8th Altenberg Workshop in Theoretical Biology 2002

BIOLOGICAL INFORMATION BEYOND METAPHOR:
Causality, Explanation, and Unification

11-14 July 2002

organized by Werner Callebaut and Sahotra Sarkar

Konrad Lorenz Institute
for Evolution and Cognition Research
Altenberg, Austria
&
Program in the History and Philosophy of Science
University of Texas at Austin

The topic

The emergence of the new biological subdisciplines of biocomplexity and bioinformatics at the end of the twentieth century marks the continuing attempt to provide a new potentially unifying framework for biology based on concepts of information. Informational thinking first entered biology through neurobiology in the 1920s; it entered genetics, developmental biology and evolution in the 1950s. It was central to the formation of molecular biology as a discipline with an epistemological framework identifiably different from predecessors such as biochemistry. "Information" lies at the basis of the views of morphogenesis based on the idea of a genetic or developmental program. In recent years even ecological complexity has been characterized informationally. With the completion of several genome sequences "information" has come to have a dual aspect: following the long-standing practice of molecular biology, DNA sequences have been interpreted as carriers of biological information; and new techniques have had to be developed to attempt to make sense of the astronomical quantities of information (qua sequences as strings of symbols) that have emerged from sequencing projects.

In spite of such ubiquitous use of informational thinking in biology, the concepts of information that have been invoked have remained murky, and their use controversial. In recent years some theoretical biologists and philosophers have challenged the informational framework as being no more than a metaphor that masquerades as a theoretical entity; others have argued even more strongly for the centrality of information in biology. This workshop is designed to explore these issues, as they are formulated today, as well as how they have been articulated in different contexts during the last century.

This workshop will bring together biologists with philosophers of science who have specialized in the conceptual and mathematical foundations of biology.

Program

CONCEPTUAL FOUNDATIONS

Werner Callebaut Information, Causation, Explanation, and Unification
William F. Harms Four Concepts of Information
John Collier Information Expression in Biological Hierarchies

MOLECULAR INFORMATION THEORY

Hans-Jörg Rheinberger Notions of Language, Information and Regulation in the Writings of François Jacob
Antoine Danchin Living Turing Machines
Thomas D. Schneider Twenty Years of Delila and Molecular Information Theory

NEUROSCIENCE AND ANIMAL COMMUNICATION

Justin Garson The Introduction of Information into Neurobiology
Jonathan D. Victor Approaches to the Analysis of Information in Neural Activity
Jessica Pfeifer Applying Information Theory to Animal Communication: Some Lessons From Social Insects

MECHANISMS AND PHENOTYPIC EVOLUTION

Lindley Darden Flow of Information in Molecular Biological Mechanisms
Peter Stadler Genotype-Phenotype Maps
Eörs Szathmáry Information and Replication

Abstracts

WERNER CALLEBAUT
werner.callebaut@kli.ac.at
Konrad Lorenz Institute for Evolution and Cognition Research
Adolf-Lorenz-Gasse 2
A.3422 Altenberg
Austria
phone: +43 2242 32 390 18
fax: +43 2242 32 390 4

Information, Causation, Explanation, and Unification

Theoretical biologists, although notoriously disagreeing on the issue of the autonomy vs. reducibility of the biological sciences, have mostly taken some variety of the early twentieth-century idea of a unified science for granted. The two major contemporary philosophical accounts of scientific explanation are the causal-mechanistic view and the view, which may but mustn't be cast in terms of Hempel-type subsumption, that theoretical unification is the (main) function of explanation. They are usually understood as complementing rather than conflicting with each other. Concepts of information and the theories in which they are embedded have often been understood as powerful unificatory tools for biology and science in general. In the paper I attempt to systematize the interrelations between present-day views of causation and explanation, and their bearing on our view of information, against the background of recent criticisms of the unity of science "ideology" and a consistently naturalistic philosophical approach to science.

JOHN D. COLLIER
john.collier@kla.univie.ac.at
Konrad Lorenz Institute for Evolution and Cognition Research
Adolf-Lorenz-Gasse 2
A.3422 Altenberg
Austria
phone: +43 2242 32 390 19
fax: +43 2242 32 390 4

Information Expression In Biological Hierarchies

The term 'information' is used in multiply ambiguous ways. Information carrying capacity is the consequence of a system having certain causal (dynamical) properties. These properties can be defined at a single scalar level, allowing but not implying arbitrarily high degrees of specifiability. Real specifiability requires real alternatives. This is possible for a system that has the properties of what I call a physical information system. The expression of information, however, requires at least a second dynamically defined level. This is sufficient for at least a form of protoreference, but, I will argue, not a real pragmatics, and hence not a real semantics. A physical information system is composed of a number of relatively stable units that can combine more or less freely to form somewhat less stable structures has a capacity to carry information in a more or less arbitrary way. All physical information systems have certain general dynamical properties. DNA can form such a system, but so can, to a lesser degree, RNA, proteins, cells and cellular subsystems, various immune system elements, organisms in populations and in ecosystems, as well as other higher-level phenomena. These systems are related hierarchically with respect to the expression of lower level information at higher levels. This allows a distinction between macro and microstates within the system, with entropy driven dynamics, including the possibility of self-organization, system bifurcation, and the formation of higher levels of information expression. Although lower-level information is expressed in an information hierarchy, this in itself is not sufficient for reference, function, or meaning. Nonetheless, the expression of information is central to the realization of all of these. 'Biological information' is thus ambiguous

between syntactic information in a hierarchical modular system, and functional information. However, the dynamics of hierarchical physical information systems is of interest to the study of how functional information might be embodied physically.

ANTOINE DANCHIN
adanchin@hkucc.hku.hk
HKU-Pasteur Research Centre
1/F Dexter H.C. Man Building
8 Sassoon Road
Pokfulam, Hong Kong
phone: +852 2816 8414
fax: +852 2168 4427

Living Turing Machines

Despite their origin as technical or political programs in the majority of cases, some genome programs can be associated with a deep investigation of what life is. As an example, this goal was explicitly at the core of the setting up of the Bacillus subtilis genome program. The underlying concept is that the alphabet metaphor that lies at the basis of the study of genomes is highly adequate to account for many of the most surprising properties of life. This metaphor, which can be deeply rooted in Number Theory, carries with it a reassessment of the concept of information, and this will be the topic of our discussion. Naturally we cannot be highly technical here, and we shall take the view of the experimental biologist to reinvestigate the nature of the concept.

LINDLEY DARDEN
darden@carnap.umd.edu
Department of Philosophy
University of Maryland
1125 A Skinner Building
College Park, MD 20742
USA
phone: +1 301 405 5699
fax: +1 301 405 5690

Flow of Information in Molecular Biological Mechanisms

In 1958, Francis Crick distinguished the flow of information from the flow of matter and the flow of energy in the mechanism of protein synthesis. Crick's claims about information flow and coding in molecular biology are viewed from the perspective of a new characterization of mechanisms (Machamer, Darden, Craver, Philosophy of Science 2000) and from the perspective of information as holding a key to distinguishing work in molecular biology from that of biochemistry in the 1950s-70s (Darden and Craver, Studies in History and Philosophy of Biological and Biomedical Sciences 2002). The flow of information and coding in molecular biological mechanisms are to be distinguished, on the one hand, from formal information theory, and, on the other, from information as used in cognitive neuroscience, where information and representation are often coupled.

JUSTIN GARSON
j.garson@mail.utexas.edu
Program in the History and Philosophy of Science
University of Texas at Austin
Waggener Hall 316
Austin, TX 78712-1180

USA
phone: +1 512 232 7101
fax: +1 512 471 4806

The Introduction of Information into Neurobiology

The first use of the term “information” to describe the content of nervous impulse occurs 20 years prior to Shannon’s (1948) work, in Edgar Adrian’s The Basis of Sensation (1928). Although, at least throughout the 1920s and early 30s, the term “information” does not appear in Adrian's scientific writings to describe the content of nervous impulse, the notion that the structure of nervous impulse constitutes a type of message subject to certain constraints plays an important role in all of his writings throughout the period.
The appearance of the concept of information in Adrian’s work raises at least two important questions: (i) What were the relevant factors that motivated Adrian’s use of the concept of information? (ii) What concept of information does Adrian appeal to, and how can it be situated in relation to contemporary philosophical accounts of the notion of information in biology?

The first question involves an account of the application of communications technology in neurobiology as well as the historical and scientific background of Adrian’s major scientific achievement, which was the recording of the action potential of a single sensory neuron. My historical thesis is that the centrality of communications technology in neurobiology in the 1920s made the informational analogy readily available to Adrian, and the discoveries afforded by that technology made the similarities between the transmission of messages and the transmission of nervous impulse perspicuous enough to merit the application of the concept of information to nervous conduction in a way that was not predetermined by the use of the technology alone.

The response to the second question involves an explication of Adrian’s concept of information and an evaluation of how it may be situated in relation to more contemporary philosophical analyses of the concept of biological information. I present an explication of Adrian’s concept of information insofar as it can be reconstructed from the relevant texts, and show how it can be used to derive a concept of arbitrariness or semioticity that provides a way of resolving some of the problems associated with recent attempts in the philosophy of biology to restrict the notion of information to those causal connections that can in some sense be referred to as arbitrary.

WILLIAM HARMS
billharms@billharms.com
Centre for Applied Ethics
University of British Columbia
227-6356 Agricultural Road
Vancouver, B.C.
Canada V6T 1Z2
phone: +1 206 352 7851 (home)

Four Concepts of Information

The concept of information tempts us as a theoretical primitive, partly because of the respectability lent it by highly successful applications of Shannon’s information theory, partly because of its broad range of applicability in various domains, and partly because of its neutrality with respect to what basic sorts of things there are. This versatility, however, is the very reason why information cannot be the theoretical primitive we might like it to be. “Information,” as it is variously used, is systematically ambiguous between whether or not it deals with intrinsic or relational properties, whether it involves semantic or merely causal relations, and whether it quantifies structural or probabilistic properties of systems. Furthermore, in none of its guises is information a conserved quantity, nor does it free us from dependence on preconceptions regarding what sorts of things there are in the world. This does not mean that the concept is without utility, however. Comparing four interrelated concepts of information emphasizes distinctions that will be critical to theoretical success, both in exploring the evolution of

cognition (and epistemology) and in the unification of theoretical biology. We shall consider Shannon's "mutual" information (and its close relative "self"-information), structural information, instructional information, and the overtly semantic "information that." Gene pool examples provide convenient cases where most of these kinds of information are present. We will see that, while information is not in itself the kind of single thing that can play a significant unifying role, analyzing its ambiguities may help make clear what unification will require.

JESSICA PFEIFER
pfeifer@umbc.edu
Philosophy Department
University of Maryland, Baltimore County
Fine Arts 541
phone: +1 410 455 2014

Applying Information Theory to Animal Communication: Some Lessons from Social Insects

In 1954, J. B. S. Haldane and H. Spurway applied Claude Shannon's information measure to data collected by K. von Frisch to quantify the amount of information about the location of food sources transmitted by the waggle dance of the honeybee Apis mellifera carnica. In 1962, E. O. Wilson compared this with the amount of information transmitted by the odor trail of the fire ant Solenopsis saevissima. Wilson found that these two very different systems transmit approximately the same amount of information. This provided an early and apparently clear example of the utility of applying Shannon's information measure to animal communication systems. Its use in this case seems to have revealed a similarity that would otherwise not have been apparent. It is therefore initially surprising that information theory has played only a minimal role in the analysis of social insect communication. However, when one looks more closely at how information theory is used in the study of animal communication, it becomes clear that its application is more problematic than it initially appears.

In this paper, I discuss how information theory has been used in the study of animal communication, as well as some of the difficulties faced in this application. Shannon's information measure is used primarily for the construction of optimality models of animal communication systems. However, to use the information measure in this way imposes certain requirements that make its application particularly difficult. To make these problems clear, I will focus on the transmission of information by the ant's odor trail and the honeybee's waggle dance. Both of these systems are relatively simple and well understood, and the application of Shannon's information measure to these systems seems initially very promising and relatively straightforward. They are therefore particularly suitable for revealing the benefits and difficulties of applying information theory to physical systems in general and animal communication systems in particular.

HANS-JÖRG RHEINBERGER
rheinbg@mail.mpiwg-berlin.mpg.de
Max Planck Institute for the History of Science
Wilhelmstrasse 44
10117 Berlin
Germany
phone: +49 30 22667 0
fax: +49 30 22667 299

Notions of Language, Information and Regulation in the Writings of François Jacob

My contribution will consist in a historical case study consisting in a close reading, analysis and reconstruction of the three fields of metaphors: 1) language and text, 2) information, instruction, and program, and 3) feedback, regulation, and control in the work of François Jacob. I will examine how these metaphors emerged in the formative work on phage and bacteria of the 1950s which led to the concepts of messenger RNA and of operon. Then I will trace how they were refined during the 1960s, and finally, how they came to shape Jacob's particular view of the "logic of life" around 1970. In the workshop, a few preliminary ideas on this program will be discussed.

TOM D. SCHNEIDER
toms@ncifcrf.gov
Molecular Information Theory Group
Laboratory of Computational and Experimental Biology (LEBC)
at the National Institutes of Health (NCI) - Frederick Building 469, Room 144
Frederick, MD 21702-1201
USA
phone: +1 301-846-5581
fax: +1 301-846-5598

Twenty Years of Delila and Molecular Information Theory

A brief personal history is given about how information theory can be applied to binding sites of genetic control molecules on nucleic acids. The primary example used is ribosome binding sites in \emph{Escherichia coli}. Once the sites are aligned, the information needed to describe the sites can be computed using Claude Shannon's method. This is displayed by a computer graphic called a sequence logo. The logo represents an average, and the mathematics easily allows one to determine the components of the average. That is, given a set of binding sites, the information for individual binding sites can also be computed. One can go further and predict the information of sites that are not in the original data set. Information theory also allows one to model the flexibility of ribosome binding sites, and this led us to a simple model for ribosome translational initiation in which the molecular components fit together only when the ribosome is at a good ribosome binding site. Since information theory is general, the same mathematics applies to human splice junctions, where we can predict the effect of sequence changes that cause human genetic diseases and cancer. The second example given is the Pribnow `box' which, when viewed by the information theory method, reveals a mechanism for initiation of both transcription and DNA replication. Replication, transcription, splicing and translation into protein represent the central dogma, so these examples show how molecular information theory is contributing to our knowledge of basic biology.

PETER STADLER
peter.f.stadler@univie.ac.at
Theoretical Biochemistry Group (TBI)
Institute of Theoretical Chemistry
University of Vienna
Währingerstrasse 17
A-1090 Vienna
Austria
phone: +43 1 4277 52 737
fax: +43 1 4277 527 93

Genotype-Phenotype Maps

The current implementation of the Neo-Darwinian model of evolution typically assumes that the set of possible phenotypes is organized into a highly symmetric and regular space equipped with a notion of distance, for example, a Euclidean vector space. Recent computational work on a biophysical genotype-phenotype model based on the folding of RNA sequences into secondary structures suggests a rather different picture. If phenotypes are organized according to genetic accessibility, the resulting space lacks a metric and is formalized by an unfamiliar structure, known as a pretopology. If recombination is taken into account, an even weaker structure, known as neighborhood space, must be

used.
Patterns of phenotypic evolution — such as punctuation, irreversibility, modularity — result naturally from the properties of the genotype-phenotype map, which, given the genetic accessibility structure, defines accessibility in the phenotype space. The classical framework, however, addresses these patterns by exclusively invoking natural selection on suitably imposed fitness landscapes. We extend the explanatory level for phenotypic evolution from fitness considerations alone to include the topological structure of phenotype space as induced by the genotype-phenotype map. The topological framework allows us to consider, e.g., the continuity of an evolutionary trajectory in an unambiguous way.

Lewontin's notion of ``quasi-independence'' of characters can be formalized as the assumption that a region of the phenotype space is a represented by a product space of orthogonal factors. In this picture each character corresponds to a factor of a region of the phenotype space. We consider any region of the phenotype space that has a given factorization as a ``type'', i.e., as a set of phenotypes that share the same set of phenotypic characters. Using the notion of local factorizations we develop a theory of character identity based on the continuity of common factors among different regions of the phenotype space. We also consider the topological constraints on evolutionary transitions among regions with different regional factorizations, i.e., for the evolution of new types or body plans.

EÖRS SZATHMÁRY
szathmary@zeus.colbud.hu
Collegium Budapest (Institute for Advanced Study)
2 Szentháromság u.
H-1014 Budapest, Hungary
phone: +36 1 224 8300
fax: +36 1 224 8310

Information and Replication

Replicators of interest in chemistry, biology and culture are briefly surveyed from a conceptual point of view. Systems with limited heredity have only a limited evolutionary potential, because the number of available types is too low. Chemical cycles, such as the formose reaction, are holistic replicators since replication is not based on the successive addition of modules. Replicator networks consisting of catalytic molecules (such as reflexively autocatalytic sets of proteins, or reproducing lipid vesicles) are hypothetical ensemble replicators, and their functioning rests on attractors of their dynamics. Ensemble replicators suffer from the paradox of specificity: while their abstract feasibility seems to require a high number of molecular types, the harmful effect of side reactions calls for a small system size. No satisfactory solution to this problem is known. Phenotypic replicators do not pass on their genotypes: only some aspects of the phenotype are transmitted. Phenotypic replicators with limited heredity include genetic membranes, prions, and simple memetic systems. Memes in human culture are unlimited hereditary, phenotypic replicators, based on language. The typical path of evolution goes from limited to unlimited heredity, and from attractor-based to modular (digital) replicators.

JONATHAN VICTOR
jdvicto@pop.med.cornell.edu
Department of Neurology and Neuroscience
Weill Medical College of Cornell University
1300 York Avenue
New York, NY 10021
USA
phones: +1 212 746 2343 (ofc)
+1 212 746 6521 (lab)
fax: +1 212 746 8984

Approaches to the Analysis of Information in Neural Activity

Understanding how neurons transmit, represent, and manipulate information is perhaps the most fundamental issue in neuroscience. While it is natural to look to information theory as a framework for attacking this problem, technical and theoretical considerations make such a program less than straight-forward. Consequently, a range of information-theoretic approaches to the analysis of neural data have been developed, that differ in generality, rigor, goals, and practicality

These ideas will be exemplified by examining responses of single neurons and local clusters of neurons in the primary visual cortex of the macaque monkey. Information-theoretic analysis is based on a metric-space approach that provides a rigorous test for simple hypotheses concerning coding, such as, that the precise times of spikes, or the intervals between them, or the neuron of origin, are informative. These analyses lead to the conclusion that spike trains are more than estimators of a firing rate, and that the detailed pattern of neural activity within individual spike trains and across neurons cannot be ignored. Additionally, the philosophy behind the metric-space approach, its relationship to the gamut of information-theoretic approaches, and open issues involved in broadening its applicability, will be discussed.