe, then the goal would be to maximize the probability of $m$ conditional to the input $s$

$$\Pr(m/s)$$

This ideal complete description is not available, but an approximation can be obtained by Bayes' rule

$$\Pr(s/m) = \frac{Pr(s/m)P(m)}{\sum_{m'} P(s/m')}$$

All these probabilities are conditioned on the choice of the model class. Furthermore, all terms on the right-hand side refer to a single data stream $s$. Here $\Pr(s/m)$ is the probability that a model $m$ produces the given data. With sufficient effort $\Pr(s/m)$ can be estimated. Finally, the normalization in the denominator depends only on the given data and so can be dropped as a constant.

The most likely explanation corresponds to the shortest code of length

$$|s| = -\log_2 \Pr(m/s)$$

There are two criteria for a good explanation:

(i) $x$ must explain $s$, that is, $O_2$ should resynthesize the original data $s' = s$

(ii) the bit length $|s| = |m| + |e|$ must be minimized.

The efficiency of an explanation is given by the compression ratio

$$C(m, s) = \frac{|s|}{|s|} = \frac{|m| + |e|}{|s|}$$

$C$ is a cost function. The optimal model minimizes this cost.

There are two limit cases. When the model is trivial ($|m| = 0$) the entire data are on the error channel: $|e| = |s|$. On the contrary a tautological model $m = s$ has no error: $|e| = 0$.

ARECHI: COMPLEXITY, COMPLEX SYSTEMS

The reconstruction of a state space out of an assigned time series and the assignment of transition probabilities among states is a task that can be faced in many ways. A suitable topological machine can be foreseen that amounts to a labeled graph $G = \{V, E\}$ made of vertices $v$ connected by labeled edges that assign a probability of going from vertex $v$ to $v'$ on observing a symbol $s$. Out of this labeled graph, suitable complexity indicators can be extracted.

According to the title of this section, here we have explored the complexity of a given symbolic sequence either from the computational point of view, aimed at reconstructing the single item, or from the probabilistic point of view, aimed at selecting a model within a class.

Preliminary to that, however, the problem arises of how we have obtained a given sequence, and this implies a critical analysis of the measurement apparatus. In the forthcoming section, we show how measuring apparatuses are suitable formalizations of the everyday knowledge expressed in the ordinary language, and we will provide an adaptive strategy to optimize the measurement performance in view of some specific goals.

3. FROM THE ORDINARY LANGUAGE TO THE SCIENTIFIC LANGUAGE

The word of the ordinary language is polysemic. In general it denotes a large variety of different situations, that we call "events," distributed on a "semantic space" (Fig. 1a). If the word is the name of an object, we usually do not mean the isolated object (which would be a mental abstraction) but the object embedded in different environments.

In a given language, the same word can be attributed with different degrees of appropriateness to different events (object plus environment). The different attributes have a different probability, as it emerges from a perusal of a historical dictionary providing for each word the frequencies of occurrence of different connotations in the literary texts of that language. In fact, the histogram is a finitistic approximation, a kind of coarse graining, due to the limited number of available texts.

If, however, we consider the everyday use, the continuous probability curve is more appropriate, since the environment includes the observer with his (her) own mood of the moment, hardly can be reduced to a countable number of states. Thus, whereas an artificial cognitive agent (a collection of detectors with fixed resolution feeding the input of a universal computer) would extract a histogram, thus justifying a finitistic approach, instead finitism seems excluded from the human everyday experience, as we reflect on the variety of nuances that qualifies a poem, or even a private conversation. Whence comes the problem of interpretation, that is, of what is the right meaning to be attributed to a word, within the wide support subtended by its probability distribution? In Indo-European languages, a quasi-univocal, or narrow range space of meaning $g$ is obtained by a "filtering" operation, which consists of supplying the word with a sufficient number of attributions or specifications, as sketched in Figure 1a. A discourse, seen as a flow of different words connected by grammar rules, appears as a wide riverbed within which everybody can cut out a different interpretation (Fig. 1b). As well known, there is no unique sense of a given
discourse, but one must refer to other sources of information, besides the text itself, in order to narrow the semantic range of each term.

Such ambiguities of the ordinary language were discussed at length by many Renaissance philologists and were well known to Galileo. He provided\(^{16}\) a way out of the ambiguities through his suggestion of “naming” an event via the number extracted by a suitable measuring apparatus \(M\) applied to the event itself (FIG. 2a), that is, limiting himself to quantitative “affections” rather than attempting to grasp the “essence.” This procedure apparently provided univocal meanings, since it filtered out a single denotation, clearing away all the context. As an example, a physicist does not speak of a “table” but of the “weight” or the “length” of the table. In this new language, the syntactical rule connecting two words becomes a mathematical relation connecting the output numbers from two measuring apparatuses related to two “objects.”

This provides a solid framework for any scientific description, in terms of well-established existence uniqueness theorems. Thus, the flow of scientific discourse consists of sharp, necessary connections among pointlike objects of different semantic spaces, corresponding to different measurements as shown by the solid line in FIGURE 2b. That solid line seems to be a great progress compared to the wide riverbed of FIGURE 1b. It means that the scientific language is free from interpretational ambiguities. The most crucial aspect of Galileo’s self limitation is the apparent arbitrariness in placing \(M\) over the semantic space, that is, the large number of dif-

---

**FIGURE 1.** (a) Polysemantic word represented as a probability curve over a semantic space. Constraining the word by further attributions narrows its semantic range, that is, reduces the probability spread on the semantic space (“oligo-semantic” word). (b) A discourse, as the connection of different semantic spaces, yields a wider riverbed, allowing for different interpretations. (From the point of view of a linguist, this sketch is a rough caricature, since it reduces the semantic narrowing to a precontextual session).

**FIGURE 2.** (a) A fixed measuring apparatus selects a univocal denotation, attributing to the event a membership property to a suitable set. (b) A scientific discourse is a chain of necessary connections between univocal terms, thus leading to a unique pattern.
different univocal words extractable for a class of events usually denoted by a single word in an ordinary language. This proliferation of MS has avoided for centuries the question of complexity. Suppose, for instance, (Fig. 3) that by M1 we look at the cell dynamics. Thereby, we build the specific science of “cytology” with its own words. We may now realize that many cells form an organ, but the organ is observed by a different M2, which provides different words and hence a different science: “physiology.” Similarly, if we look at the biomolecules we build a “biochemistry.”

We have thus obeyed Occam’s razor (Entia non sunt multiplicanda praeter necessitatem) in a economic way, that is, changing language whenever it is no longer appropriate. A more strict obedience would consist in a reductionistic approach. As shown in Figure 4, we can build a hierarchy from large to small and say that the behavior of smaller objects should determine that of larger ones. But here a perverse thing, already hinted by Anderson, occurs. If our words were a global description of the object in any situation (as the philosophical “essences” in Galileo’s letter) then, of course, knowledge of elementary particles would be sufficient to make predictions on animals and society. In fact Galileo’s self-limitation, to some “affections” is sufficient for a limited description of the event, but only from a narrow point of view. Even though we believe that humans are made of atoms, the affections that we measure in atomic physics are insufficient to make predictions on human behavior.

The fact that higher levels in the ladder of Figure 4 display features not predictable from the lower ones is what colloquially we call complexity. This way complexity is not a property of things (like being red or hot), but it is a relation with our status of knowledge, and for modern science it emerges from Galileo’s self-limitation.

A theory is successful if it is a “compressed” description of the world, that is, if the length of its initial assumption is much shorter than a detailed description of the events themselves. At the start, a physical theory is just mathematics. It becomes a model, that is, it acquires semantic values, whenever we interpret the objects of the theory as elements of reality. Therefore, a scientific theory must be considered as a set of primitive concepts (defined by suitable measuring apparatuses such as M of Figure 2a) related by axioms. The deduction of all possible consequences (theorems) provides predictions that have to be compared with the observations. If the observations falsify the expectations, then one tries with different axioms. The deductive process is affected by a Gödel undecidability like any formal theory, in the sense that it should be possible to build a well-formed statement, but the rules of deduction are unable to decide whether that statement is true or false. Besides that, a second drawback is represented by intractability, that is, by the exponential increase of possible outcomes among which we have to select the final state of a dynamic evolution. Figure 5a sketches a bifurcation tree well familiar to computer scientists, because one has to perform a complex calculation with branch points implying multiple choices of the type “if–then.” I rather consider it to be the bifurcation tree of a complex nonlinear dynamics, as one changes a suitable control parameter α. Going back to the reductionistic tentative of explaining reality out of
its constituents, then we find an exponentially high number of possible outcomes, when only one is in fact that observed. This means that, while the theory, that is the syntax, would give equal probability to all branches of the tree, in reality we observe an organization process, whereby only one final state has a high probability of occurrence. It is here necessary to recall some descriptive elements on the bifurcation of the stable branches of a dynamics for different settings of a control parameter. These are the necessary ingredients of any complex dynamics. Notice that dynamical bifurcations in a system of interactive identical particles display specific symmetries (Fig. 6a). Only external gradients break this symmetry (Fig. 6b). Thus, during the course of a dynamical evolution, either because some control parameters \( \{ \alpha \} \) are tuned from the outside to assume different values or because internal feedbacks change the \( \{ \alpha \} \) set in course of time, starting from some initial conditions we expect an exponential increase of final states. Whenever there has been organization, this means that at each bifurcation vertex of Figure 5 the symmetry was broken by an agent external to the system under investigation. We can thus stipulate the following things:

(i) A set of control parameters

\[ \alpha_1, \alpha_2, \ldots, \alpha_N = \{ \alpha \} \]

is responsible for successive bifurcations leading to an exponentially high number (in the example, of the order of \( 2^N \)) of final outcomes. If the system has no boundary effects (considered of infinite size), then all outcomes have comparable probabilities, and we call complexity the impossibility of predicting which one is the state we will observe at the end of the chain of bifurcations.

(ii) A set of external forces

\[ A_1, A_2, \ldots, A_N = \{ A \} \]

applied at each bifurcation point breaks the symmetries, biasing toward a specific choice and eventually leading to a unique final state.

We are in the presence of a conflict between (i) syntax represented by the set of rules (axioms) \( \{ \alpha \} \) and (ii) semantics represented by the intervening \( g \) external agents \( \{ \alpha \} \). The syntax provides \( 2^N \) legal outcomes. But if the system is open to the
external world, the presence of which is expressed by \( A \), then it organizes to a unique final outcome. Once the syntax \( \{ \alpha \} \) is known, evidence of an unique final result implies that the set of external events \( \{ A \} \) must have occurred. Therefore we can take \( \{ A \} \) as the element of reality in which our system is embedded. We define "certitude" the correct application of the rules \( \{ \alpha \} \) and "truth" as the combination \( \{ \alpha, A \} \) of those a priori rules with external influences \( \{ A \} \) that perform the choices. However, the same final outcome would be reached by a different set of rules \( \{ \beta \} \). In such a case, retracing back the new tree of bifurcations, we would reconstruct a set \( \{ \beta \} \) of external agents. Thus, it seems that truth, \( \{ \alpha, A \} \) or \( \{ \beta, B \} \), is language dependent. Furthermore, the "emergence" of organization means that we can even build a set of axioms \( \{ \varepsilon \} \) that succeeds in predicting the correct final state without external perturbations, that is, \( \{ \varepsilon \} = \emptyset \) (Fig. 7). This is indeed the pretension of the so-called "autopoiesis," or "self-organization," to which I have opposed the term "heteroorganization."^20

From a cognitive point of view, the theory \( \{ \varepsilon \} \) can be reputed to be a "petitio principii," a tricky formulation tailored for a specific situation and not applicable to slightly different situations. Rather than explicitly listing the elements of reality, as for example \( \{ A \} \) \( \{ \alpha \} \), the use of language \( \{ \varepsilon \} \) has already exploited at a preformalized level the elements of reality, and has made good use of them in planing the axioms \( \{ \varepsilon \} \). From a cognitive point of view, an ad hoc model may be appropriate for a specific situation, but in general it lacks sufficient breadth to be considered as a general theory. However, in describing the adaptive strategy of a living species, or a community, or so forth, a "self-organization" may be the most successful action. In other words, once the environmental influences have been known, better to incorporate this knowledge in the model from the beginning, this assuming the fast convergence to a given goal. These pieces of knowledge which precede axiomatization have received different names, such as "abduction"^21 or "tacit dimension."^22 Some of them have been memorized as universal tools either in our genetic heritage or dur-

![Figure 7](attachment:image7.png)

**FIGURE 7.** Different theoretical models may explain the same final state. The backward reconstruction of the dynamical history will then retrieve different classes of external agents.
ing \( M \) means changing the "point of view" under which we observe the world and hence making a different theoretical model.

In the previous section, we show an indeterminacy in the reconstruction of the elements of reality \([A, B, C, \ldots]\) that modify the dynamical \([\alpha, \beta, \gamma, \ldots]\) of a mentally isolated system. As a result, the truth is represented by a combination of a model developed for the subsystem, plus the external gradients, or boundary conditions, once the subsystem is embedded in a suitable environment.

Which one is the most appropriate among the pairs

\[ \{a, A\}, \{\beta, B\}, \{\gamma, C\}, \text{ etc.} \]

The question is equivalent to asking: Among all possible measuring apparatuses \( M \) applicable to an event, which one is the most appropriate? If we prefer not to decide, we independently make use of different \( M \)s and correspondingly define different sciences (Fig. 3).

In real life, we have to face problems overlapping different separate sciences. For instance, a cardiac disease may be due to a global offset of the pacemaker or to some local cytopatology or even to a drug effect acting at a biomolecular level.

Figure 8 shows the difference between fixed and adaptive \( M \). In the first case we sharply define three separate sciences. What can be exchanged among different specialists is not technical words, which are specific to each science, but just the residual metaphorical part, not filtered into the technical word. Should we say that two scientists of different areas always communicate by metaphors? A tentative way out (fuzzy logic) is to avoid sharp definitions, so that different disciplinary terms have regions of overlap.

The most natural approach, however, seems to start with \( M \) at very low resolution, covering all the disciplinary areas, and then—depending on a specific problem—to zoom toward one of the other narrow points of view. A successful approach of adaptive measurement has been worked out in my group.\(^{11,12}\) It consists in developing a measuring procedure \( M \) whereby we observe a system only at "almost equal" geometric separations in state space. As a consequence, \( M \) is activated only for short time intervals at irregularly distributed times, separated from each other by unequal intervals. The stroboscopic sequence of these intervals has an information content that provides fast, reliable answers to the following questions:

1. recognizing chaotic dynamics (Lyapunov exponents, different unstable periodic orbits (UPOs));
2. discriminating determinism from stochastic noise;
3. controlling chaos, that is, stabilizing one of the UPOs contained within a chaotic attractor.

When applied to the control of chaos,\(^{12}\) this adaptive algorithm is effective for values of the stroboscopic times much larger than the Runge-Kutta integration steps and smaller that the periods of all UPOs. In other words, the method introduces a natural adaptation time scale that is intermediate between the minimum resolution time of the dynamics and the time scale of the periodic orbits.

The effectiveness of this adaptive method is rooted upon a drastic simplification. Rather than looking at all degrees of freedom of a \( D \)-dimensional object, that is, to all the information contained in the invariant set where the motion is confined, this approach extracts only the information contained in a one-dimensional string of data, namely, the stroboscopic series of observation time intervals \([T]\). A lesson seems to emerge from this example, that is, any successful way of approaching the world relies on reducing the complexity by limiting oneself to a simplified description. This seems peculiar of any scientific description that is by no means holistic, but, sharply confined to a particular point of view.

6. CONCLUSION

In this conclusion we discuss the truth value of scientific statements on the basis of the considerations of the previous sections. In the scientific investigation, we select a quantitative feature by application of an apparatus \( M \) at a particular point of the semantic space. The emerging description of reality represents an observation "from one point of view." Due, however, to the variety of possible \( M \)s, we must justify at a metascientific level why we have selected that \( M \) rather than another one. This is a general question dealing with the role of those elements of reality which are preliminary to one particular program. In Section 5 we have seen that adaptation is a slalom among different sets or rules \([\alpha], [\beta], \ldots\), under the guidance of a preferential set of external elements that has nonempty intersections with \([A], [B], \ldots\), but does not coincide with either of them.

In the face of the truth problem we can take three attitudes, namely,

(i) Assume the adaptive strategy and its associated reality set as a kind of privileged reference frame. Indeed, being the result of an optimization process, it appears more appropriate than any particular theory, \([\alpha]\) or \([\beta]\).

(ii) Consider the truth problem as a metatheoretical problem. At this metalevel, the set of all sets of truth values \([A], [B], \ldots\) has to be considered as the truth, but, with the stipulation that any individual set makes sense only if associated with the corresponding theory.

(iii) A more fundamental approach recovers the polysemy of the ordinary language as a virtue, not a drawback. More than questioning the power of any specific theory we put into question the same set-theoretical approach to the fundamental concepts of the physical description. Going back to Figure 2, we have seen that \( M \) provides a sharp distinction that allows to classify any observed entity within all appropriate sets. Hence the set-theoretical character of all modern sciences, with the consequent antinomies of modern logic after Cantor, Russel, Gödel, and so forth transferred into the heart of the scientific language.

An adaptive \( M \) means that the localization in semantic space is no longer as sharp as whenever it is defined by a precise stipulation as for the \( sets\).

This degree of smoothness seems to me to be going back to the polysemy of ordinary language. Hence Epimenides Creton says, "all Cretons are liars" is no longer an antinomy, since Epimenides is not bound to be always a liar, but a liar in general, even though sometimes even he can tell the truth!
Going back to the title, the truth versus certitude issue can be summarized by the following scheme (Fig. 9). \( R \) stands for reality (whatever this means), \( S \) for a symbol interpreter (an intelligent being or even a Turing machine!), and \( M \) is the measuring apparatus.

In modern science, \( M \) is usually not questioned, and the elaboration takes place on the output of \( M \). This was called as the (A) complexity approach, leading eventually to certitude, when the \( S \) machine has optimized the explanation necessary to retrieve the \( M \) output sequence, according to the discussion of Section 2. From a gnoseological point of view, if the \( M \) box corresponds to our senses, as suggested by Hume, then \( S \) (called by Descartes res cogitans) has to face not the world, but the representation already coded by \( M \), and this is a grammatical problem solvable by a machine. This means that Descartes’ mind is equivalent to a Turing machine, as already suspected by many experts of artificial intelligence. This strong association of \( M \) on the side of \( R \) is equivalent to what Atmanspacher called “the Cartesian cut.”

On the contrary, the (B) complex system approach regards world’s knowledge as an endeavor globally faced by the observer \((S + M)\) through an adaptive procedure \( M \), for which, however, a linguistic foundation does not exist, because any linguistic formulation is subsequent to the operation of \( M \). During the scientific operation, \((S + M)\) acts in an entangled way. A meta-level of investigation (psychology of cognition) is required to disentangle \( S \) from \( M \). In summary, there is a nonlinguistic residue in the scientific operation which then precludes a Turing machine from acting as a creative scientist.

**ADDENDUM**

The connection from \( M \) to \( S \) in Figure 9, rather than being considered as unidirectional, must be taken as a feedback loop that provides sensory messages \( M S \) renormalized by the mental expectations \( S M \), and not just bare sensations, the way they would be tested in a laboratory session of a behaviorist psychology department.

A hot issue debated by early sociologists (Marx, Durkheim) is whether science means knowing the world or trying to modify it. In fact, if the link from \( R \) to \( M \) is also considered a feedback loop, that would mean that reality \( R \) is continuously modified during the interaction with the observer \((S + M)\), and hence only the global \((R + M + S)\) entangled system should be considered. This would happen if, for instance, a smart theoretician had discovered the dynamical rules of the stock market and, rather than writing a paper on it, he/she tries to take advantage of that knowledge by playing in the market. As a consequence, the \( R \) situation just observed gets modified, and one must give up on an “objective” (i.e., observer-independent) description. In fact, this seems to be the general strategy of living species or communities, seen as adaptive systems.

Finally, the above-mentioned entanglement problems play a role in microscopic quantum mechanics, but these are still open problems, which represent an intellectual challenge, even though decoherence due to the environment destroys the problems in most practical cases.

**REFERENCES**