

On the Origins of Information and Its Relevance for Biological Complexity

Kepa Ruiz-Mirazo and Alvaro Moreno

Department of Logic and Philosophy of Science
University of the Basque Country
San Sebastian-Donostia, Spain
kepa_ruiz@ehu.es
alvaro.moreno@ehu.es

Information as a Naturalized Organizational Property

In Shannon and Weaver's (SW) theory (1949), the idea of information derives from fixing or selecting a set of discrete states (a "sequence") among other possible sets. States are allowed to change and be combined according to certain *syntactic* rules in a way that is independent of their specific material basis. Under those conditions, (the amount of) information is defined in terms of the probability of appearance of a given sequence in relation to the total number of possible sequences that could, in principle, appear (the lower probability, the higher informational content of a sequence). SW theory analyzes why a particular set of discrete states constitutes a "message" and not "noise" but does not address what is behind this fundamental distinction. Similarly, it includes (internally) not well-definable distinctions or constructs like "source," "encoder," "decoder," "receiver," etc., which implicitly subsume the processes of generation, interpretation, and function of information. In brief, it is a theory that provides a rigorous operational measure of information by focusing on its formal side and disregarding, to a large extent, its *semantics*.

Although this approach may be very convenient from the point of view of a communication engineer, it will never be satisfactory from a naturalizing perspective, because it presupposes the existence of (intelligent) agents that produce, understand, and use information. Therefore, in order to discuss the significance of this concept in biology we need a different framework that, being fully congruent with SW formal requirements, is also capable of encompassing other major connotations of the term. This unavoidably demands the introduction and explanation of semantic constructs, which were left implicit in the SW account and are fundamental to appreciate

the organizational nature of the concept. But, above all, it requires building a solid theoretical bridge from a world in which physics and chemistry (with their own primitive concepts) are sufficient to account for natural phenomena to a world that cannot do without discrete sequences and meaning.¹

Generally the term "information" is used when something has a referential capacity, regardless of its physical properties. To say that A is a *message* whose referent is B implies that B is causally connected with A through its *form*² within a specific *interpretation* context. Thus, the idea of information immediately brings about the need for a formal type of causality, which is at odds with the traditional approach of causality in physics, where strictly causal connections are those fully governed by the fundamental laws of interaction and, thus, depend on the specific values of mass, charge, energy, speed, etc., of the elements involved. On the contrary, the idea of "formal causation" implies (i) arbitrary/conventional relations among inert interchangeable elements, which can be subsequently rearranged according to particular rules (*compositionality*), and which (ii) have causal effects only by virtue of their form (*sequential order*). Information entails, therefore, the creation of a new domain of processes and structures causally linked through formal (rule-based) relations, which emerges out of the fundamental physical world where relations are based on the intrinsic properties of matter (Moreno 1998).

The constitution of a, so to speak, "syntactic domain" cannot take place without an underlying "dynamic or physical domain" (which provides, at least, part of the necessary conditions, like the equivalent—energetically degenerate—states required for compositionality) and, in turn, has important consequences on it. In fact, the objects to which syntactic structures/sequences refer belong to the physical domain, and their causal effects also pertain to it. So information requires a special type of material structures (discrete, inert, reconfigurable components) where it is "recorded" and through which it can be formally "processed," but also a dynamic domain (constituted by other functional components) where it is, somehow, "expressed." This is mainly why the idea has an inherent organizational character, being grounded on the complex

integration of those two domains, which are simultaneously *decoupled* and *interdependent*. Such a complementary relationship is possible because the decoupling occurs from the point of view of the dynamics of the system, whereas the link is established through higher-level connections or constraints (which are largely independent of physical causal relations).

The question that obviously arises at this point is the following: if the link between informational “records” and their “referents” (or “effects”) is, in material-energetic terms, contingent, how can it actually be established and maintained? In the case of artificial informational systems, the link is clearly enforced by human beings and their own functional needs. But this is not a way to solve the problem of the origin of information, or of its naturalization. Instead, we would face an infinite regress (since human mind ultimately requires the concept of information). Therefore, the first condition to advance in our understanding of information as a natural phenomenon is to conceive it as a self-maintaining system of relations, as an *organization* whose functioning cannot be explained by resorting to any intelligent external agent. And the second is to examine carefully the implications of the idea in causal terms, i.e., as an organizing principle/mechanism that radically changes the functional-operational structure of the system where it is realized.

What kind of material-dynamic organization would, then, be a good candidate to become informational? In order to make the naturalization project consistent and search really for the minimal forms of organization that could be informational, one should start from a physicochemical scenario where complex (far from equilibrium) reaction networks achieve a considerable degree of autonomy in their component production dynamics. The development of those “proto-metabolic” networks would be crucial to solving basic problems regarding the material-energetic implementation of a minimally robust, self-maintaining organization, on top of which informational mechanisms may later appear and be anchored.³ In other words, the naturalization of information leads to a prerequisite: the appearance of some autonomous form of organization, without which it makes no sense to speak of information in the first place. But what does it mean operationally that an autonomous organization be informational? What are the functional and evolutionary consequences of becoming informationally organized for previous or simpler versions of autonomy?

What Do Autonomous Organizations Need Information For?

We can imagine that chemical evolution led to the appearance of proto-metabolic systems, capable of managing the flow of matter and energy through them in order to achieve a self-constructive dynamics. However, the organization of these “basic autonomous systems” (Ruiz-Mirazo et al. 2004)

does not solve the problem of preserving (for long periods) the new complexity that it can generate, and therefore, nor does it solve the problem of how this complexity could grow indefinitely. Of course, the productive—and possibly also reproductive—dynamics of basic autonomous systems would allow the maintenance of a certain level of complexity. But that—still rudimentary—functional dynamics cannot ensure that the components (together with the way they are organized) remain unaltered for much longer than their typical lifetimes (or that of the organization) and the system faces a bottleneck: as the complexity in it rises, its preservation becomes more and more difficult.

Now, how to overcome that initial threshold of complexity? As Szathmáry and Maynard-Smith (1997) have pointed out, the way to preserve the specificity of an increasingly complex organization is through what they call a mechanism of “unlimited heredity.” We understand this mechanism as a link between the sequential structure of certain stable and self-replicating components (“modular templates”) and the most complex functional properties of the system, which makes possible the “storage” or “recording” of those functions (and increasingly complex/specific ones). This, in turn, makes possible—through the controlled replication of the templates—a more reliable way of reproducing the whole system. Interestingly, the introduction of modular templates linked to the most complex part of the functional organization can also be a way to achieve system robustness, for a highly distributed and holistic organization would be rather fragile—especially in evolutionary terms (Simon 1973).

Presumably, the organization of the first systems endowed with modular templates was based on a single type of polymer, supporting at the same time replication and catalytic functions (a kind of “RNA world” embedded in cellular proto-metabolisms). However, this cannot generate an open-ended increase in complexity, because there is a trade-off between efficient catalytic and storage/replication functions (Benner 1999). Accordingly, not until the organization is based on (at least) two different types of polymers, specialized respectively in template and catalytic tasks, do these systems become free to develop unlimited versatility and complexity (Ruiz-Mirazo et al. 2004).

The new form of organization is thus based on two quite different, but complementary, forms of operation: one involved in the fundamental productive-metabolic processes (i.e., dynamic, “rate-dependent” processes), and the other (genetic, “rate-independent” processes), partly decoupled—or “off-line”—from all that muddle of chemical reactions, being responsible for the storage and long-term preservation of the functional properties of the system (Pattee 1977). The “in-forming” action of genetic components is not directly determined by the physical properties of the molecules involved, but can only take place with the contribution of

the functional-metabolic organization on which they operate. Hence the adequacy of speaking in terms of a *translation* or *interpretation* code, since the causal connection between those two different operational modes can be established only in an indirect way only, by means of a whole set of metabolic intermediates and nonlinear processes. So, despite being dynamically decoupled, both modes of operation in the system are deeply intertwined and none of them can work without the other. Such a circular, code mediated, interrelation between two decoupled levels in the system (genotype and phenotype) is the core of a naturalized view on the problem of “symbol interpretation” and constitutes (the minimal form of) what Pattee (1982) considered the universal organization of an informational system.

Final Remarks

Information brings about, on the one hand, an increase in the robustness of the processes of self-maintenance of necessarily previous autonomous organizations and, on the other hand, the potential for an evolutionary growth in their complexity. Genetic records, shaped through a collective and historical process (natural selection), make possible the rearrangement of subsets of sequential structures so that highly organized systems are generated and preserved. Ultimately, the decoupling between genetic material and metabolic dynamics is the expression of the insertion of organisms, as autonomous systems, into a historical-collective network (an *ecosystem*) where slower processes of creation and modification of informational patterns take place, and where an additional circular relation of cause and effect is established between individual organizations and the eco-evolutionary global dimension. The origin of information takes place precisely when the link between both dimensions is articulated, giving way to an open-ended process of production and evolution of complexity in nature.

Acknowledgments

KRM holds a postdoctoral fellowship from the Basque Government. Funding was also provided by the Spanish Department of Education, grants BFF 2002-03294 and BMC2003-06957, and by the University of the Basque Country grant 9/upv 00003.230-15840/2004.

Notes

1. In contrast with the strong generalizing tendencies of information theorists (who try to extend the concept to the realm of physics) or even with the followers of the “(bio-)semiotic” school (who ultimately consider “signs” as a kind of physical primitive, too), we defend a position that reserves the concept of information for explaining the transition toward new phenomenological domains in nature.
2. By *form* we do not mean here a geometrical shape (like the spatial configuration or “folding” that specifies the catalytic action of an enzyme) but an abstract, purely sequential order, which can be rewritten, through encoding, in many different, still equivalent ways.
3. There is an alternative and, in fact, more widely accepted scenario for the origins of information, founded on the dynamics of populations of self-replicating molecular species. However, in so far as this approach does not include a proper “*system* organization” (i) through which material-energetic resources are channeled to achieve a realistic autonomous implementation and (ii) on which the informational content of molecular sequences can be expressed in functionally diverse enough ways, it will not provide a solid basis to explain the emergence of a minimal informational system.

References

- Benner S (1999) How small can a microorganism be? In: *Size Limits of Very Small Microorganisms (Proceedings of a Workshop)*, 126–135. Washington, DC: National Academy Press.
- Moreno A (1998) Information, causality and self-reference in natural and artificial systems. In: *Computing Anticipatory Systems* (Dubois DM, ed), 202–206. Woodbury, NY: American Institute of Physics.
- Pattee HH (1977) Dynamic and linguistic modes of complex systems. *International Journal of General Systems* 3: 259–266.
- Pattee HH (1982) Cell psychology: An evolutionary approach to the symbol-matter problem. *Cognition and Brain Theory* 4: 325–341.
- Ruiz-Mirazo K, Peretó J, Moreno A (2004) A universal definition of life: Autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere* 34: 323–346.
- Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication*. Urbana: University of Urbana Press.
- Simon HA (1973) The organization of complex systems. In: *Hierarchy Theory: The Challenge of Complex Systems* (Pattee HH, ed), 1–27. New York: Braziller.
- Szathmáry E, Maynard Smith J (1997) From replicators to reproducers: The first major transitions leading to life. *Journal of Theoretical Biology* 187: 555–571.